Deliverable D2.1
Title: Requirements and Use Cases

Dissemination Level:	**PU**
Nature of the Deliverable:	**R**
Date:	**4/4/2019**
Distribution:	**WP2**
Editors:	**ICCS**
Reviewers:	**UTP/THALES**
Contributors:	**ALL**

Abstract: This document identifies the scenarios, use cases and requirements (both user and system requirements) that must be covered by the SocialTruth framework. It establishes a common specification framework to proceed with the design and implementation of the SocialTruth tools and services. Furthermore, trust and blockchain aspects are elaborated.

Disclaimer

Revision History

Date	Rev.	Description	Partner
14/01/2019	0.1	TOC	ICCS
25/01/2019	0.2	Requirements elicitation webinars	ALL
4/02/2019	0.5	Partners assignments	ICCS
8/02/2019	0.6	Final deliverable structure and guidelines	ICCS
15/02/2019	0.6.1	Contribution added (Chapter 6)	Z&P
15/02/2019	0.6.2	Contribution added (Subchapter 4.2)	INFOC
18/02/2019	0.6.3	Contribution added (Subchapter 2.4)	LSBU
18/02/2019	0.6.4	Contribution added (Subchapters 2.1, 2.2, 2,3)	ESF
19/02/2019	0.6.5	Contribution added (Subchapters 4.1, 5.1)	ADNK
19/02/2019	0.6.6	Contribution added (Chapter 7)	TECOMS
20/02/2019	0.6.7	Updates in Subchapter 4.2	INFOC
20/02/2019	0.6.8	Contribution added (Subchapter 2.6)	THALES
25/02/2019	0.6.9	Contribution added (Subchapter 4.4)	DEASC
26/02/2019	0.7.1	Contribution added (Subchapter 4.3)	QWANT
26/02/2019	0.7.2	Contribution added (Chapter 1, Subchapter 2.7, Chapter3)	ICCS
05/03/2019	0.8	Integrated version	ICCS
01/04/2019	0.9	Contribution added (in Chapter 8)	UTP
02/04/2019	1.0	Final draft sent for peer review	ICCS, TECOMS, UTP
05/04/2019	1.1	Final version submitted the EC	ICCS
13/03/2020	1.11	Review and alignment with D2.1 based on 1st review recommendations	ICCS
23/03/2020	1.12	Adaptations based on partners feedback	ICCS, ALL
27/03/2020	1.13	Peer review and Pre-final version v.2	UTP
02/04/2020	1.14	Final version	ICCS

List of Authors

Partner	Author
ICCS	Pavlos Kosmides, George Koutalieris, Ioannis Loumiotis, Konstantinos Demestichas
Z&P	Giulia Venturi
INFOC	Lazar Iuliana, Păun Ana
DEASC	Sebastiano Bagarottto, Karen Nahum
LSBU	Manik Gupta
ESF	Sonia Collada
ADNK	Giovanni Parise, Antonacci Filippo
TECOMS	Guido Villa
THALES	Sebastien Keller
QWANT	Stan Assier, Christophe Servan, Noel Martin

Executive Summary

The Deliverable 2.1 is entitled Requirements and Use Cases and it is the outcome of the corresponding task T2.1. The deliverable identifies and describes the use cases and defines the end-users' requirements for the SocialTruth platform. These will guide all the subsequent research, development and demonstration activities of the project. Use cases are provided by the end-users of the SocialTruth platform, since the deliverable is focused on collecting and analysing user requirements related to SocialTruth's architectural blueprint. Specifically, the end-users have identified 4 use cases:

1. Use Case 1: Checking sources in the production process
2. Use Case 2: Digital companion for content verification
3. Use Case 3: Search engine rankings & advertising prevention for fraudulent sites
4. Use Case 4: External sources reliability check in the educational domain

In this deliverable, the actors, technologies, processes and relations of SocialTruth use cases are identified. An initial description of a set of relevant use cases has already been provided by corresponding users – members of the consortium - in the DoA - Part B [1]. In this document, these descriptions have been refined and mapped to specific system requirements, described in a formal and structured manner.

The sources used to create the initial list of requirements during the process of their elicitation, are:

- SocialTruth DoA, where the main functionalities of the SocialTruth solution have been described and general user needs analysed,
- Review and update of the technological State of the Art and the technologies that will provide the implementation of the SocialTruth Use Cases (presented in Section 2)
- Discussions, webinars, teleconferences carried out with the strong involvement of the project end-users (use-case actors) during the first months of the project,
- Detailed use-cases description provided in section 4 of the current document, and results of their analysis,
- End-user workshop session during the Bydgoszcz Plenary Meeting, 26-27 March 2019.

Based on these activities, the SocialTruth use cases have been further refined and are presented in Section 4. Human factors related to the use cases are elaborated in Section 5, reporting activities implemented as part of T2.4 "Sociotechnical and Human Aspects". The use cases have been further specified using Unified Modelling Language (UML), to ensure interoperability along all development stages of the project. The requirements have been prioritized using the MoSCoW methodology, based on active end-users engagement and contribution. This methodology has led to a complete set of system requirements described in a structured manner and in UML in Section 7. Minimum background information related to the MoSCoW methodology, for example the approach employed to distinguish functional and non-functional requirements, is also provided. This is especially important with respect to non-functional requirements related to security, privacy and ethical needs of the SocialTruth solution, which are monitored throughout the project lifecycle in T1.4, T1.5 and reported in three versions of "Data Management and Ethical Guidelines" reports on M6, M18 & M36.

Due to the highly innovative character of the SocialTruth research, new requirements, refining existing functionalities, will be considered during the roll-out of the use case demonstrators. In this way the overall SocialTruth solution will be informed based on expert users' feedback and subsequent research,

development and demonstration activities will be tightly linked with the actual stakeholders' needs, making it a solution that is mature enough to address real market needs and opportunities.

D2.1 was updated on February 2020 in order to further specify trust and blockchain aspects. These have are included in the new chapter 8 in this version, aiming to further refine these specific aspects as the basis for the design of the software architecture (D2.2, D2.3) and implementation of these services in WP4 and WP5.

Table of Contents

Index of figures

Index of tables

Glossary

CCTV	Closed Circuit Television
CFA	Color Filter Array
CMC	Computer-Mediated Communication
CIT	Critical Incident Technique
DCG	Discounted Cumulative Gain
DJPEG	Double JPEG
DoA	Description of Actions
EC	European Commission
HLG	High-Level Group
HCS	Hybrid Classifier Systems
KB	Knowledge Base
KBL	Knowledge-Based Learner
LL	Lifelong Learning
NLP	Natural Language Processing
SA	Sentiment Analysis
SCT	Social Cognitive Theory
SRM	Steganalysis Rich Model
TKM	Task-Based Knowledge Miner
UML	Unified Modelling Language
U&G	Uses and Gratifications
WBD	Weighted Belied Distribution
WP	Work Package

1. Introduction

During the last decade, there has been an unprecedented revolution in how people interconnect and socialize. **Social networks, media and platforms** are becoming the standard way of communication in our societies. However, the lack of control over the **digital content** being published and shared, has led the **information veracity** to be in dispute. This problem is targeted by SocialTruth and has been getting larger and larger over the years, with fake news and falsified articles spreading fast. Because of this high rate of false information spread, large **media organisations** face increasing pressure to respond quickly and accurately to breaking news stories.

SocialTruth advocates that: a) content verification cannot be entrusted to a single centralised authority; b) the aim should not be to devise the "single most perfect verification algorithm", since even the most sophisticated deep learning classification model is optimized at the time it is created –and as a result its accuracy deteriorates as new sources of fake news arise every day and the writing style of fake news changes, in order to successfully "game" and bypass verification checks; c) content verification should be easy and flexible to use "as a service" by individual users and professional organisations alike.

In response to these unmet challenges, SocialTruth takes into consideration the existing approaches, but provides a different focus: on creating an open, democratic, pluralistic and distributed ecosystem that enables easy access to various verification services, ensuring scalability and establishing trust in a completely decentralized environment.

The approach of the entire project is to keep media value chain stakeholders constantly involved, in order to produce a platform driven by their needs, with their close involvement in order to provide unique, in-depth knowledge of requirements that are not yet met with the existing solutions. In fact, the media value chain stakeholders of the consortium are providing fundamental contributions in WP2 – "*Requirements and Architectures*", in which this deliverable D2.1 is located. In this document the definitions of specific user requirements and use cases are proposed. These are formalized with the help of the technical partners, which will also specify the architecture of the SocialTruth Platform.

In particular, four complementary and representative use cases, led and coordinated by the media value chain stakeholders of the consortium, will be used to demonstrate the project goals and outcomes: (1) Develop a distributed content verification solution with a complex-free Digital Companion for online credibility verification of digital content found on web and social media; (2) Compose a digital content analytics and verification ecosystem with support for text, image and video, open to third-party service providers; (3) Leverage blockchain technologies to establish distributed reputation and trust in digital content sharing; (4) Deploy a distributed and thoroughly validated architecture (TRL-7) for the delivery of SocialTruth credibility evaluation services; (5) Test and evaluate the SocialTruth architecture and services in real-life use cases with the close involvement of end-users; and, (6) Introduce innovative business models for news, web and social media stakeholders, and provide support to the EU strategic agendas and policies. All of them are designed to address the real and current challenges posed to the media value chain actors, each will be demonstrated using the open data model, and the technologies and tools developed by the project. Such categories of use cases are used as reference for media and content verification to give a practical example of specific needs in different applications. Moreover, all

participating media industry actors are specifying the technologies that are currently available to them and the new capabilities to be implemented as part of the SocialTruth platform development. Technical partners have formalized such requirements as structured and interoperable outputs.

1.1 Objectives of this deliverable

D2.1 "Requirements and Use Cases" aims to report on the use cases and requirements and human-related aspects that should be covered by the SocialTruth solution. Since some types of end-users might have some very similar or very different requirements, D2.1 specifically caters to extract the common vs. specific requirements (e.g. about the semantic engine, data crawling, streaming analytics, open interfaces, etc.), properly encodes and prioritize them.

Presentation of the core technology of the project is elaborated in: Software Architecture D2.2 "Distributed System Architecture, Data Modelling and Interfaces" [M6]; D2.3 "Refined Distributed System Architecture" [M16]; and in the respective implementation reporting, e.g. in D4.1 "SocialTruth Blockchain" [M17] for the trust and blockchain supporting components. D2.1 will aim at contributing to an improved presentation of the technological aspects in these deliverables, when the relevance and role of the various components in detecting fake news within the SocialTruth, will be highlighted. Especially with respect to fake news detection, the employed techniques will be benchmarked (which of them constitute new and which improve upon existing methods) and innovation potential will be justified.

Thus, the scope of the deliverable is to identify and define the end-user requirements for the SocialTruth platform, that will determine the SocialTruth's architectural blueprint. Such requirements will be derived by proper use cases, where each media value chain actors will report a plausible content verification scenario as a reference for the development of the platform. An initial description of a set of representative use cases has already been provided by the Consortium in DoA - Part B [1]. In this deliverable, this description has been refined and mapped to specific system requirements, described in a formal and structured manner.

In D2.1 each Use Case has identified, actors, technologies, processes and relations to fully specify the SocialTruth functionalities. Both functional and nonfunctional requirements have been derived, taking into account potential areas where human factors are of relevance (based on the output of T2.4 "Sociotechnical and Human Aspects").

The use cases depicted by SocialTruth consortium have generated have been further refined using in Unified Modelling Language (UML), to ensure interoperability, and the requirements have been prioritised using the MoSCoW methodology [2]. With respect to the MoSCoW methodology we provide essential background information, however to a minimum and where it is absolutely relevant and necessary. For instance, we the difference between functional and non-functional types of requirements is summarised in two paragraphs of text in section 3.3, in order to make clear to the reader our approach regarding the analysis selection of functional and non-fuctional requirements. This project aims to "to create an open, democratic, pluralistic and distributed ecosystem that allows easy access to various verification services (both internal and third-party), ensuring scalability and establishing trust in a completely decentralized

environment". Objectives 2 and 3 also reinforce the idea of a blockchain-supported, open eco-system of distributed reputation and trust.

Given the complexity and innovative characteristics of the SocialTruth solution, our focus in this deliverable is to conceptualise the anticipated functions and uses of the developed applications. Non-fuctional aspects that are related with security, privacy and ethical considerations are addressed ex-post based on the specified functional characteristics that are presented in D2.1. For this purpose the project workplan foresees parallel processes that are active throughout the project lifecycle (Task 1.4: Data Management and Protection ; Task 1.5: Ethics and Gender Management) and are reported in three reports D1.3 : Ethics and Data Management Mid Report [18]; D1.4 : Ethics and Data Management Final Report [36]; D1.4 Ethics and Data Management Final Report [M36]. Findings of these investigations are to be reported in the Pilot implementation plans (D5.1 Overall Evaluation Plan [M14]) and recommended changes will be reported in D5.7 Completed Hands-on Trials and Recommendations R1.0 [21]; D5.8 Completed Handson Trials and Recommendations R2.0 [27]; D5.9 Final Lessons Learned and Overall Impact Analysis [34].

Still, D2.1 is providing a preliminary conceptualisation framework for the non-functional aspects of trust and blockchain support. The core project technology with respect to blockchain will be further analysed in D2.3 and D4.1. During the presentation of the core technology related to blockchain, efficiency and performance aspects will be addressed, for example, the SocialTruth DataModel and the blockchain data management aspects, including off-chain and on-chain storage considerations.

However, these technical aspects of the current design and implementation of the SocialTruth eco-system are not part of D2.1 scope.

1.2 Deliverable structure

The output of tasks T2.1 and T2.4 are presented in this document with the following structure. Chapter 2 provides the updated state of the art for the technologies that will be used for the implementation of the SocialTruth platform and services. These are the core technologies that will materialise the ambition of SocialTruth and demonstrate it in the Use Case pilots. In this way, chapter 2 specifies the technological needs of the planned Use Cases, in terms of tools already available to participating stakeholders and anticipated future functionalities supported by SocialTruth. In chapter 3 we provide our methodology for the requirement analysis, for the definition of the use cases and also the process of formalization of requirements derived from corresponding use case. In chapter 4 details we provide detailed and refined presentation of the SocialTruth use cases. This work was produced after a sequence of partners interactions (i.e. face to face meetings, webinars, online documents collaboration, email threads) to further elaborate on the DoA descriptions and update the vision for each use case. In chapter 5 the work of T2.4 is channelled to provide Socio-technical assessment and priorities for the platform and per use case. In chapter 6 Ethical, Privacy and security concerns are presented and a link is established with the relevant activities coordinated by the Ethics Manager. Chapter 7 ppesents the core outcomes of the deliverable in terms of the first batch of requirements to be implemented by the SocialTruth platform and the corresponding global UML Use cases.

Finally, chapter 8 provides and overview of the presentation of the core technology for the project, specifically considering the aspects of trustworthiness in a decentralized ecosystem such as the one SocialTruth aims to prototype. The role of blockchain technological capabilities to support the decision making of the verification process is also addressed.

2. Pilots technological needs assessment

2.1 Natural Language Processing (NLP) baseline

For the implementation of the aforementioned advanced Natural Language Processing (NLP) and semantic functionalities, SocialTruth will use as a starting point the semantic engine COGITO[1] (Figure 1) of partner ESF. This will allow SocialTruth to provide **advanced semantic solutions**, including semantic search, text analytics, ontology and taxonomy management, automatic categorization, automated self-help solutions, extraction of unstructured information and natural language processing. COGITO is powered by two main components:

- The *Disambiguator*, a multi-level linguistic engine able to disambiguate the meaning of a word by recognizing the context of occurrence of that word. Disambiguation can be described as the process of resolving conflicts that arise when a term can express more than one meaning, leading to different interpretations of the same string of text. The ultimate aim of such after lexical, grammatical and syntactical analysis, is to associate word to several concepts among those available in Sensigrafo. The disambiguator begins skimming the list of candidates for each word by considering the context in which every word appears to determine its meaning.

- The Sensigrafo, a semantic network that represents and stores the different semantic relations between the words of a language. Unlike traditional dictionaries where words are listed in alphabetical order, words contained in this database are arranged in groups of items expressing identical or similar meaning. These groups are connected to each other by millions of logical and language-related links. Our knowledge graph provides a domain independent representation of knowledge (or language lexicons) through concepts and their related meanings and the different relationships that exist among concepts. It is designed to interact with our semantic engine to resolve the ambiguity in meaning of each word, a fundamental step in the text analytics process. In the same way that human knowledge can be improved by learning new things, Cogito's knowledge may also be expanded through the acquisition of new knowledge from subject matter experts.

The COGITO semantic engine, which will be used as part of the SocialTruth solution, will start from this consolidated architecture, but will require further development to customize the semantic engine in the domain of disinformation and fake news, within the SocialTruth scope, in terms of taxonomies/categories definition, entities to be extracted, as well as emotions and writeprint to be identified from text. In order to apply this dedicated tuning to the semantic network, a manual or semiautomatic approach can be used. The second approach is preferable and is related to the use of machine learning techniques to create new ontologies automatically, which can be validated by human beings. This is a unique technique that can be used also in the opposite direction, so that the output of the semantic analysis serves as an input to the machine learning algorithms; these bidirectional processes allow to create a truly high-quality hybrid engine.

---

[1] http://www.expertsystem.com/cogito/
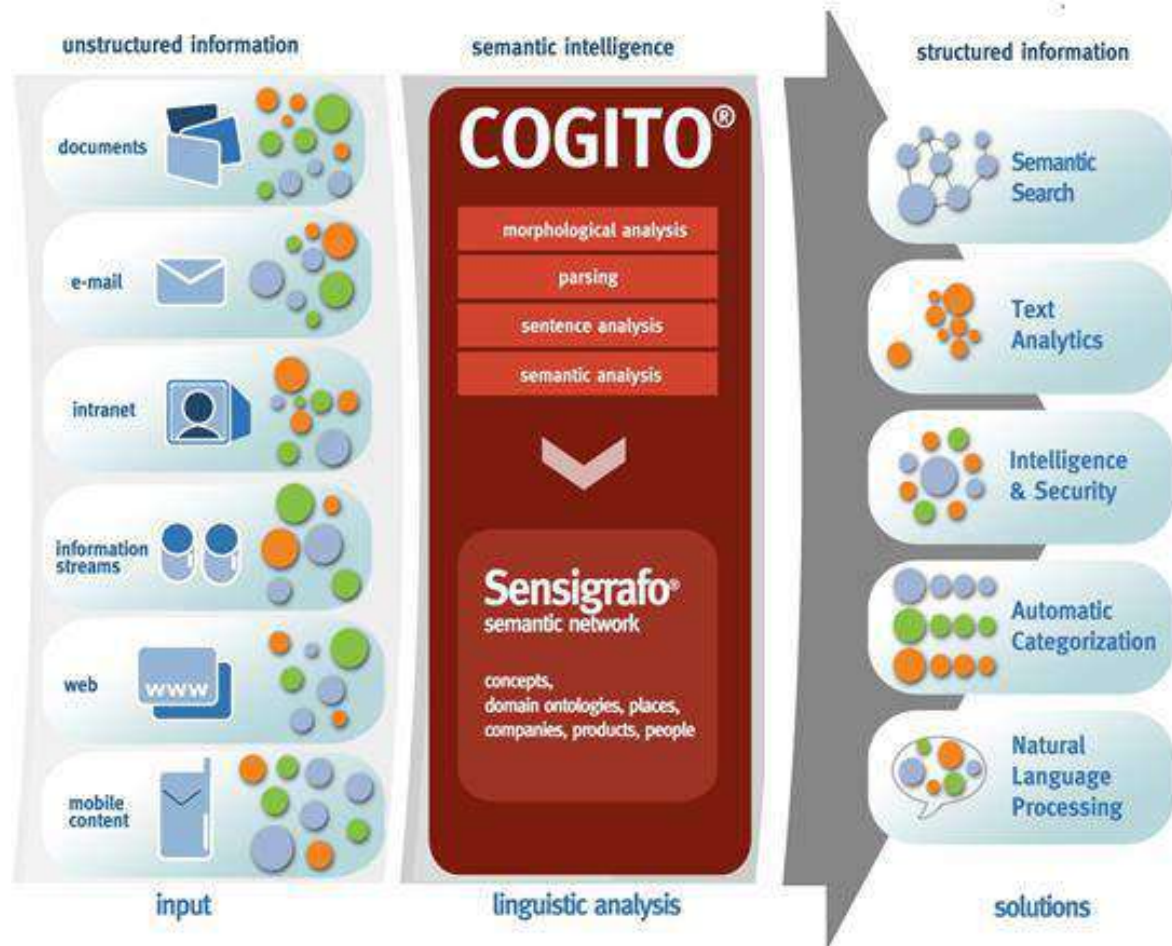
Figure 1. COGITO framework.

The Natural Language Processing with Cogito technology allows to:

- Add semantic features to your systems for knowledge management and Big Data analysis
- **Extract topics**, entities, **emotions**, facts, relationships between data automatically and integrate them into third-party systems
- Perform **temporal analysis** with time references extracted from the text
- Through a deep understanding of meaning and context in the processing of data, Cogito distinguishes between the high volume and various streams of information, resolving ambiguities more effectively than keyword or statistics-based solutions.
- Assess **authorship identification, stylometric characterization, writers' intentions**, text sources and formats.

2.2 Stylometric and semantic analysis

Natural Language Processing (NLP) support for both stylometric and semantic analysis, i.e. capability to detect author/writing style, similar content.

2.2.1 Stylometric analysis: Understanding Author and Writing Style

The semantic technologies of SocialTruth will be able to understand a story's author and writing style. This will help in the credibility evaluation process. To address this, SocialTruth will use as a starting point and extend the semantic technology of partner Expert System (ESF), namely its product COGITO, which facilitates deep understanding of language.

SocialTruth will extend this capability through advanced and innovative features like "writeprint" or stylometric analysis, which will make it possible to analyze the style of writing behind each story acquired from the web, social networks and any other textual source, with the scope to:

- Understand if an individual who is publishing a story has a style of writing that can be related and mapped to another style from a historical database.
- Understand if contents published on the web using different accounts or nicknames are actually related to the same person (i.e. clustering of different virtual IDs).

This stylometric analysis shall be based on a series of technical parameters, such as usage of short words, conjunctions, vocabulary richness and complexity, lexical differentiation, etc., which are strictly related to specific human factors and behavior, thus effectively defining a sort of "fingerprint" (writeprint) of the writer.

The Writeprint provides statistical and semantic text readability indexes with the purpose of targeting biometry and authorship assessment. Thanks to a proprietary algorithm, Writeprint can outline a document's readability and the level of education necessary to understand it; it also provides a full set of grammatical and structural analysis indexes.

Figure 2. Stylometric analysis.

2.2.2 Semantic analysis and Clustering of similar news

Semantic analysis describes the process of understanding natural language–the way that humans communicate–based on meaning and context.

The semantic analysis of natural language content starts by reading all the words in content to capture the real meaning of any text. It identifies the text elements and assigns them to their logical and

grammatical role. It analyzes context in the surrounding text, and it analyzes the text structure to accurately disambiguate the proper meaning of words that have more than one definition. Semantic technology processes the logical structure of sentences to identify the most relevant elements in text and understand the topic discussed. It also understands the relationships between different concepts in the text.

Using the semantic analysis capability of COGITO as a starting point, the goal is to identify the primary and secondary subjects of articles and stories, as well as other elements relevant for classification and entity extraction (such as People, Organizations and Places). Thus, SocialTruth will be able to classify the text according to a detailed, customizable tree of categories, enabling modification according to end-user requirements. These capabilities shall allow to:

- Make the information management automatic, more efficient and independent of subjective criteria.
- Immediately identify useful information and reduce search time, by simplifying access to content and enabling search by subject/topic.
- Cluster information according to customizable taxonomies and categories.

Clustering using categorization and entities extraction is one point. Extracting more semantic tags, such as relations, verbs and events, can lead the proposed solution to find correlations based on different analogy or similarity criteria. This is valuable, since, for example, the following two sentences are *different* from the statistical point of view, but *similar* from the semantics point of view: "*John is driving his car to home*" versus "*John is reaching home by his vehicle*".

2.3 Sentiment/Emotional Analysis

SocialTruth will also have the capability to extract a full set of emotions from textual content, not just using a standard positive/negative/neutral evaluation approach (sentiment) but also providing the capability to have deeper granularity on different kinds of feelings (i.e. stress, fear, trust, anger, etc.). When applied to sources like social networks and the web, this feature can be used, amongst others, as a bias estimator (e.g. detecting users that may have bad intention towards an event, a person, an infrastructure, etc.) or for assessing the public opinion about a specific topic of interest.

2.4 Multimedia verification support for images/photos and video

This particular task looks at examining the image content in online news articles and social media stories. The main focus will be to make use of the image forensic features to detect tampering. Image tampering can be of three different types- image splicing, copy-move attack, and resampling/histogram operations [3][4]. The method of copying one part of an image and inserting it into another refers to image splicing operation. The copy-move attacks practice taking a part of an image and copying it in the same image, to conceal something or to falsely add information. The third type of tampering called resampling/histogram operations is well intended and rarely changes image semantics.

There are several novel image forensics methods, which can discriminate different types of image manipulations carried out on images. One of the proposed methods employs a deep siamese CNN [5], which takes a pair of image, patches as input and decides whether they are identically or differently

processed. It is known that the CNNs are capable of learning accurate image editing features automatically from the training data. However, instead of learning features to classify image patches to different manipulation classes, the proposed method learns the features, which can discriminate different image editing operations.

Further research on image forensics consists of various approaches to detect the low-level tampering artifacts within a tampered image, including double JPEG compression, Color Filter Array (CFA), color array analysis and local noise analysis [7]. The ability of detecting whether an image has been compressed once or twice provides paramount information toward image authenticity assessment. Therefore, the capability of CNNs to capture double JPEG (DJPEG) artifacts directly from images has been exploited in [6]. CFA based methods analyse low level statistics introduced by the camera internal filter patterns under the assumption that the tampered regions disturb these patterns. Recently, local noise features based methods, like the steganalysis rich model (SRM) have shown promising performance in image forensics tasks. These methods extract local noise features from adjacent pixels, capturing the inconsistency between tampered regions and authentic regions.

2.5 Data crawling and Social Mining

This activity will provide a unified and flexible open Application Programming Interface (API) to enable transparent request of crawled data from the Web and Social media (Twitter, online blogs and fora). It will be based on QWANT's web crawling, indexing and retrieval capabilities used in its search engine technologies, as well as on its various types of information descriptors (content descriptors, social web descriptors, emotional descriptors, social influence sphere).

2.6 Blockchain in online identities and content traceability

Blockchain technology is for sure one of the most disruptive concepts that have emerged in the last years, with an impressive raise of interest from both the financial and technological communities. Initially intimately associated with Bitcoin, the first crypto-currency introduced in 2008 following the initial work of (Pseudonymous?) Satoshi Nakamoto [8] (who also published a first version of the related Open Source SW in 2009), the scope of applicability of the Blockchain technology has dramatically evolved, as well as its technological basis.

The disruptive nature of the Blockchain resides mainly in the new way it addresses the Trust problem, the most essential value for the development of the digital economy. In very broad terms, Blockchain aims at providing a string answer to one of the most difficult questions related to our digital world "*how can you trust people that you don't know?*". In this respect the Blockchain technology moves away from the classical pyramidal approach used in conventional systems.

In such pyramidal systems, some Roots of Trust are visible enough to be controlled by "the public".

- Examples (from the IT world) are: OS vendors, CPU designers, CPU manufacturers, AppStores, and of course PKI roots.
- Other examples are of more organisational nature such as government systems (justice, police, social benefits, etc.), and delegated authorities.

What makes these systems trustable is that their reputation prevents them from blatantly endorsing bad guys and that the users can "punish" them in case of failure (democracy, boycott, switch to alternate suppliers, aso..).

The Blockchain introduces a completely different paradigm for Trust based on a fully distributed approach. The Blockchain approach is based on three founding principles:

- All transactions are recorded in a Distributed Trusted Repository (*Distributed Ledger*) controlled by "the **public**".
- The control comes from the assumption that the good guys are more "**powerful**" than the bad guys (in Bitcoin, the "power" is the computing power of the "miners").
- The trust comes from the fact that the content of the repository is **publicly validated** by "the public itself", and that this permits to avoid having everyone validating everything.

Many variants of the Blockchain technology have been defined, which differ notably by the definition and the implementation of the three above mentioned "**public**", "**power**", and "**validation**" principles. But the main pillars stay[2]: 1. Security, 2. Scalability, 3. Decentralized/P2P, 4. Immutability and Data integrity. With these pillars and the above principles, different usages have been proposed in different domains.

In SocialTruth, the main objectives will be to provide a trust architecture to store the analysis performed by different actors and used by the same or other actors. All actors should have the possibility to check the evaluations performed and so, all actors will trust data stored inside the Blockchain.

2.7 Verification in social media through deep learning

Fake news detection is a non-trivial task, which requires multi-source information such as news content, social context, and dynamic information[3,4,5].
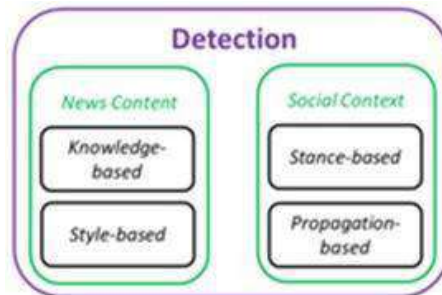
Figure 3. Fake news detection on social media: detection can be based on news content or social context.

---

[2] https://bravenewcoin.com/insights/the-5-pillars-and-3-layers-to-enterprise-blockchain-solution-design

[3] Shu, Kai, Amy Sliva, Suhang Wang, Jiliang Tang, and Huan Liu. "Fake news detection on social media: A data mining perspective." ACM SIGKDD Explorations Newsletter 19, no. 1 (2017): 22-36.

[4] Vedova, Marco Della, Eugenio Tacchini, Stefano Moret, Gabriele Ballarin, Massimo DiPierro, and Luca de Alfaro. "Automatic Online Fake News Detection Combining Content and Social Signals." In Proceedings of the 22st Conference of Open Innovations Association FRUCT, p. 38. FRUCT Oy, 2018.

[5] Castillo, Carlos, Marcelo Mendoza, and Barbara Poblete. "Information credibility on twitter." In Proceedings of the 20th international conference on World wide web, pp. 675-684. ACM, 2011.

Fake news detection methods have been recently classified into two categories - news content models and social context models. The methods belonging in the first category focus on the content of the news, i.e. the body-text, the title, and few additional meta-data and are referred to as content- based methods. Methods belonging in the second category focus on social features and signals, such as the engagement and interaction of users with given news on social media (e.g. "liking" a news on Facebook, "re-tweeting" it on Twitter, etc.) and are referred to as social-based methods.

Social-based methods constitute a more recent and promising strategy for fake news detection on social media and make use of additional information other than the news content[6] Examples of features, which have been used for this purpose, are the characteristics of users (e.g. registration age, number of followers, etc.) - or their opinions and viewpoints, exploited to assess the credibility of content. An alternative social-based strategy for fake news detection on social media is based on mapping the diffusion pattern of information. The rationale behind this strategy lies in the dynamics of social media sharing and interactions.

This task will concentrate on the social context based features and models that aim to utilize user social engagements as auxiliary information to help detect fake news. Stance-based approaches utilize users' viewpoints from relevant post contents to infer the veracity of original news articles. In addition, propagation-based approaches reason about the relations of relevant social media posts to guide the learning of credibility scores by propagating credibility values between users, posts, and news. The veracity of a news piece is aggregated by the credibility values of relevant social media posts.

2.8 Lifelong Learning Meta-verification methods

One of the current trends in machine learning is to develop intelligent systems that are able to learn consecutive tasks and to transfer knowledge from previously learned contexts to learn new tasks. Such capability is termed as Lifelong Learning Machines (LL) [9] and tries to mimic "human learning".

The motivation behind LL stems from sentiment analysis (SA) where aspect or opinion target extraction from opinion documents, is a fundamental task [10]. However, SA methods are not sufficient for building an intelligent system that can learn continually to achieve close to the human level of intelligence. LL aims to make progress in this very direction, and is becoming increasingly important because for systems that have to interact with humans and/or other systems, learn constantly in the process, and retain the knowledge learned in their interactions.

The LL conceptual architecture is provided in Figure 4. This is a general architecture and not all existing systems use all the components. SocialTruth will encompass only the appropriate LL algorithms that are specific to text verification.

---

[6] Kumar, Srijan, and Neil Shah. "False information on web and social media: A survey." arXiv preprint arXiv:1804.08559(2018).
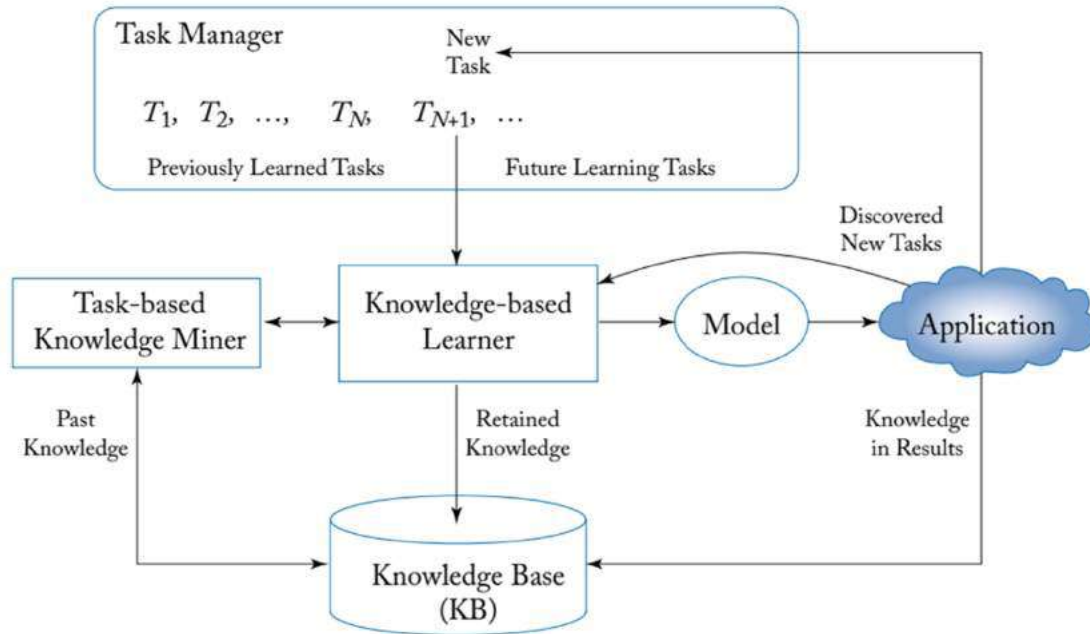
Figure 4. LL conceptual architecture [9].

In this conceptual architecture Knowledge Base (KB) is mainly for storing the previously learned knowledge; Knowledge-Based Learner (KBL) uses prior knowledge in learning. KBL leverages the knowledge in the KB to learn a new task; Task-based Knowledge Miner (TKM) is the module that mines knowledge from the KB specifically for the new task; Model is the learned model, which can be a prediction model or classifier in supervised learning, clusters or topics in unsupervised learning, etc; Application is the real-world application for the model, in our case will be the SocialTruth Meta-verification services; and, Task Manager (TM) receives and manages the tasks that arrive in the system.

There are also competing approaches such as DARPA L2M - Lifelong Learning Machines [11] aiming at fostering research in these directions. Therefore, the project's ambition is to use the Lifelong Learning paradigm to address challenges in fake news detection. In particular, hybrid classifier systems (HCS) and ensemble learning will be considered, as these have already been successfully applied to solve complex machine learning problems in various other domains.

In the context of SocialTruth, the Digital Companion will use at least one meta-verification engine powered by lifelong learning machines.

3. Methodology

The outcome of the deliverable has been obtained by following a precise methodology, which has been reported and described as follows. In fact, in order to proceed with the SocialTruth platform development, a set of system requirements shall be provided to developers by end users. To achieve this target, Use Case representatives and participating consortium partners have been asked to refine the Use Cases presented during the proposal and to further refine these initial concepts.

During January 2019 two (2) specifically designated Webinars were organised with the participation of all consortium partners with the sole purpose of validating and refining the four (4) SocialTruth Use Cases. Following these webinars, further bilateral communications have been implemented to further Ti ua These user requirements analysis webinars followed a semi-structured approach, aiming to elicitate new prospects and insights to the existing DoA Use Cases, following an exploratory co-creation approach that involved the active participation of partners.

The method we used required that all Use Case stakeholders would elaborate on Open Questions based on the "Think about the how, where, when, who, what, and why of the Use Cases". The following set of questions have been addressed to all Uses Case discussions, enabling a common requirements elicitation framework. The pool of questions we used was the following:

How

- How will you use this feature?

- Is this feature a process and, if so, what are the steps? Or, what questions can I ask to ascertain the steps?

- How might we meet this business need?

- How might we think about this feature a bit differently?

- How will we know this is complete?

Where

- Where does the process start?

- Where would the user access this feature?

- Where would the user be located physically when using this feature?

- Where would the results be visible?

When

- When will this feature be used?

- When do you need to know about…?

- When will the feature fail?

- When will we be ready to start?

Who

- Who will use this feature?

- Who will deliver the inputs for the feature?

- Who will receive the outputs of the feature?

- Who will learn about the results of someone using this feature?

- Who can I ask to learn more about this?

What

- What do I know about this feature?

- Or, what assumptions am I making about this feature that I need to confirm?

- What does this feature need to do?

- What is the end result of doing this?

- What are the pieces of this feature?

- What needs to happen next?

- What must happen before?

- What if….? Think of all the alternative scenarios and ask questions about what should happen if those scenarios are true.

- What needs to be tracked?

Why (validation: *are we building the right system?*)

- Is there any other way to accomplish this?

- Does this feature meet the business need and solve the problem we're trying to solve?

Following this requirements elicitation process, examples of use cases refinements have been identified and were used as reference scenarios for the set of features that the SocialTruth platform should provide. Such extended scenarios, were formalized in a structured manner, to embody the **new versions of the SocialTruth use cases,** in which media value chain stakeholders report their technological needs by describing a realistic operative scenario in which SocialTruth is supposed to be beneficial. The new

versions of the SocialTruth Use Cases contain several additional elements, in order to derive corresponding system requirements, specifically:

- **Factors identification** (actors/technologies/processes/relations)
- **Requirements** (functional/non-functional)
- **Prioritization** (must/should/could/won't)

3.1 Use case structure

The use case has to be described by both a static and a dynamic point of view. In the static part, end users should illustrate the context scenario of the verification situation, while in the dynamic part the sequence of verification steps has to be specified.

3.1.1 Static Scenarios

The static part of the use case description should be focused on introducing the environment where the verification of content is necessary. Possibly in this section should be reported general information that can be helpful to fully understand the context, such as type of content, media channels, falsification risks. These details have to be reported with the purpose of giving an overview on the environment for which the verification services will be beneficial.

3.1.2 Dynamic events

The dynamic part of the use case description should contain the timeline of the events related to the content verification, in order to highlight the relevant actions that are required in order to provide the verification services to end-users. The descriptions reported in this section will be useful to provide an overview of the particular data sources selected for the specific verification that the SocialTruth platform has to be able to support.

3.2 Factor Identification

The factors identification includes the enumeration and the description of the major elements that compose and characterize a specific investigation. Those are:

- Actors
- Technologies
- Processes
- Relations

By properly describing such elements in the corresponding use case, each end-user can help in defining the SocialTruth platform's requirement. Thus the developers will be aware on what kind of features are mostly expected to be delivered, and how the system will be likely used in real investigations.

3.2.1 Actors

In the actors' section, all relevant subjects involved for any reason in the investigation should be listed. They can be people or institutions, whether they belong to end-user or not. All the mentioned actors have to cover a clear and precise role in the use case.

3.2.2 Technologies

End-users should mention all technologies available at moment for the investigation in the corresponding use case. The various toosl should be reported in a self-explicative manner, in order to be identified with its primary function. Separately, also desired technologies that end-users would like to be implemented into the SocialTruth platform have to be reported.

3.2.3 Processes

In this section all relevant processes applied during the case resolution should be described. A process is intended to be any sort of action performed by an actor to progress with the use case (e.g. deploying information sources).

3.2.4 Relations

Once the previous elements (actors, technologies and processes) have been identified, in this section there will be reported relations among them, by explicating how they support the investigation.

3.3 Requirements mapping process

Based on reported information in the use cases description, end-users have to define a set of requirements for the SocialTruth platform, by identifying the particular feature needed from their own point of view. Both functional and non-functional can be specified.

3.3.1 System functional requirements

A functional requirement, as by software system engineering, is a description of the feature required by the end user. Functional requirements deal with what the system should do or provide, the operations and activities that a system must be able to perform. They include, for instance, description of the required functions, outlines of associated reports or queries, and details of data to be held in the system. Functional Requirements can include the descriptions of data to be entered into the system, descriptions of operations performed by each module, descriptions of work-flows or reports

3.3.2 Non- functional requirements

Non-Functional Requirements, on the other hand, are a description and, where possible, target values of associated non-functional requirements. Non-functional requirements detail constraints, targets or control mechanisms for the new system. They describe how, how well or to what standard a function should be provided. For example, levels of required service such as response times, security and access requirements, technical constraints, required interfacing with users' and other systems, and project constraints such as implementation on the organisation's hardware/software platform [12].

3.4 Output standardization and organisation

Requirements have to be reported by following the MoSCoW prioritization. Once use cases have been structured in a formal manner, the technology partners will elaborate a UML description that will be integrated in D2.1 report and then submitted to the partners involved in the other WPs

3.4.1 MoSCoW prioritization

By using this method of prioritization, user requirements are marked as

- *must*, if they are mandatory,
- *should*, if they are strongly preferred,
- *could*, if they can be omitted in case of technical issues and
- *won't*, if they don't have to be implemented at all.

3.4.2 UML format

The requirements, both functional and non-functional have been analysed using UML Requirements Diagrams. The Requirements model is a structured catalogue of end-user requirements. These are represented as either Requirement or Feature elements. The model is divided into two sub-catalogues:

- The Functional requirements package contains requirements and features that represent functional behaviour and features that the system under development must support.
- The Non-functional requirements package contains constraints and performance levels the system must meet. For example, non-functional requirements include response times, transactions per second, security strength.
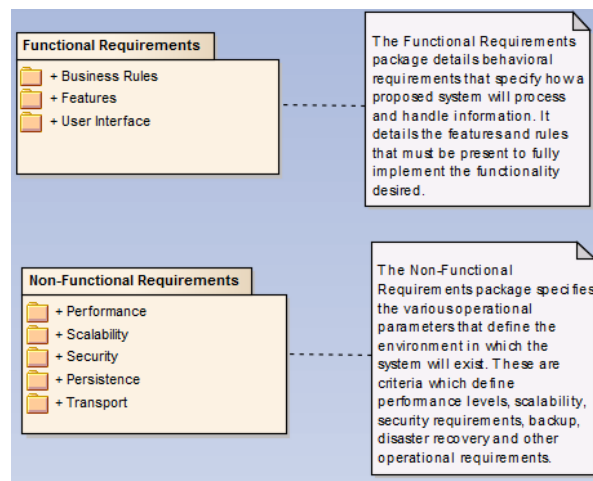
Figure 5. The general overview of Requirements UML Diagram.

Further the Functional Requirements can be analysed from different perspectives using following (and more detailed) diagrams:

- Business Rules – a catalogue of explicit business rules which are required to be implemented within the current project. Business Rules are typically executed during program execution and control the processing of information and transactions.
- Features – typically describe discrete pieces of functional behavior that yield a specific result.
- The User Interface – contains high level descriptions of end-user visible screens and forms which are required to support the proposed system.

Similarly, non-functional requirements can be group according to following views:

- The Transport package defines constraints and requirements affecting the transmission of information between nodes. Networks, relays, protocols, quality of service and even transmission of physical media are included here.

- The Persistence package details operational and performance criteria relating to the storage of information, including where relevant, redundancy, back-up, database system, files and other persistent storage mechanisms.
- The Security package details requirements with respect to both data access (information security) and physical security (access to servers and other critical hardware).
- Scalability requirements define the operational parameters with respect to system size, number of transactions, capacity, number of users and node distribution.

Performance requirements define parameters such as transactions per second, network latency, form load times and other measurable aspects of the system that govern overall speed and responsiveness

4. Use Cases Categories

4.1 Use Case 1: Checking sources in the production process

Adnkronos, as a press agency, is a primary source and occupies the first step in an ideal information staircase. It is therefore essential that it produces verified and reliable news for other media and final users. As pilot leader, Adnkronos will approach SocialTruth from a journalistic point of view, by exploiting its output to shorten the production workflow. This process involves checking other and more reliable sources, looking for a direct confirmation from the original source, fact-checking the statements, and so on.

4.1.1 Motivation (AS-IS)

Sometimes the source of a story is a subject that is not clearly identified or identifiable. In similar situations, the verification of the source reliability becomes much more complex and can occupy a significant portion of the journalist's working day. Moreover, Adnkronos operates in a market in which it is essential to be timely. In these cases, it is necessary to contact other sources, known and accredited, to check the information and decide if such information can become news, to be offered to other media and final users.

4.1.2 Pilot targets (TO-BE)

From SocialTruth Adnkronos journalists, as end-users, should expect to be able to analyze a much higher number of sources, filtering them through a credibility index, that goes beyond mere statistical similarity or semantic relevance, eliminating fake news in advance and flagging this up to the user, through filters that also verify the credibility and authoritativeness of the information

4.1.3 New concepts identified during requirements elicitation

The checking sources process takes place outside the CMS, so a full integration of SocialTruth functionalities in the core editorial software is not an absolute priority for this use case. On the other hand, it will be more useful to provide the journalists with a SocialTruth-empowered web browser plugin. Below we address some specific usage scenarios that come from real-world issues that our journalists are facing.

4.1.3.1 *1.1 Use case 1: collection of instances and sub-use cases*

Since there is no fixed and reproducible pattern in the activity of verifying the journalistic sources, rather than delineating a hypothetical use case instances, we will examine real cases that could be sorted out by using SocialTruth, along with other traditional means.

Example one: unaccredited tweet (real case)
A tweet reported a fact that lately became the source of one of the main news reported by Italian media. It came from a user who has about 6,500 followers and is not a verified account. It introduced itself as an Independent support for people crossing the Mediterranean Sea to the EU. The tweet referred to the presence of about 100 migrants on a boat in the Mediterranean. The news was extremely relevant, if confirmed. However, this was not a verified profile. It was not even an organization of the highest level, as can be guessed from the number of followers so this tweet could automatically generate a news.

Some other checks and confirmation by institutions that represent reliable sources were needed (the step that will involve SocialTruth). Once contacted, the Italian coast guard and the Libyan coast guard didn't manage to give enough information to confirm the whole story, but the UN agency for migration did. Only at that point the fact became a news. This verification process allowed us to consider the source reliable and, in relation to a given topic, will allow us to use it more easily in the future.

Example two: unconfirmed news (real case)
We have received the news about two Italian divers missing in the waters of Santo Domingo. Some Dominican sites and Facebook pages have spread the unverified news concerning the death of divers. The news was too soon considered true and reliable by some Italian newspapers. As a press agency, we've called the Italian authorities in Rome, the Italian consulate in Santo Domingo and the local authorities involved in the research operations. The news has been defined 'not true yet'. No official source confirmed the diver's death. We've spent some hours to get the truth (the step that will involve SocialTruth). Sites and newspapers that have spread the news have not performed an adequate fact checking. So they can be considered as unreliable sources.

4.2 Use Case 2: Digital companion for content verification

The rapidly increasing influence of the social media and the shift from the classic media sources to the online information, without professional rules and ethics, created conditions for the appearance of new and malicious ways to influence many persons to act, to buy, to vote using fake and fabricated information. Many scandals on manipulated elections, useless products promoted as wonder solutions for different applications, fake charity campaigns are more and more found in our day-by-day activity.

False news nature is established by untrue content in news format as well as image digital file formats or other media formats and by a capacity to travel even more, than the authentic ones [13]. Misinformation is spread faster through social networks, especially when news is about crisis and health problems. False news were often created to influence the public for ideological and/or financial gain [20].

There is global worry over false news and the probability that it can impact on political, economic, and social well-being [26]. A data set of rumor cascades on Twitter from 2006 to 2017 were investigate on "*how false news spreads and the outcomes highlighted that false news reached more people than the truth*" [26]. The researchers found that "*false news was more novel than true news, which suggests that people were more likely to share novel information. Whereas false stories inspired fear, disgust, and surprise in replies, true stories inspired anticipation, sadness, joy, and trust. Contrary to conventional wisdom, robots accelerated the spread of true and false news at the same rate, implying that false news spreads more than the truth because humans, not robots, are more likely to spread it*" [26]. Rumors tend to be posted and re-posted quickly on social media, especially when there is no information from the authorities. If measures are not taken quickly, misinformation can result in more serious damage than the disaster itself [28].

Social media companies and state governments try now to implement new politics to stop this negative phenomenon, but the results are not conclusive, and many efforts must be made to solve this issue.

For example, European Commission high-level group (HLG) published its final report on misinformation, relying on the input of experts from around the world.

According to this report the best responses to disinformation are multi-dimensional approach based on 5 supports and involving of actual, inter-dependent activities [17]:

1. "*enhance transparency of online news, involving an adequate and privacy-compliant sharing of data about the systems that enable their circulation online;*

2. *promote media and information literacy to counter disinformation and help users navigate the digital media environment;*

3. *develop tools for empowering users and journalists to tackle disinformation and foster a positive engagement with fast-evolving information technologies;*

4. *safeguard the diversity and sustainability of the European news media ecosystem*

5. *promote continued research on the impact of disinformation in Europe to evaluate the measures taken by different actors and constantly adjust the necessary responses*".

The HLG recommends, as a first step, a "*self-regulatory approach based on a clearly defined multi-stakeholder engagement process and a binding roadmap for the implementation of specific actions*" [17].

From Jang perspective who inspected the root content, producers of original source, and progress patterns "*to improve the flow of quality data and combat fake news, it is essential to identify the origins and evolution patterns of false information*" [19].

One of the main objectives of the InfoCons association is to protect consumers from any product or service that may affect their personal life in one way or another. In the last time, the development of online communication and social media expose all consumers to fake news which can affect their decisions and can affect them emotionally or even material. InfoCons is in close contact with everyday Internet users, devoting large efforts in informing and educating consumers about existing risks and threats that might affect them. In addition, InfoCons provides advice to users and conducts local studies and campaigns in the areas of the Internet, food, health, financial education, environment and sustainable development, as well as on other topics that concern modern life in terms of fair, safe, real values, ethics and integrity.

Even InfoCons is affected by this increasing number of fake news. Significant effort is spent in preparing and verifying the actions and content that will be used in relation to consumers, as it is crucial for the association to maintain its credibility and reputation.

The content verification is necessary both for consumers and professionals, but at this moment is an individual, difficult and time-consuming process.

4.2.1   Pilot targets (TO-BE)

To approach our target group better and attain our project's aims, we might involve relevant stakeholders [17], including:

- o **online platforms,**
- o **news media organisations,**
- o **journalists,**
- o **fact-checkers organisations,**
- o **independent content creators**
- o **the advertising industry**

Map out our stakeholders and evaluate in which steps to be involved are important actions.

Stakeholders' roles and tasks will be grouped into a Stakeholder Communications Worksheet [25] (including communications approach, key interests and issues, desired project roles, etc.).

User communities reachable through the InfoCons association will be asked to participate in the trials by installing and using the Digital Companion as a plugin in their web browsers. In their daily routine, by making use of the Digital Companion from within their browser, they will be able to request the automatic verification of a digital article or social media post that interests them. The Digital Companion does not calculate the credibility score itself but merely communicates with the meta-verification engine of SocialTruth, which in turn takes care of the rest. These normal everyday users will be asked for their feedback on the usability and utility of the SocialTruth solution.

In parallel, the staff of InfoCons will also be able to use the Digital Companion (in the form of a web front-end) for content verification simultaneous with 1-3 fact-checkers organisations to compare the outputs.

As professional users, the Digital Companion is expected to save them time in the preparation of their studies, providing verified results. In addition, the professional user will be able to configure and finetune the individual verification services applied to the content, thus providing a more tailored user experience.

4.2.1.1 *Individual users Use Case*
Generalities:

Testing will be carried out on the following areas of consumer protection:

a.   Food safety and food security

b.   Safety and security of non-food products

c.   Financial services

Requirements: The requirements of a trial destined to consumers will be:

A. Usability and accessibility testing

B. Reaction and opinion of end-users testing

C. Identify solutions to improve the Digital Companion

Accessibility and Inclusion

Accessibility by all users as target group is a very important issue for all organizations. Thus, the project partners will ensure that accessibility of Digital Companion will be the same for all users.

Primary Actor: Independent content creators' as Reader/Reproducer/Author

Goal: The individual user uses the Digital Companion to verify information and give feedback. The solution must be very easy to use, user-friendly and reliable.

Story:

- Step 1: The individual user installs the Digital Companion plugin in their web browser(s). The solution must be compatible with the most used browsers like Chrome, Safari, Firefox, UC, Opera, IE, Samsung Internet, Edge, etc.;
- Step 2: The individual user requests the verification of information (news/post). When the user found an article on an internet page or a post on social media account can request the verification of the confidence of the information in a very simple way (one-two clicks);
- Step 3: The Digital Companion returns the result of verification. In the pilot stage, the result of the verification will be not very elaborated, but good enough to can be compared with the results obtained by the individual user from a personal verification;
- Step 4: The individual user makes his own verification. To can appreciate the accuracy of the Digital Companion instrument, the individual user will use the classic methods to verify the information: search engines, related social media posts, search for the original information; in par
- Step 5: The individual user gives feedback on automatic verification. Based on his personal investigation, the user will appreciate using a score scale the reliability of the information generated by Digital Companion, for each investigated news/post.

4.2.1.2 *InfoCons Use Case*
Primary Actor: InfoCons as Professional User

Goal: InfoCons use the Digital Companion to verify news/posts and give feedback

Story:

- Step 1: InfoCons use the Digital Companion as web front-end in all current documentation activities. The application will be used to test all the sources used or potentially to be used for different documents production, posts or information campaignes;
- Step 2: InfoCons compare the results generated by the Digital Companion with the results of the fact-cheking verification. As an association acting for the consumer protection, InfoCons verify all information used in his activity and coming from not verified sources. Even if the information is coming from trust sources, the verification is necessary many times, when this is possible. During this pilot phase, the fact-checking verification will continue and the comparation process will have only three possible results: the result of verification using Digital Companion is correct, is not correct, or the result cannot be confirmed or partial confirmed; fact-checking [18] verification by InfoCons:
  - Distinguish the exact claim that is to be confirmed;

- o Contact the person or organization who makes the claim to ask them about the source or proof of their claim. Ask the person making the claim for evidence. It's basic fairness to tell someone you are fact-checking them, so they can give their side of the story!
  - o Inspect opinion of other fact-checkers have found before you;
  - o Do a Google's advanced search settings allow you to look at specific sites and time periods; use proper keywords and file type (file:pdf) or domain name
  - o Contact own experts in the field to argue their point of view and/or do a Databases advanced search (Deep Web);
  - o Write the reports based on scientific argue; highlighted the evidence step by step and inserting the sources used;
  - o An independent editor will revise the manuscript draft; if one verdict is fully agreed by author/s and editor/s, the article will be published.
- Step 3: InfoCons give feedback on automatic verification and propose improvements and fine tunes of the application. In the case when the result offered by the verification of the Digital Companion instrument is not correct, InfoCons will offer not only the negative mark, but also will indicate the reasons for this, helping the production team to improve the analysis algorithm. In parallel, the staff of InfoCons will also be able to fact-check to compare the outputs.

4.2.2 New concepts identified during requirements elicitation

In addition to the requirements presented in this deliverable, new needs will be investigated during the use case demonstrations through a process of elicitation. Eliciting requirements is part of the definition of requirements before their validation. Usually used elicitation procedures are the stakeholder consultations or workshops to debate their perception of the requirements [24].

For the collection of feedback during the Digital Companion demonstrations, the chosen methodology will be more often a combination of several methods [21].

InfoCons will organized online public campaign which are of course a form of debate media, interviews on television, semi-structured questionnaire, brainstorming, use cases and user's behavioral data.

Stakeholders desires and expectations will be grouped into a set of clear and concise need declarations that are suitable as a starting point for new requirements with proper adequate definitions and actions attached to them [15]. The new approach employs a wider range of possible techniques and captures more requirements information on the needs of a consumer. Card sorting [14][16] and Delphi method [23], Weighted Belief Distribution (WBD) [27] and critical incident technique (CIT) [22] can help to decide new software requirements.

4.3 Use Case 3: Search engine rankings & advertising prevention for fraudulent sites

4.3.1 Motivation (AS-IS)

One of the most common ways for a fake news website to increase its traffic (and, thus, its advertising-based revenues), is to promote itself through various social media and search engines. In particular, the

easiest way for a website to gain readers and become recognizable, is through the available advertising features provided by popular social networks (e.g. Facebook) and search engines (e.g. Google). By advertising fake news to targeted user groups, the fraudulent site can increase user hits and simultaneously raise its visibility, either by appearing in higher places within search results or by being re-posted from users who read their fake news.

Qwant.com results are ranked to be displayed so as to optimize end user experience and satisfaction. Hence ranking is crucial for a Web search engine. Over pertinence it takes into account multiple criteria such as popularity, consistency, topical which consists in moderation of popularity according to the thematic proximity, and other signals which fall under the industrial secret of each search engine. The so ranked results are then displayed on search engine result page such as the one hereunder.
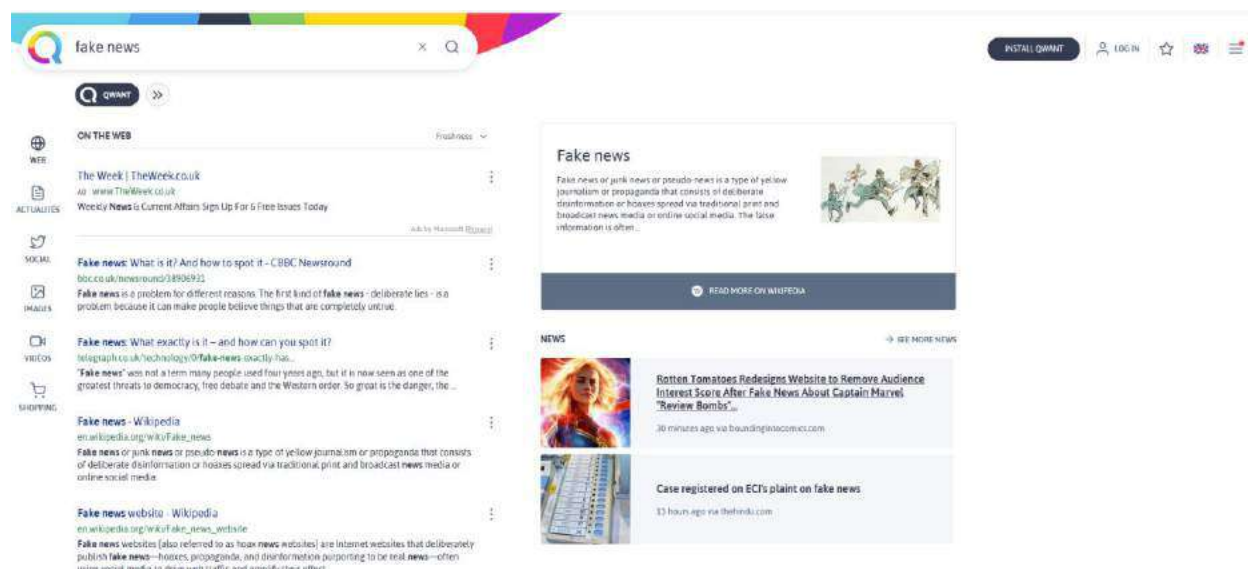
Figure 6. Ranked results displayed on QWANT search engine result page.

4.3.2   Pilot targets (TO-BE)

Qwant will include into qwant.com ranking criteria which will be used either to rank results within the standard ranking process of qwant.com, and/or, depending on the specifications defined with end users UX UI specification groups, mark the results with a fraudulent likeliness indicator. Results might then be displayed on search engine result page such as the shown hereunder.
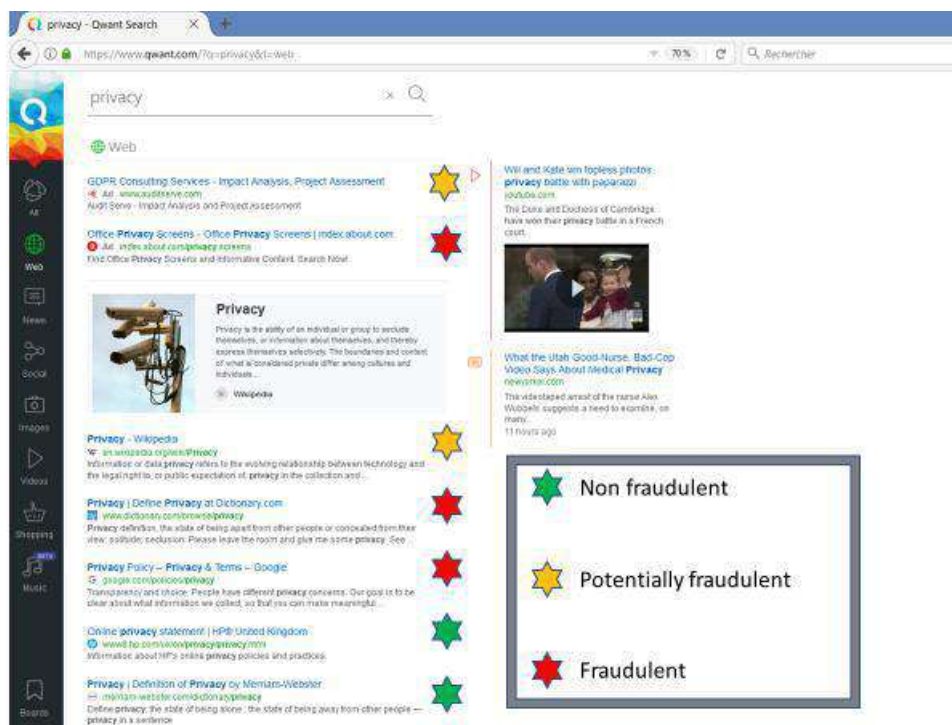
Figure 7. Proposed representation of QWANT search engine result page including fraudulent likeliness indicator.

Depending on the SocialTruth platform response time, QWANT will use part of its audience related to its research projects to test the fraudulent likeliness criteria. There will be two different test modalities. As a first step a dedicated UI will be displayed and then if all the conditions are meet one may propose to integrate it within the complete UI.

4.3.3 New concepts identified during requirements elicitation

Real time verification of website fraudulent likeliness thru the social truth platform and real-time reranking of results with the fraudulent likeliness criteria. In order to optimize resources fraudulent likeliness of popular websites will be store in cache memory.

4.4 Use Case 4: External sources reliability check in the educational domain

4.4.1 Motivation (AS-IS)

DeAgostini Scuola has created several portals and applications that support teachers in the creation of original contents (e.g. lessons) and in the combination of existing resources into workable teaching material. In particular, **the CreaLezioni platform** is a tool implemented by DeAgostini Scuola to support teachers of different grades in the creation of their lessons and training material (on multiple disciplines), exploiting and combining available material, lesson templates, traditional contents (provided by the editor), and resources/content/information retrieved from external sources, i.e. online articles and multimedia content.

As of today, besides the structured contents available on books, teachers increasingly make use of online-available resources to tailor/improve their everyday lessons (e.g. YouTube videos, references to online articles).

This approach allows teachers to be more effective with today's students (digital natives) and better engage them during frontal lessons. Besides standard lessons (e.g. history ones) there are several subjects that request thematic insights (e.g. civics), for which the use of online sources is essential.

In this context, if fake news or inappropriate sources (poorly referenced/misleading/inaccurate) are used/linked during lessons, there is a considerable impact for the quality of education of pupils and youths, which typically have lower critical capability to distinguish fake news/inaccurate information in comparison to adults' audience.

4.4.2 Pilot targets (TO-BE)

The use of the SocialTruth content verification solution in this domain can help to prevent the distribution of fake/inaccurate information in the school domain. In this light, the target use case will be focused on the analysis of external resources linked by teachers in the design of their lessons, of their reliability and of the credibility of the story, news or information reported therein.

Thus, the main goal of this use case it to prove the effectiveness of the project's solution in the educational environment. This aim will be realized by addressing three mutually reinforcing objectives, namely to:

1) test the project methodology and IT capabilities in an editorial environment (*technological validation*), 2) to validate the online contents linked by teachers in ad-hoc tailored lessons elaborated through DeAgostini Scuola platforms (*content verification*)

3) to improve the teachers' understanding about the appropriate use of online sources within the educational domain (*awareness raising among specific stakeholders*).

4) educate pupils on how to recognise trustable content online, also by means of disseminating the SocialTruth project purpose and tools (e.g. fake news)

According to the aforementioned four goals, the target use case will deal with the following processes:

1) ***Testing of the project methodology in an editor-relevant environment***, using the project's ecosystem for establishing distributed trust and validating the project methodology covering all the steps expected (e.g. author and writing style detection, text mining and semantic analysis, sentiments and emotions detection, image/photo and video verification, as well as blockchain infrastructure); under this process, the end-users involved will also assess the possibility to define **a sort of "quality label"** for editors that provided contents verified through the project's ecosystem;

2) ***Using the content verification services to validate the online contents linked by teachers in ad-hoc tailored lessons***, created through the use of the DeAgostini Scuola platforms; this step will allow to experiment and gather feedback on the actual effectiveness, efficiency and benefits that the proposed solution can provide in the educational domain, by verifying in near-real time the sources (e.g. news,

articles, videos, webpages) that a teacher is linking/reporting while designing its lesson through the DeAgostini platforms; further details on the implementation and benefits of this process.

4.4.3   New concepts identified during requirements elicitation

The Main Stakeholders identified are Teachers and Students.

User Experience

User experience of the teacher desktop was presented, with SocialTruth functionality as an add-on. Trustability/reliability of educational resources would be added as a quality characteristic for the product.

Integration

- **Phase 1: Integration with CreaLezioni**
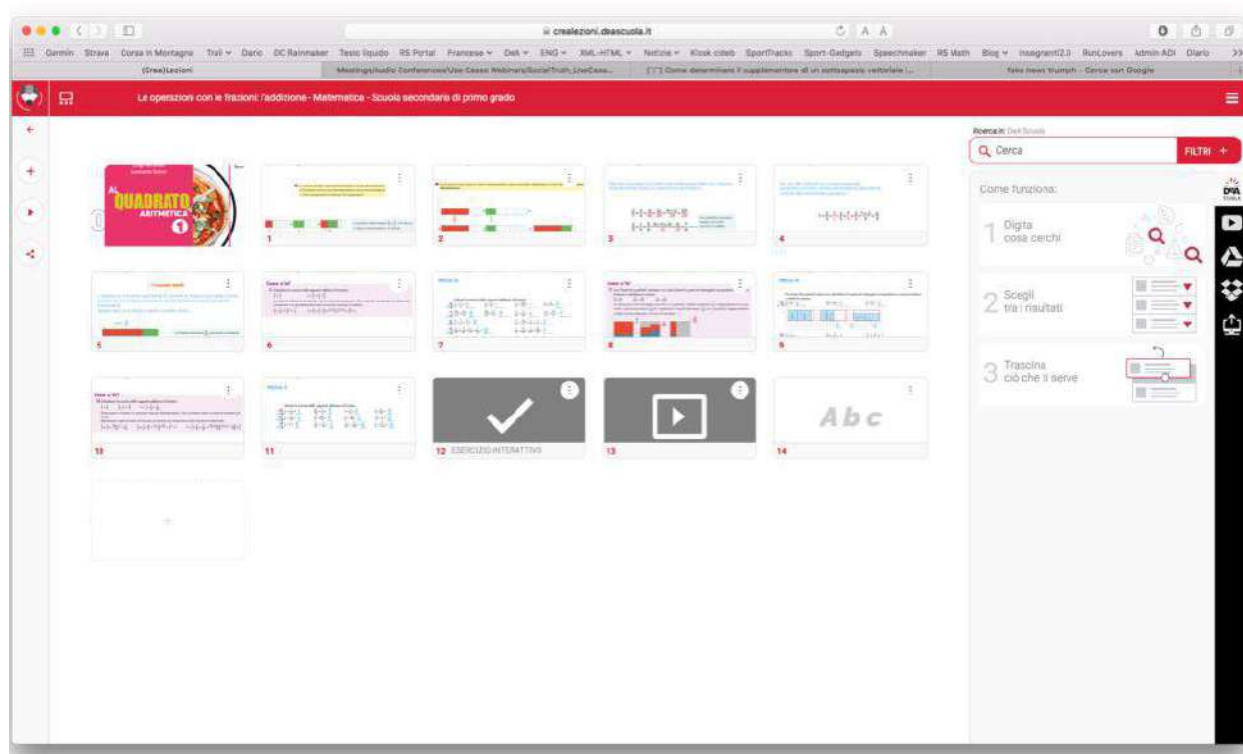
Figure 8. The CreaLezioni platform.

The Phase 1 would be the integration with the CreaLezioni. The platform has integrated Google and Youtube as web content and movies search engines so ideally *CreaLezioni web link search engine* should be integrated with the SocialTruth Plugin *(widget?)*. The user experience can be the follow:

- typing a web address the user will get the query results with a sort of trust identity card of the content measured with some parameters/variables (tbd)

-It would be important to highlight the reasons why the content could be not trustable (depending on the algorithm implemented by the consortium)
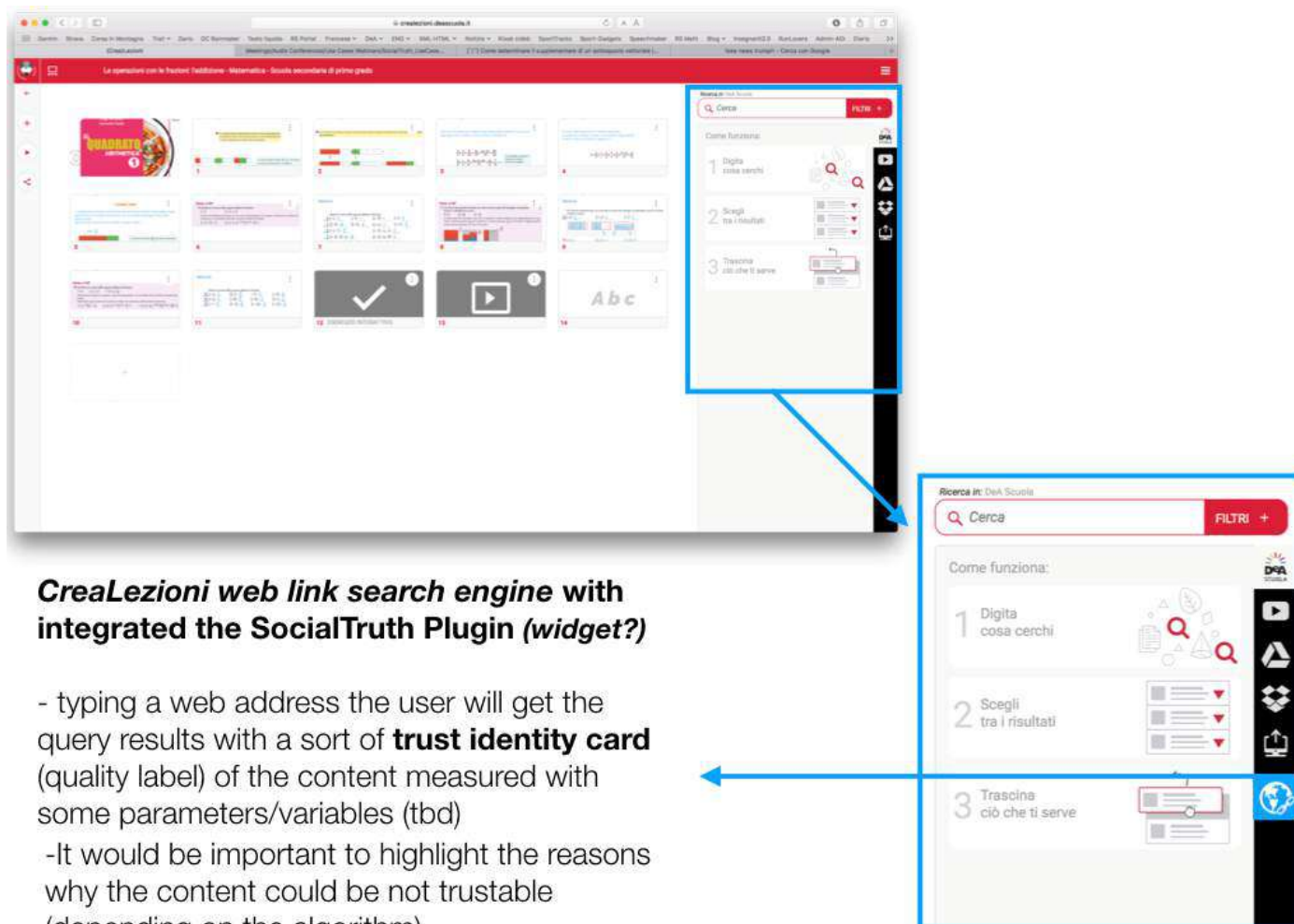
Figure 9. Integrating SocialTruth with the CreaLezioni platform.

The Trust identity card, **a sort of "quality label"** that provided contents verified (as specified on 4.4.2) could help teachers and pupils to understand if a content (or video) is appropriated or not and in case not the reasons why it should be refused and not added to the lesson.

It would contain all the variables and parameters that the consortium will decide to use to measure it.
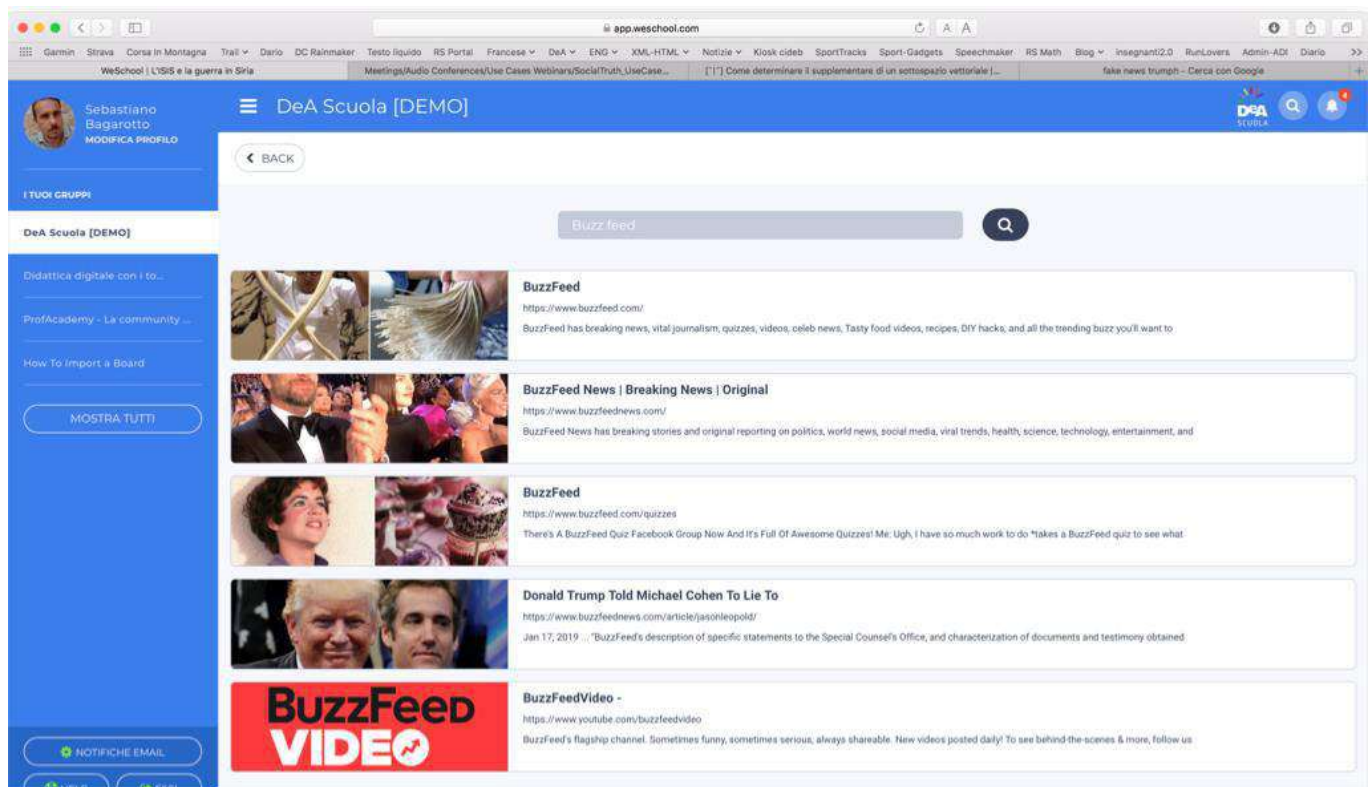
Figure 10. Screenshot from DeA Scuola partner LMS platform.
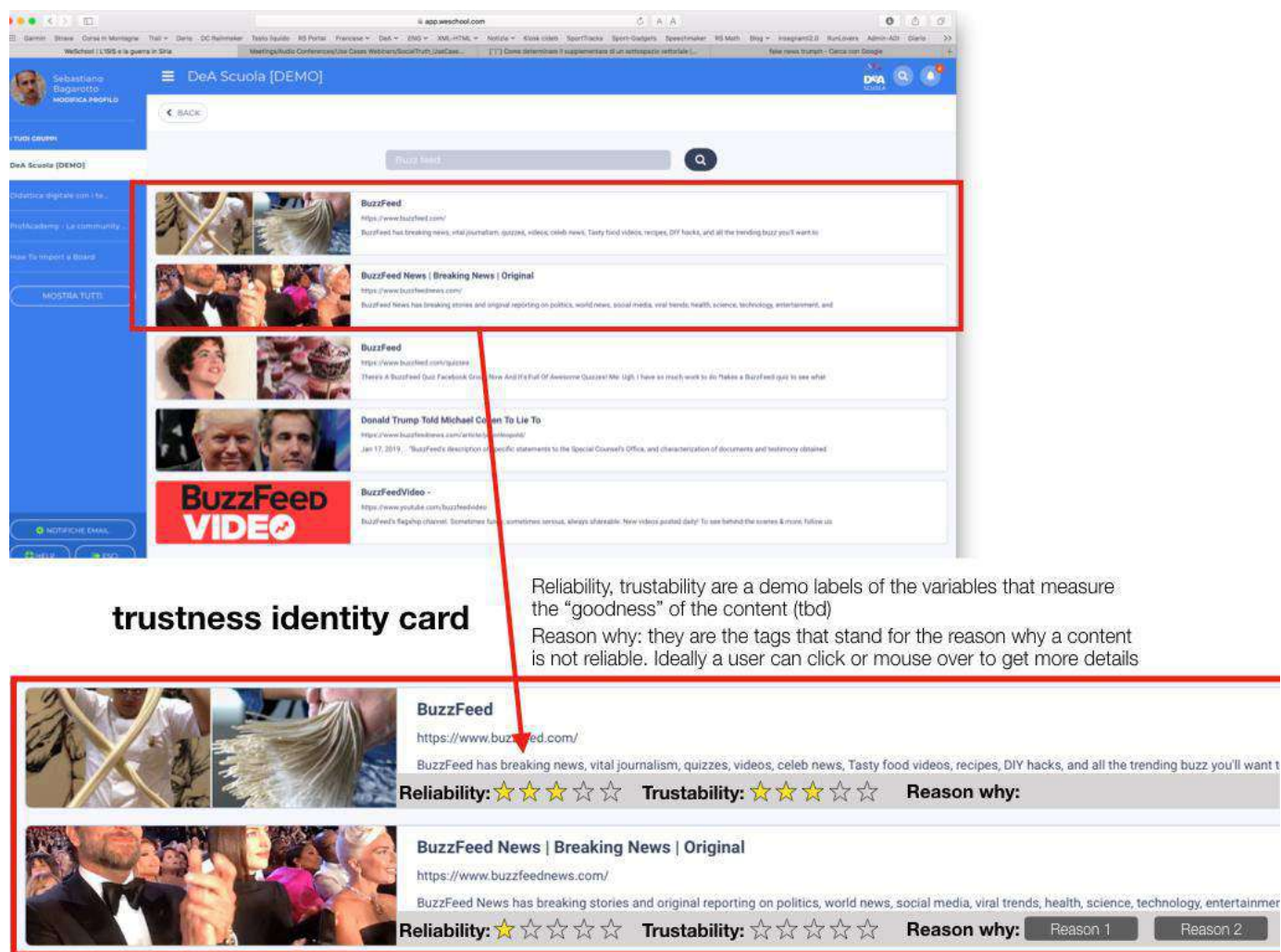
Figure 11. Introducing reliability and trustability in DeA Scuola partner LMS platform.

In these screenshots of a DeA Scuola partner LMS platform it is shown a search tasks user experience , please keep in mind those examples are made with just a functional purposes (no aim to suggest any layout/uix features). In particular:

- Reliability, trustability are demo labels of the variables that measure the "goodness" of the content (tbd)
- Reason why: they are the tags that stand for the reasons why a content is not reliable. Ideally a user can click or mouse over to get more details

Phase 2: Integration with DeA Scuola Web Portal
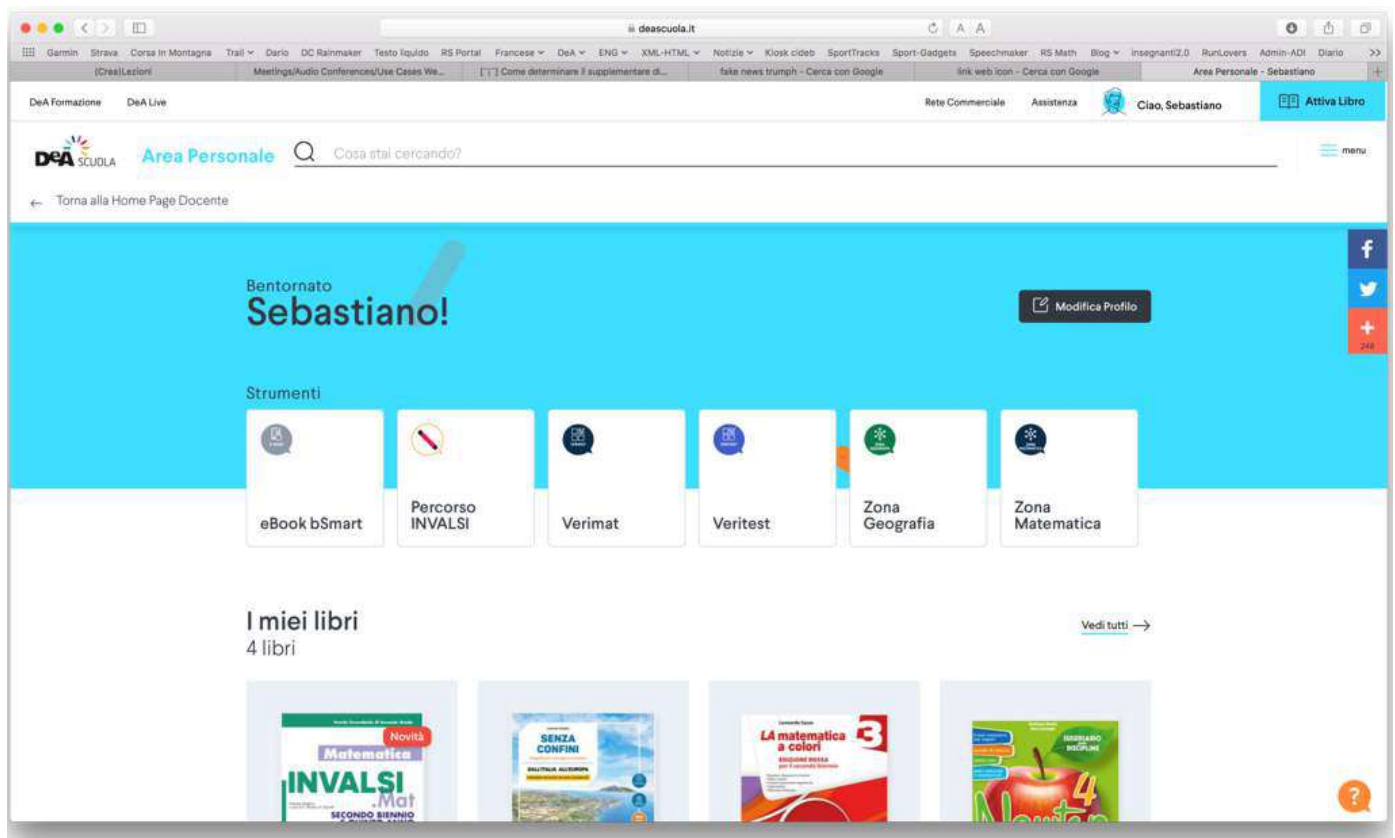
Figure 12. Screenshot from DeA Scuola Web Portal.

The Phase 2 would be the integration with the DeA Scuola web portal. The user experience can be the follow:

- the social truth plugin will we available everywhere on the web portal as a widget

- the user can check the content whenever he wants

Figure 13. Integrating SocialTruth with DeA Scuola Web Portal.

Some overall besides notes:

- Make aware both the teachers and the students of the capabilities of the systems and ways they could integrate into their activities
- No variation of requirements for different school grades
- There should be different validation services for web articles and videos. Mostly web articles are needed on this Use case.
- During the testing phase of the user case we will ask a quality feedback on usability

5. Socio-technical & Human Aspects

This chapter aims at defining SocialTruth's socio-technical methodologies for behaviour modelling and user engagement according to well-established literature and empirical studies in online writing and content sharing in news sites and social networks.

To this end, three key aspects are considered, namely domain knowledge (e.g. ethical and deontological aspects in online writing and content sharing), motivational affordances (e.g. hypertextuality, interactivity, multimediality) and behavioural outcomes (e.g., level of engagement, possible reactions to fake news events).

In addition, social and human aspects in online content creations are also investigated, with the purpose of building engaging and user-centred services. Recommendations for different user categories (journalists, citizens, educators) are proposed, in a multi-perspective approach.

Finally, a common vocabulary for the subject matter is proposed in the last paragraph.

5.1 The public function of the media: ethical aspects in online writing and content sharing

Correctness, sincerity, exhaustivity and transparency all have a central role in social trust and cooperation, and in the democratic participation; a culture based on the truth is a key point not only in the political field and within public institutions, but more in general in human relationships, associations, enterprises and every-day life [29]. As reported by Vosoughi et al., foundational theories of decision-making, cooperation, communication and markets all view some conceptualisation of truth or accuracy as central to the functioning of nearly every human endeavour [30].

If the concept of truth is essential for individuals, for the safeguard of dignity and for the auto determination, it gains even more importance when it comes to the citizen's right to participate directly or indirectly in collective decision-making, especially through free elections, the choice of political representatives and the power to hold elected officials accountable. If citizens are to exploit these rights to the fullest, they must have free access to information that will give them sufficient basis for making enlightened judgements and informed political choices [31].

Such information is diffused, historically and traditionally, by the media. Democracy requires a well-informed, inclusive and pluralistic public sphere, and the media are the creators as well as the editors of this public sphere. Indeed, the production and distribution of reports on events and facts mainly reckons on the work of journalists, who gather information, use literary techniques to make sense of it and disseminate it, by means of print media (newspapers, magazines…), television news channels, radio, digital platforms etc.

Ideally, journalists should support and ward their own work by building a solid reputation and credibility, checking rumours through their network of connections and reporting truthful descriptions of the events they write about. The veracity and completeness of their work products should, in principle, be the main parameters used to assess their journalistic skills. To this purpose, professional journalists adhere to specific ethical rules.

The Declaration of Principles on the Conduct of Journalists (also known as the Bordeaux Declaration) by the International Federation of Journalists [32], provides the standard of professional conduct for journalists engaged in gathering, transmitting, disseminating and commenting on news and information in describing events, summarised by the following principles:

1) respect for truth and for the right of the public to truth is the first duty of the journalist;
2) in pursuance of this duty, the journalist shall at all times defend the principles of freedom in the honest collection and publication of news, and of the right of fair comment and criticism;
3) the journalist shall report only in accordance with facts of which he/she knows the origin; the journalist shall not suppress essential information or falsify documents;
4) the journalist shall use only fair methods to obtain news, photographs and documents;
5) the journalist shall do the utmost to rectify any published information which is found to be harmfully inaccurate;
6) the journalist shall observe professional secrecy regarding the source of information obtained in confidence;

7) the journalist shall be aware of the danger of discrimination being furthered by the media, and shall do the utmost to avoid facilitating such discrimination based on, among other things, race, sex, sexual orientation, language, religion, political or other opinions, and national or social origins;

8) the journalist shall regard as grave professional offences the following:
   - plagiarism;
   - malicious misrepresentation;
   - calumny, slander, libel, unfounded accusations;
   - acceptance of a bribe in any form in consideration of either publication or suppression;

9) journalists worthy of the name shall deem it their duty to observe faithfully the principles stated above; within the general law of each country the journalist shall recognize in professional matters the jurisdiction of colleagues only, to the exclusion of every kind of interference by governments or others.

Given this, not all journalists are equally able or willing to answer to the high expectations of objectivity placed upon them by a hopeful and trusting public, and not all journalism is intended to be entirely objective in the first place. Indeed, the content of the journalistic message and its degree of correspondence to anything resembling an "objective truth" have been reported to be dependent on the following factors [31]:

- professional qualifications and integrity of the journalists;
- adequate working conditions[7];
- subtle or direct economic pressures exercised by the owners of the media or their advertisers;
- technology used to carry out the professional activity;
- defamation laws potentially used to restrict the freedom of the press;
- potential conflicts of interest arising from journalists' closeness to business interests.

In addition to these threads, at present the diffusion of the traditional forms of news dissemination is decreasing, while a major role has been gained by digital media, which can be created, viewed, distributed, modified and preserved on digital electronics devices.

This makes new questions arise on the ethical and deontological aspects of online writing, since the capacity to spread ideas and reach an audience is no longer limited by access to expensive, centralized broadcasting infrastructure: it is limited instead by one's ability to garner and distribute attention [33]. Moreover, those who create and spread ideas and news digitally are to a great majority non-professional and must not undergo any ethical code.

Every user of the internet may now publish news and opinions, being neither privileged nor bound by journalistic rules and professional ethics, but acting simply on his or her right of freedom of expression and the rules and limitations resulting out of that fundamental right, equally guaranteed by national

---

[7] Working conditions for journalists are not mainly about a pleasant environment of plant pots and air conditioning, but about issues that are crucial to building a structure for quality and ethical media. These include having enough people and time to do the job properly and creating an atmosphere in the newsroom which respects professional ethics and experience.

constitutions of member states, the European Convention of Human Rights and the EU Charter of Fundamental Rights [31].

The patterns of the spread of digital content often resemble word-of-mouth, that is people talking about something or telling people about something. Word-of-mouth has gained power in recent years due to the increase of online content sharing: with the arrival of interactive and social media, the patterns, role and impact of word-of-mouth have evolved, and new forms of online communities were shaped. In the social media age word-of-mouth can travel fast. The scale, speed and real-time nature of the social Web and the increased possibilities for people to share and express themselves play a crucial role in this.

The absence of any editorial legacy and the speed of the online content's spread highlight one main issue related to online news searching and content sharing: the level of reliability and veracity of the content.

In this frame it is becoming increasingly difficult, for both professional journalists and regular users of the web, to assess the reliability of news and posts. New social technologies, which facilitate rapid information sharing and large-scale information cascades, can enable the spread of misinformation[29].

This can undermine the public right to have access to quality information, since the distinction between "true" news and misleading, false, fabricated news is becoming increasingly difficult to recognise, and it is when audiences mistake it as a real news that false news is able to play with journalism' legitimacy [34].

Moreover, even professional journalists face the need to find new strategies for making preliminary educated guesses on the trustworthiness of the upcoming breaking news coming from external sources, before starting to fact-check and verify them in the view of quick publication.

The publication and spread of false news can have different origins and reasonings, from harmless satire/parody to misleading commercial advertising, from propaganda to social or economic destabilisation.

In this frame, the SocialTruth project aims at helping to dilute all the non-truthful news spreading on the web by detecting and labelling it, thus promoting objective, true, and verified news.

5.2 Motivational affordances for online content sharing

The online environments allow users to keep up with friends, network with colleagues, and share their personal views and observations with others, but the news use is also becoming a significant component of the overall social network site activity [35].

Social media are becoming central to the way people experience news: networked media technologies are extending the ability of users to create and receive personalized news streams. Editorially, the traditional gatekeeping function of the media is weakened as a significant proportion of news consumers turn to family, friends and acquaintances to alert them to items of interest. Essentially, a person's social circle takes on the role of news editor, deciding whether a story, video or other piece of content is important, interesting or entertaining enough to recommend [36].

The Customer Insight Group of the New York Times [37] investigated the **psychology of sharing** and found that there are five main reasons that led people to share online:

- to bring valuable and entertaining content to others;
- to grow and nourish relationships;
- to define ourselves to others;
- to provide self-fulfilment;
- to get the word out about causes or brands.

Along with these general reasons, one can also use the **Uses and Gratifications (U&G) theory** in attempt to explain what social and psychological needs motivate audiences to select particular media channels and content choices, as well as the subsequent attitudinal and behavioural effects. From a U&G perspective, with regard to news reading, it is assumed that people actively choose among news sources owing to the sources ability to gratify their different needs. As reported by Lee *et al.* [38], the perceived gratifications of online news appear to be entertainment, information search, peer acceptance, relationship maintenance, socialising and self-status seeking.

Apart from these reasons, though, great contribution to the fact that people increasingly share online content is also given by the motivational affordances[8] made available by the platform, website, social network or digital media. The key characteristics which determine this motivational "added value" are reported to be mainly hypertextuality, interactivity and multimediality [39][40].

Hypertextuality is a concept originally envisioned by sociologist and computer programmer Ted Nelson in the 1960s. It first emerged in a 1965 article about his lecture 'Computers, Creativity, and the Nature of

---

[8] According to P. Zhang [110], the term *affordance* refers to the actionable properties between an object and an actor. When perceived, affordance allows actors to take actions that may satisfy certain needs. Motivational affordances comprise the properties of an object that determine whether and how it can support one's motivational needs. When using ICT involves our motivational needs, we feel interest (thus attend and engage). When using ICT satisfies our motivational needs, we feel enjoyment (thus want more). The ultimate goal of designing an ICT for human use is to achieve high motivational affordance so that users would be attracted to it, really want to use it, and cannot live without it.

the Written Word' [41]. Nelson defined the hypertext as "non-sequential writing, a branching text that allows the reader to make choices; it is something that can be best read in front of an interactive screen. [...] With a hypertext we can create new forms of writing that reflect the structure of what we write about; and the readers can choose different paths according to their attitudes and the stream of their thoughts, in a way until now believed to be impossible" [42]. Practically, hypertextuality refers to a text which contains links to other texts (hyperlinks), internal or external to the primary text source.

Multimediality is described as the result of the convergence of diversified media modalities or the sum of different media formats [43]: the online news media, for example, offered little more than written text and pictures at the initial stage, but numerous sites currently offer audio and video content, including streamed and real-time signals. On the other hand, nowadays social networks allow people to share text, images, video content and audio content, also in real time.

Regarding **interactivity**, with the words of Kumiko Aoki [44], in the context of interpersonal communication and computer-mediated communication (CMC), Rafaeli [45] defined interactivity as "an expression of the extent that, in a given series of communication exchanges, any third (or later) transmission (or message) is related to the degree to which previous exchanges referred to even earlier transmissions". Rafaeli refined this definition later, stating that interactivity is "a process-related, variable characteristic of communication settings", and "not a characteristic of the medium" [46].

In this sense, interactivity of a medium refers to a characteristic of communication settings a medium can create that allows users to interact. The categorization proposed identifies two major categories of interactivity on the Web: interacting with the Web and interacting through the Web. Interaction with the Web is enabled by rich media aspects of the Web and personalization of the information offered by a website. This interaction exemplifies the user control aspect of interactivity. Interaction through the Web includes interaction between the publisher of a website content and users of the website and interaction among the users of a website. The former interaction can be coined as publishing and the latter interaction can be referred as community building. Both are interactions for the purpose of communication.

As pointed out by Wang *et al.* [46][47], the users' cognitive involvement and perceived social presence translates the effects of interactivity into gratification, satisfaction with the website or platform, purchase intention and spreading word-of-mouth.

Apart from hypertextuality, multimediality and interactivity, other motivational affordances are emerging in the context of online content sharing, with regard to mobile technologies: portability, availability and locatability [48].

The **portability** affordance is defined as the perception of the physical characteristics such as size and weight, as well as those evaluated through use, such as battery life. Smartphones may now have processing power on par with computers, but portability is what fundamentally differentiates mobile media from desktops. *The very point of the mobile phone's affordances is that the user is able to move in the world* [49].

Regarding **availability**, as reported by Schrock [48], mobile phones were initially thought to enable the potential for "perpetual contact" [50] between individuals and their social networks. Yet, the negotiation of availability is subtler [51]. Strategies of disconnection or partial connection [52] on mobile media become necessary to navigate being constantly connected. In other words, individuals navigate the affordance of availability for specific goals. Availability is, like a radio, "tuned" [53] within a user's comfort zone. Affordances can make communication possible, but it is up to individuals to use these affordances in more or less strategic ways.

Again, as reported by Schrock [48]**, locatability** refers to location-based services such as locative and mobile social networks [54]. These services were often positioned as enabling fundamentally different forms of communication; as Rheingold observed, "knowing our exact geographic location is one form of context awareness in which machines are better than humans" [55]. Location delivers new ways for individuals to form relationships and participate in place-making activities, and geographical coordinates acquire dynamic meaning as a consequence of the constantly changing location-based information that is attached to them. Eric Gordon made an ontological argument that location was able to give greater visibility to "local knowledge produced within the context of located information" [56]. Finally, Jason Farman used location to enable "site-specific storytelling"—narratives produced by small groups to encourage reflection [57].

5.3 Level of engagement with digital media in Europe

Nowadays, the diffusion of the traditional forms of news dissemination is decreasing, while a major role has been gained by digital media, which can be created, viewed, distributed, modified and preserved on digital electronics devices.

In this frame, it is important to recall that:

> In 2018 in Europe, with a total population of 843 million people, there are **647 millions of internet users** (with a penetration of nearly the 80%), 448 millions of active social media users (53% penetration), and 376 millions of active mobile social users (45% penetration), and the annual digital growth rate is increasing [58].
>
> **The sources of news for the European citizens are mainly TV (74%) and online sources (82%), including social media**.

As shown in the following Figures, the Europeans mainly search for news online (including on social media) in all European countries excepting Italy, Germany and France (where TV is the main source of news)[59].
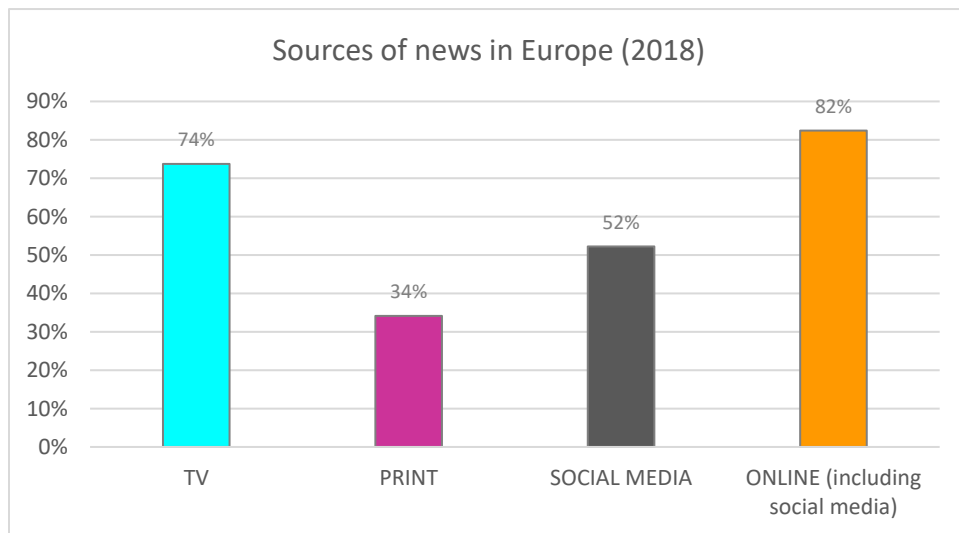
Figure 14. Sources of news in the Europe, graphs extracted from [59].
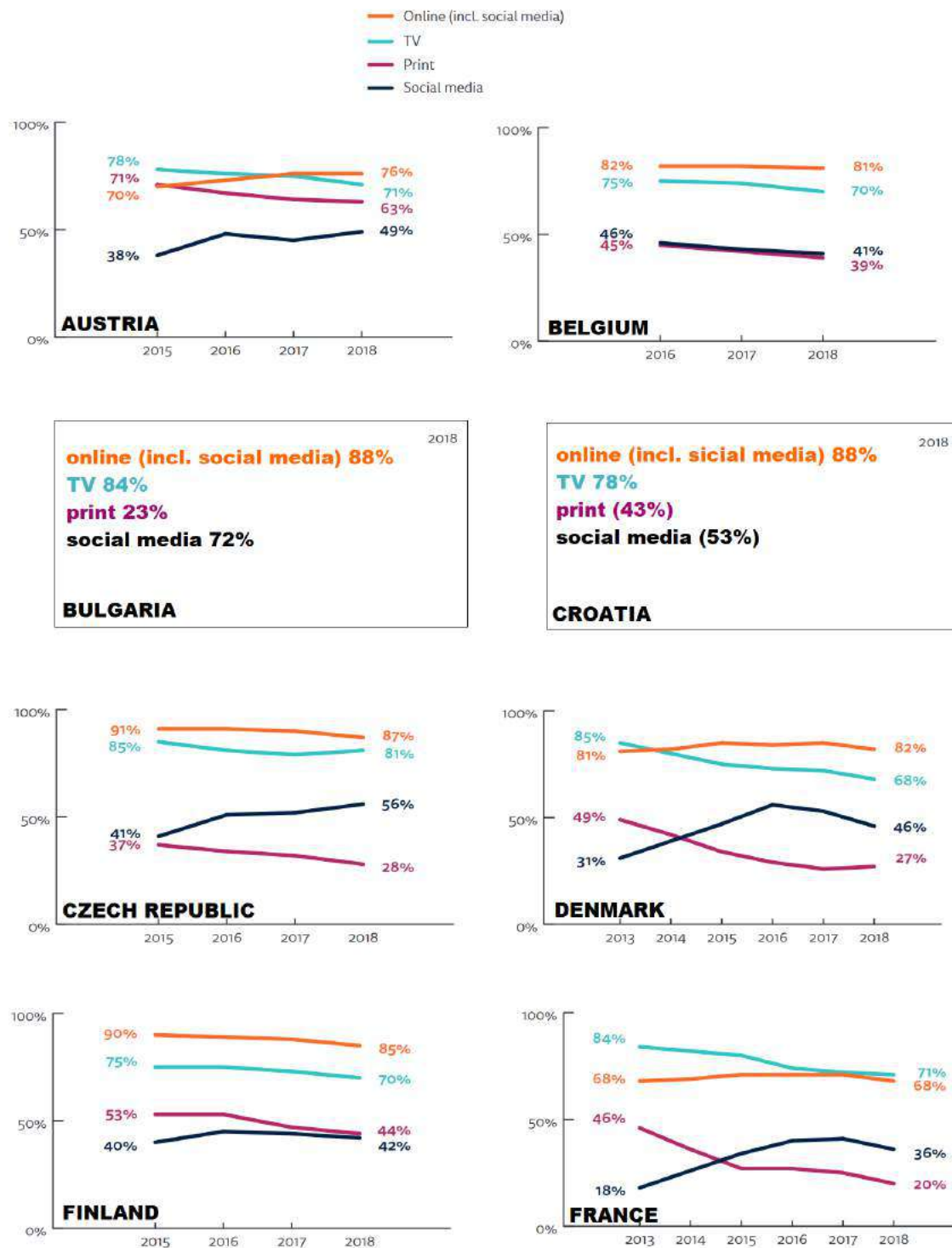
Figure 15. Sources of news in the European countries, graphs extracted from [59].
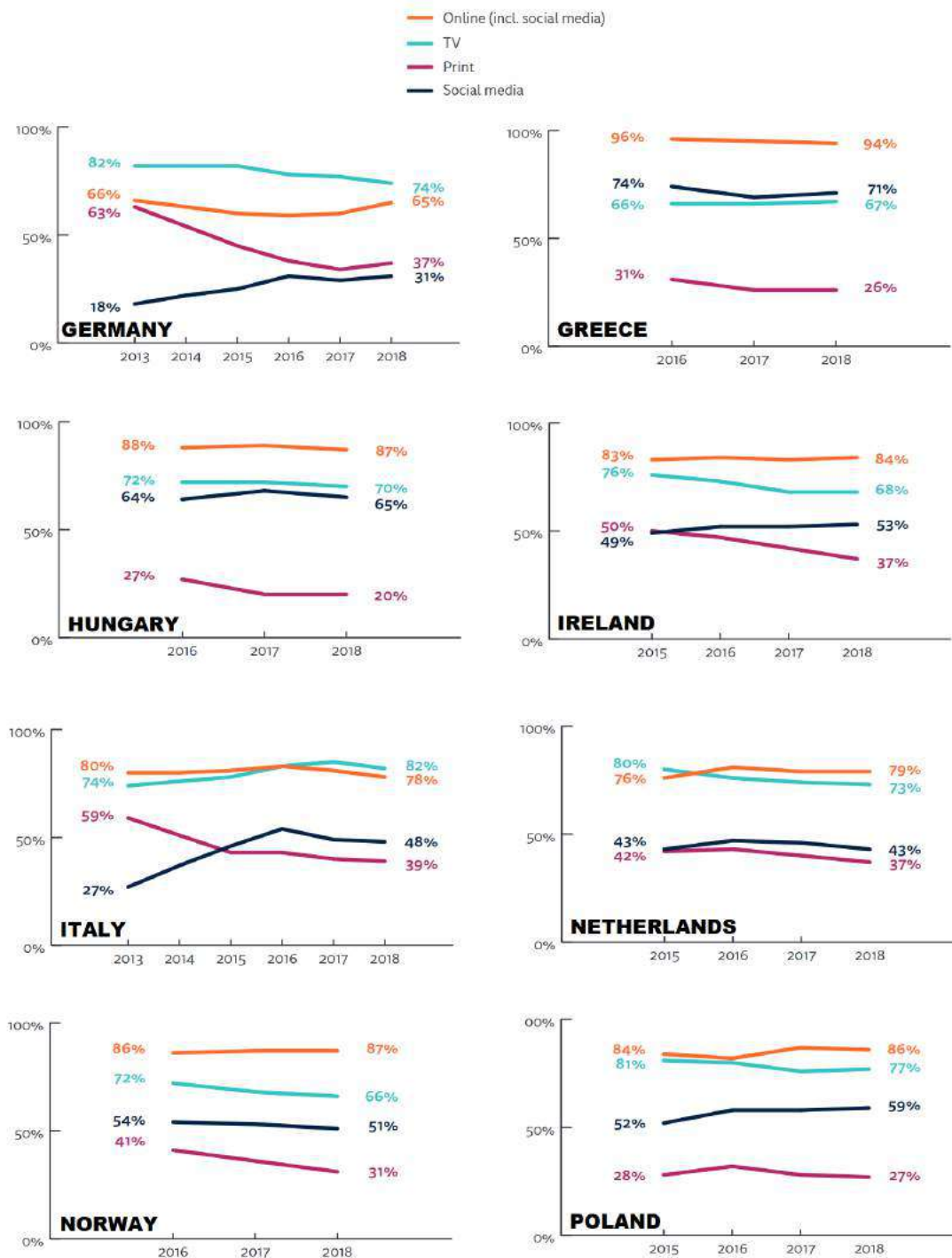
Figure 16. Sources of news in the European countries, graphs extracted from [59].
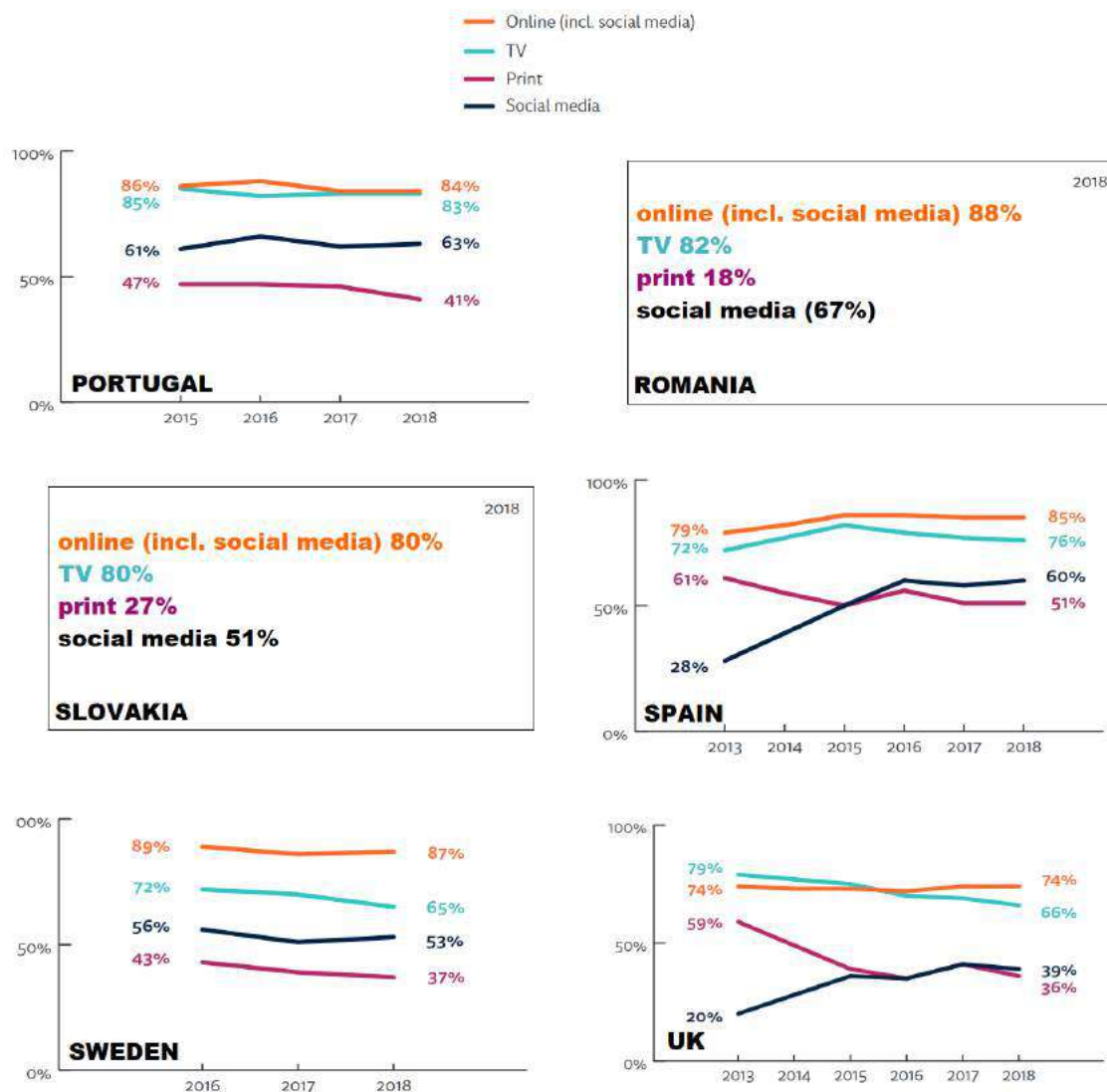
Figure 17. Sources of news in the European countries, graphs extracted from [59].

We also provide summarising graphs:

Percentage of citizens gathering news from TV (2018)
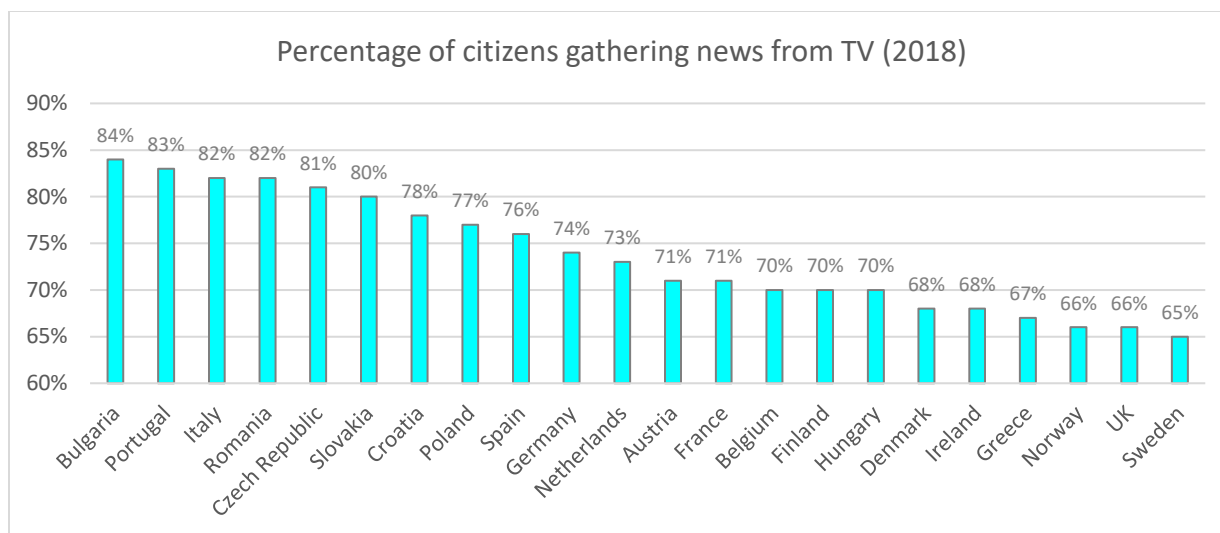
Graph 1 - Percentage of citizens gathering news from TV in 2018, based on findings of [59]

Percentage of citizens gathering news from print media (2018)
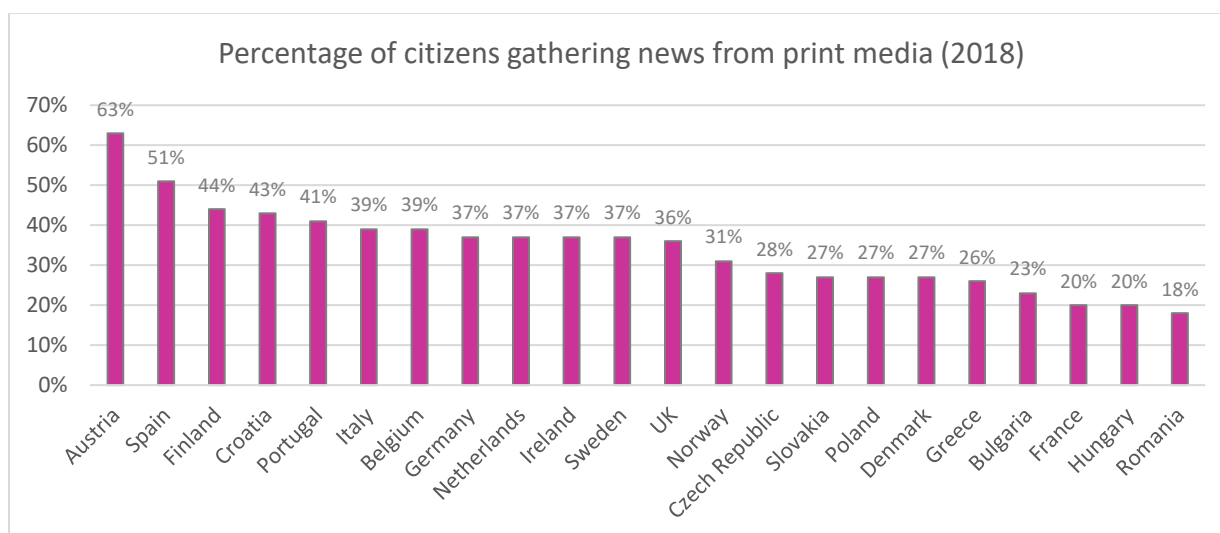
Graph 2 - Percentage of citizens gathering news from print media in 2018, based on findings of [59]

Percentage of citizens gathering news from social media (2018)
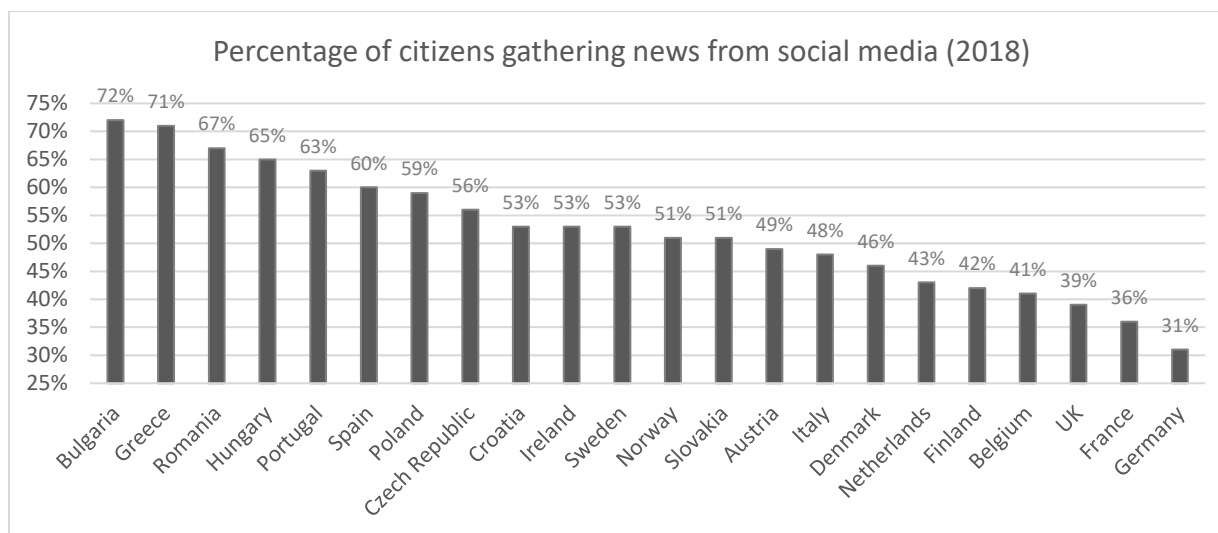
Graph 3 - Percentage of citizens gathering news from social media in 2018, based on findings of [59]

Percentage of citizens gathering news online (incl. social media) (2018)
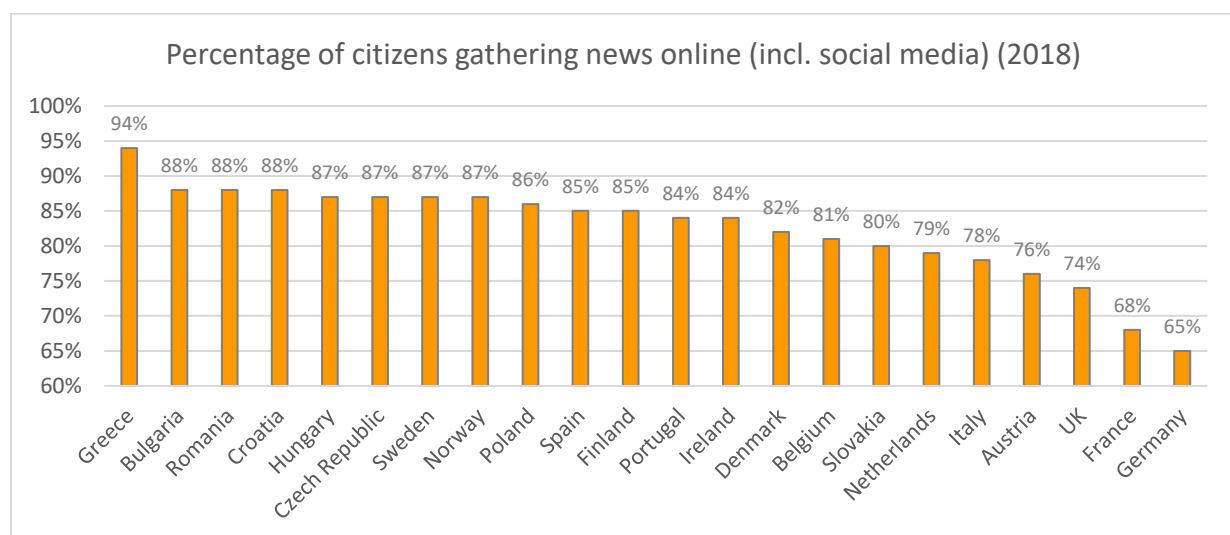
Graph 4 - Percentage of citizens gathering news online, including from social media, in 2018, based on findings of [59]

Given these numbers, it appears that online news searching, and content sharing, has become embedded in people's everyday life, arising new questions on the on-going evolution of the media context and the shaping of a collective truth.

The study also highlight the fact that, on world-wide basis, the vast majority of the respondents (65%) prefer to get news through a "side door", such as search engines, social media, email, mobile alerts and aggregators, rather than going directly to a news website or app. The percentage increases to 73% for people under 35, who prefer using mobile aggregators and social platforms and less direct access. The email, on the other hand, is mostly used by over45s.

5.4 Perception of and reaction to false news

The Flash Eurobarometer 464 [60] provides information regarding the levels of trust in news and information accessed through different channels. We hereby report the main findings:

Respondents are less likely to trust news and information from online sources than from more traditional sources

- *The majority of respondents totally trust or tend to trust news and information they receive through radio (70%), television (66%) and printed media (63%). However, less than half (47%) trust online newspapers and magazines, and lower proportions trust video hosting websites and podcasts (27%) and online social networks and messaging apps (26%).*
- *This pattern is consistent across Member States, with traditional media sources trusted more than online sources in all them.*

Most respondents say they encounter fake news at least once a week

- *More than a third of respondents (37%) say they come across fake news every day or almost every day, and a further 31% say that this happens at least once a week.*
- *In every country, at least half of respondents say they come across fake news at least once a week.*
- *Seven in ten respondents (71%) are totally or somewhat confident that they are able to identify news or information that misrepresents reality or is false (fake news), while 26% are not confident.*
- *A majority of respondents in every country are confident in their ability to identify fake news, ranging from 87% in Denmark to 55% in Spain.*
- *Respondents who use online social networks more regularly, and who come across fake news more frequently, are more confident in their ability to identify them.*

A large majority of respondents think that the existence of fake news is a problem in their country and for democracy in general

- *More than eight in ten respondents (85%) think that the existence of fake news is a problem in their country, at least to some extent. A similar proportion (83%) say that it is a problem for democracy in general.*
- *These views are consistent across Member States, with at least 70% in every country seeing fake news as a problem in their country, and at least 74% saying this in relation to democracy in general.*

Various institutions and media actors are seen as being responsible for stopping the spread of fake news

- *Respondents are most likely to think that journalists should act to stop the spread of fake news (45%), followed by national authorities (39%), press and broadcasting management (36%), citizens themselves (32%), online social*

> *networks (26%), EU institutions (21%) and non-governmental organisations (15%).*

With regard to the online sharing of content and news, in the context of the SocialTruth project it is important to recall that studies highlighted that the volumes, speeds and means of the online sharing greatly depend on the type and truthfulness of the content itself.

A research published on *Nature* by Vosoughi *et al.* [30] reports that specific features can be recognised in the ways people share true and false news online, based on verified true and false news stories distributed on Twitter from 2006 to 2017. The data comprised around 126000 stories tweeted by around 3 million people more than 4.5 million times. The news was classified false or true based on the findings of six independent fact-checking organisations.

We hereby provide the main findings of this study:

> **Falsehood diffuses significantly faster, deeper and more broadly than true news**, especially regarding politics. From their findings it appears that **false news is generally more novel** and that novel information is more likely to be shared, possibly because people feel more compellent about sharing novel news.
>
> Moreover, **the emotional reactions of recipients of false news were found to be mainly surprise and disgust**, whereas the truth inspired sadness, anticipation and trust.
>
> Also, the greater likelihood of people to retweet falsity more than the truth is what drives the spread of false news, despite network and individual factors that favour the truth. **The recommendations about misinformation-containment policies include emphasizing behavioural interventions, like labelling** and incentives to dissuade the spread of misinformation.

These findings are somehow disturbing but helps revealing the dynamics behind the spread of false news online and can act as a base for future research and development of strategies to mitigate such diffusion of falsehoods.

This appears to be essential nowadays, since the diffusion of false news can have real and severe consequences, given that information is used by people, governments and businesses to make decisions and take action [61]. We hereby provide a list of examples:

The Israel-Pakistan nuclear threats
As reported by The New York Times [62]:

In December 2016 the following false story appeared on the site awdnews.com: "Israeli Defense Minister: If Pakistan send ground troops to Syria on any pretext, we will destroy this country with a nuclear attack". The fake story about Israel even misidentified the country's defence minister, attributing quotations to a former minister, Moshe Yaalon, despite Israel's current minister of defence was Avigdor Lieberman.

The Pakistani defence minister, Khawaja Muhammad Asif, wrote a Twitter post directed at Israel after this false report — which the minister apparently believed — that Israel had threatened Pakistan with nuclear weapons: "Israeli def min threatens nuclear retaliation presuming pak role in Syria against Daesh, Israel forgets Pakistan is a Nuclear state too."

The Israeli Defence Ministry responded on Twitter to say the report was fictitious: "The statement attributed to fmr Def Min Yaalon re Pakistan was never said" and "Reports referred to by the Pakistani Def Min are entirely false".

The Pizzagate
As reported by several newspapers [63][64]:

In October 2016 false news about a pizzeria called Comet Ping Pong began appearing on social networks and websites, with the restaurant identified as being the headquarters for a child-trafficking ring led by Hillary Clinton. The articles were soon exposed as false by publications including The New York Times, The Washington Post and the fact-checking website Snopes. But the debunking did not squash the conspiracy theories about Comet Ping Pong — instead, it led to the opposite.

Twitter, Facebook and Instagram were flooded with more attacks against the pizzeria as believers in the child-trafficking conspiracy became more zealous. Within hours of the publication of one of the debunking articles, a post on Twitter by Representative Steven Smith of the 15th District of Georgia — not a real lawmaker and not a real district — warned that what was fake was the information being peddled by the mainstream media. It was retweeted dozens of times.

On YouTube, a step-by-step takedown of the Times article was viewed nearly 250,000 times and passed around on Twitter and Facebook. A surge of new fake articles amplified the original pieces, now linking the child-abuse ring — known as Pizzagate — to a global pedophilia ring reaching Britain.

Edgar M. Welch, a 28-year-old father of two from Salisbury, read the false news about Comet Ping Pong. Apparently concerned, Mr. Welch drove about six hours from his home to the pizzeria to see the situation for himself, according to court documents. Not long after arriving at the pizzeria, the police said, he fired from an assault-like AR-15 rifle. The police arrested him. They found a rifle and a handgun in the restaurant. No one was hurt.

According to the criminal complaint, he told the authorities that he was armed to help rescue children but that he surrendered peacefully after finding no evidence that "children were being

harbored in the restaurant." He was charged with four counts, including felony assault with a deadly weapon and carrying a gun without a license outside a home or business.

The killings in Plateau State
As reported by the BBC [65]:

On 23 June 2018, a series of horrifying images began to circulate on Facebook. One showed a baby with open machete wounds across his head and jaw. Another – viewed more than 11,000 times – showed a man's skull hacked open. There were pictures of homes burnt to the ground, bloodied corpses dumped in mass graves, and children murdered in their beds.

The Facebook users who posted the images claimed they showed a massacre underway in the Gashish district of Plateau State, Nigeria. Fulani Muslims, they said, were killing Christians from the region's Berom ethnic minority.

A massacre did happen in Gashish that weekend. Somewhere between 86 and 238 Berom people were killed between 22 and 24 June, according to estimates made by the police and by local community leaders. But some of the most incendiary images circulating at the time had nothing to do with the violence in Gashish. The image of the baby, which was shared with a call for God to "wipe out the entire generation of the killers of this innocent child", had first appeared on Facebook months earlier. The video in which the man's head was cut open did not even come from Nigeria, it was recorded in Congo-Brazzaville nearly a thousand miles away, in 2012.

The images landed in the Facebook feeds of young Berom men in the city of Jos, hours to the north of the rural district where the massacre was happening. Some of the Facebook posts suggested that the killings were happening right there in Jos, or that the inhabitants of the city were about to be attacked. Few stopped to question the claims, or to check the origin of the graphic pictures that were spreading from phone to phone.

Ali was a potato seller from Jos, a city of around a million people. On 24 June he went to a town called Mangu to meet some customers. It was a journey he'd made hundreds of times. He left shortly after morning prayers and expected to be back in time for dinner with his wives Umma and Amina and his 15 children.

On his way home in a shared taxi, Ali found the road blocked by a wall of burning tyres. A mob of Berom men armed with knives and machetes were interrogating drivers, looking for Fulani Muslims. Ali was dragged from his car along with another male passenger. His charred remains were found three days later near the edge of the Jos-Abuja highway. His body was so badly mutilated his wives refused to see it.

Ali was one of 11 men who were pulled out of their cars and killed on 24 June. Some were set alight. Others were hacked to death with machetes. Days later, their bodies were still being discovered across the city, dumped in ditches, behind houses and along the roadsides. Many were burnt beyond recognition.

Indian child kidnappers
As reported by The New York Times [66]:

In 1018 in India false rumors about child kidnappers have gone viral on WhatsApp. Some of the false messages on the app describe gangs of kidnappers on the prowl. Others include videos showing people driving up and snatching children.

The spread of these false news prompted fearful mobs to kill two dozen innocent people between April and July. One of the first to be killed was a 65-year-old woman named Rukmani. She and four family members were driving to a temple in the southern state of Tamil Nadu in May. As they got close to the temple, the family stopped to ask for directions. A grandmother nearby grew suspicious and called her son, who raised the alarm. The family became nervous and decided to turn back. By the time they got to the next village, a crowd was waiting for them, mistaking them for "child lifters" and assaulted them. They were stripped naked and beaten with iron rods, wooden sticks, bare hands and feet. Videos of the attack were circulated widely online.

When it was over, Rukmani was limp and lifeless. The others were left for dead. Their red sedan was crushed, and their belongings were stolen. The region's top government official said the police had gone around for weeks before the attack warning people not to believe the false kidnapping rumors. But they were no match for WhatsApp. "We could not compete," he said.

Yellow fever vaccines in Brazil
As reported by Wired [67]:

In 2016 in Brazil, perhaps driven by climate change or deforestation or both, the virus of the Yellow Fever began expanding south, even through the winter months, infecting more than 1,500 people and killing nearly 500. Meanwhile, rumors of fatal vaccine reactions, mercury preservatives, and government conspiracies surfaced with alarming speed on the Whatsapp, which is used by 120 million of Brazil's roughly 200 million residents. Brazilian health officials launched a mass campaign to vaccinate 95 percent of residents in the 69 municipalities directly in the disease's path. By the time of the announcement, the fake news cycle was already underway. An audio message from a woman claiming to be a doctor at a well-known research institute began circulating on WhatsApp, warning that the vaccine is dangerous. (The institute denied that the recording came from any of its employees). A few weeks later it was a story linking the death of a university student to the vaccine. (That too proved to be a false report).

Igor Sacramento's mother-in-law messaged him a pair of videos suggesting that the yellow fever vaccine was actually a scam aimed at reducing the world population. A health communication researcher at Fiocruz, one of Brazil's largest scientific institutions, Sacramento recognized a scam when he saw one. And no, it wasn't a global illuminati plot to kill off his countrymen. But he could understand why people would be taken in by it.

"These videos are very sophisticated, with good editing, testimonials from experts, and personal experiences," Sacramento says. It's the same journalistic format people see on TV, so it bears the

shape of truth. And when people share these videos or news stories within their social networks as personal messages, it changes the calculus of trust. "We are transitioning from a society that experienced truth based on facts to a society based on its experience of truth in intimacy, in emotion, in closeness."

People are more likely to believe rumours from family and friends. There's no algorithm mediating the experience. And when that misinformation comes in the form of forwarded texts and videos—which look the same as personal messages in WhatsApp—they're lent another layer of legitimacy. Then you get the network compounding effect; if you're in multiple group chats that all receive the fake news, the repetition makes them more believable still.

5.5 Human-related aspects for journalists' engagement in SocialTruth

As highlighted by the High Level Group (HLG) on Media Pluralism and Freedom, a free and pluralistic media is crucial for European Democracy, as citizens must have free access to information that will give them sufficient basis for making enlightened judgements and informed political choices, and the quality of the news affects the ability of citizens to contribute to and participate in the decision-making processes which concern them [31].

In this frame journalism, and in particular investigative journalism, has special responsibility and importance as a mechanism for protecting democracy and the interests of society. Thus, within the SocialTruth project, it is essential to describe the ethical and deontological aspects encased in journalism, the journalists' perception on the subject matter, and how SocialTruth can effectively help them in their everyday work.

5.5.1   Ethical challenges of digital journalism

Given the description of the ethical and deontological aspects of journalism, one question arises: how are these aspects really perceived by journalists, especially when it comes to online journalism? This matter was investigated in the project "The ethical challenges of digital journalism. A comparative analysis in three European countries: Spain, Italy and Belgium", by J.C. Suarez Villegas [68]. The project aimed at examining the journalists' perception of the ethical challenges associated to different dimensions of the practice of journalism: news production, news dissemination, their deception by the public, and the lines of action to deal with any misdoings. The results presented in the study are based on the outcomes of a 42-questions questionnaire proposed to a sample of 663 journalists from different European countries.

The first question was the following:

Table 1. Questionnaire for journalists, from [68]

What degree of influence do the following factors have in what you consider to be the ethical exercise of online journalism?
ECONOMIC FACTORS
The restructuring or convergence of media to establish a viable economic model
Transparency of the media about their sources of influence
The competition to attract users (click-based profitability)
The precarious work of journalists
PROFESSIONAL FACTORS
Work and information selection routines
Protocols to follow-up and correct errors
Protocols for the use of hyperlinks to cite and contextualise online news
Journalist's individual values
Journalist's professional training
Journalist's experience
Journalist's selected contacts and partners
IDEOLOGICAL FACTORS
The independence of the media from the pressure of political lobbyists and advertisers
The power of media owners and editors to set the agenda
Journalist's loyalty towards the ideological line of the media company they work for
TECHNOLOGICAL FACTORS
The speed of online news production and dissemination
Monitoring of sources' profiles in online forums and social networks
The constant updating of online information
The difficulty to verify and monitor some information online
SOCIAL FACTORS
Citizen participation in the production and contextualisation of information
The questioning of journalistic practices
The credibility and trust of the public in the media

Here is the complete table of results:

Table 2. Results of the questionnaire for journalists, from [68]

What degree of influence do the following factors have in what you consider to be the ethical exercise of journalism?	Very influential (%)	Fairly influential (%)	Little influential (%)	Not influential (%)	NA	Total (%)
The restructuring or convergence of media to establish a viable economic model	27.2	29.6	9.9	2.8	30.5	100.0
Transparency of the media about their sources of influence	27.7	28.2	11.3	1.9	31.0	100.0
The competition to attract users (click-based profitability)	26.8	28.6	11.3	2.3	31.0	100.0
The precarious work of journalists	41.3	15.0	10.3	2.8	30.5	100.0
The independence of the media from the pressure of political lobbyists and advertisers	39.9	20.7	8.0	0.9	30.5	100.0
The power of media owners and editors to set the agenda	33.3	25.8	6.6	3.3	31.0	100.0
Journalist's loyalty towards the ideological line of the media company they work for	22.1	18.8	23.9	3.3	31.9	100.0
The speed of online news production and dissemination	22.1	18.8	23.9	3.3	31.9	100.0
Monitoring of sources' profiles in online forums and social networks	14.6	30.0	20.7	2.8	31.9	100.0
The constant updating of online information	27.2	27.7	12.2	1.4	31.5	100.0
The difficulty to verify and monitor some information online	23.0	31.9	11.7	1.4	31.9	100.0
Work and information selection routines	27.2	30.5	8.0	2.8	31.5	100.0
Protocols to follow-up and correct errors	17.4	29.6	16.4	4.2	32.4	100.0
Protocols for the use of hyperlinks to cite and contextualise online news	15.0	27.2	19.2	6.1	32.4	100.0
Journalist's individual values	43.2	16.9	6.6	2.8	30.5	100.0
Journalist's professional training	39.4	19.2	9.4	1.9	30.0	100.0
Journalist's experience	39.0	17.8	9.4	3.3	30.5	100.0
Journalist's selected contacts and partners	36.6	20.7	10.8	0.9	31.0	100.0
Citizen participation in the production and contextualisation of information	13.6	28.6	22.1	5.2	30.5	100.0
The questioning of journalistic practices	14.6	28.2	22.5	4.2	30.5	100.0
The credibility and trust of the public in the media	26.3	22.5	17.4	3.3	30.5	100.0

The main factors influencing the exercise of ethical journalism were reported to be:

- the degree of independence of the media company for which the journalists work from the pressure of political lobbyists and advertisers (very or fairly influential for 60.6% of respondents);
- journalists' individual values (very or fairly influential for 60.1% of respondents), professional training (58.6%), selected contacts and partners (57.3) and experience (56.8%);
- the power of media owners and editors to set the agenda (very or fairly influential for 59.1% of respondents);
- the work and information selection routines (very or fairly influential for 57.7% of respondents);
- the difficulty to verify and monitor some information online (very or fairly influential for 54.9% of respondents).

The immediacy of online information was considered very or fairly influential by only the 40.9% of the respondents.

The SocialTruth validation tool should be designed to reach good usability[9] by journalists, thus it is essential to correctly assess their needs and requirements, and to properly define the affordances[10] of the solution to be designed.

A highlighted in the previous section, the necessity to quickly publish breaking news appears to be a top requirement, but although the pace of digital media requires speed, this objective should not be achieved by sacrificing the accuracy of information. This means that journalists should be provided with effective tools able to help them speed-up the fact-checking process, but the main goal should remain to correctly assess the reliability and truthfulness of the news content:

> For what concerns SocialTruth, **the difficulty to verify and monitor some information online** and **the immediacy of online information** appear to be the **most promising areas of effectiveness** in helping journalists fulfil ethical aspects while fulfilling their work duties.

5.5.2   Recommendations for journalists' engagement

In "The Elements of Journalism," Bill Kovach and Tom Rosenstiel wrote that journalism's "essence is a discipline of verification" [69].

Brandtzaeg *et al.* [70] researched that when journalists need to verify an information found online, they turn to traditional verification methods, such as phone, email and personal contact with their network. However, they are also willing to the use of content validation tools, as long as functionality and design reflect established journalistic verification procedures and needs.

In a study proposed by Backholm *et al.* [71] practices were investigated, that may contribute to a user-friendly design of a web-based online content validation toolset.

In this study, several factors were recognised to have special importance with regard to a web-based online content validation toolset, based on semi-structured interviews with news journalists from three European countries, shadowing work tasks and interviews with journalism students, and literature ([70][72][73][74]):

---

[9] Usability is the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency, and satisfaction in a specified context of use. Effectiveness is fundamental as it is about achieving the intended goal(s). Efficiency is about the resources (such as time or effort) needed by users to achieve their goals so it can be important. In addition, it is important that users are satisfied with their experience, particularly where users have discretion over whether to use a product and can readily choose some alternative means of achieving their goals.

[10] the properties of an object which allow and suggest its function.

Table 3. journalists' requirements for a web-based online content validation toolset

Journalists' requirements for a web-based online content validation toolset
to be able to **monitor the online content and identify newsworthy information**, such as emerging events or trends, sudden turns during ongoing events, new developments within already identified content areas, and key influencers affecting Social Media opinions about an event;
to be able to **handle verification of varying forms of content, such as text or videos**, and integrate content from several social media platforms into verification processes
to **offer the journalist a high level of control** over, and possibilities to filter, what type of content is tracked or presented, e.g. being able to address both content and source verification in order to answer individual journalistic needs
to be able to identify the **original source of the content** and its **geographical location**, by collecting data from several Social Media platforms, using timestamps, and creating chains of posted and reposted content
to be able to identify the **trustworthiness and quality of a source**, by carrying out comparisons of the type, quality and quantity of the source's previous activity within and across platforms, and by carrying out comparisons based on geolocation information
to have an **easy-to-use design** and visualisation of content to answer to the rapid content production requirement
to have a **user-friendly design** regardless of the technical equipment used
to offer to journalists the possibility to **verify the tool itself**, for example by understanding how central algorithms are constructed, or being able to follow a chain of searches or other tasks carried out automatically by the tool

It must be highlighted that the study explicitly refers to the following major functions required for an effective web-based online content validation toolset:

Table 4. major functions of a web-based online content validation toolset for journalists

Major functions of a web-based online content validation toolset for journalists	
automatic features	• identify the source and its trustworthiness • find new trending content and developments within already identified content areas

manual tasks	• users or team should be able to modify the tool settings according to their needs, commonly mentioned settings include:
	  ○ basic search through search engines
	  ○ sorting functions (fresh, top, trending)
	  ○ filter functions (geolocation, language of content, timeframe limitations, format – e.g. video only)

visualisation of results	• visualisation tools used as a bridge between the automated functions and the users, allowing to provide information focusing on areas of interest relevant to the current assignment, including:
	  ○ summaries of how content has been automatically compared and cross-referenced
	  ○ visual chains of automatically identified steps between reposts and the original source
	  ○ summaries of content listed according to advanced search parameters, such as geolocated posts placed on a map

technical requirements	• tool functionality across screen sizes and equipment • automatic and frequent updates of content feeds to enable rapid inclusion of the latest information • easily accessible content saving function in a format compatible with existing publication formats
team-level	• the possibility to communicate within the toolset with colleagues, in order to avoid unnecessary repetitions of tasks already carried out by other team members.

In the conclusions, the authors also highlight the importance to carefully weigh the tool content against its complexity, and to carry out practical usability tests of tool prototypes with journalists.

5.6 Human-related aspects for citizens' engagement in SocialTruth

As highlighted in the previous paragraphs, half of the European citizens say they come across fake news at least once a week and a large majority think that the existence of fake news is a problem in their country and for democracy in general [60].

Thus, within the SocialTruth project, it is essential to describe the perspective of citizens on the subject matter, with particular attention to how they perceive and trust online media, in order to evaluate how SocialTruth can effectively help them in their everyday life.

5.6.1   Level of trust in the media

The Edelman Trust Barometer Global Report may provide useful insight and information on citizens' level of trust in the media. In the Global Report 2018 [75], with more than 33000 respondents world-wide, it was highlighted that people's general trust in media (including both publishers and platforms: journalists, influencers, search engines, social media platforms, news apps and brands) was quite low in Europe:

Table 5. Trust in the media, selected European countries, 2018, from [75]

Trust in the media	
State	**2018**
France	33%
Germany	42%
Ireland	33%
Italy	45%
Netherlands	55%
Poland	34%
Spain	44%
Sweden	32%
UK	32%

Though, it was reported that the trust in journalism (traditional and online-only) was notably higher, around 59%. For example, in Germany it was reported that the general trust in journalism was 61%, whereas the trust in social media and search engines was only 40%.

One of the main concerns of people regarding news and media was reported to be the uncertainty over the truthfulness of the contents: 63% of the people world-wide agreed that "the average person does not know how to tell good journalism from rumour or falsehoods" and 59% agreed that "it is becoming harder to tell if a piece of news was produced by a respected media organisation".

The report also provides a useful infographic about the world-wide worrying about false information used as a weapon:
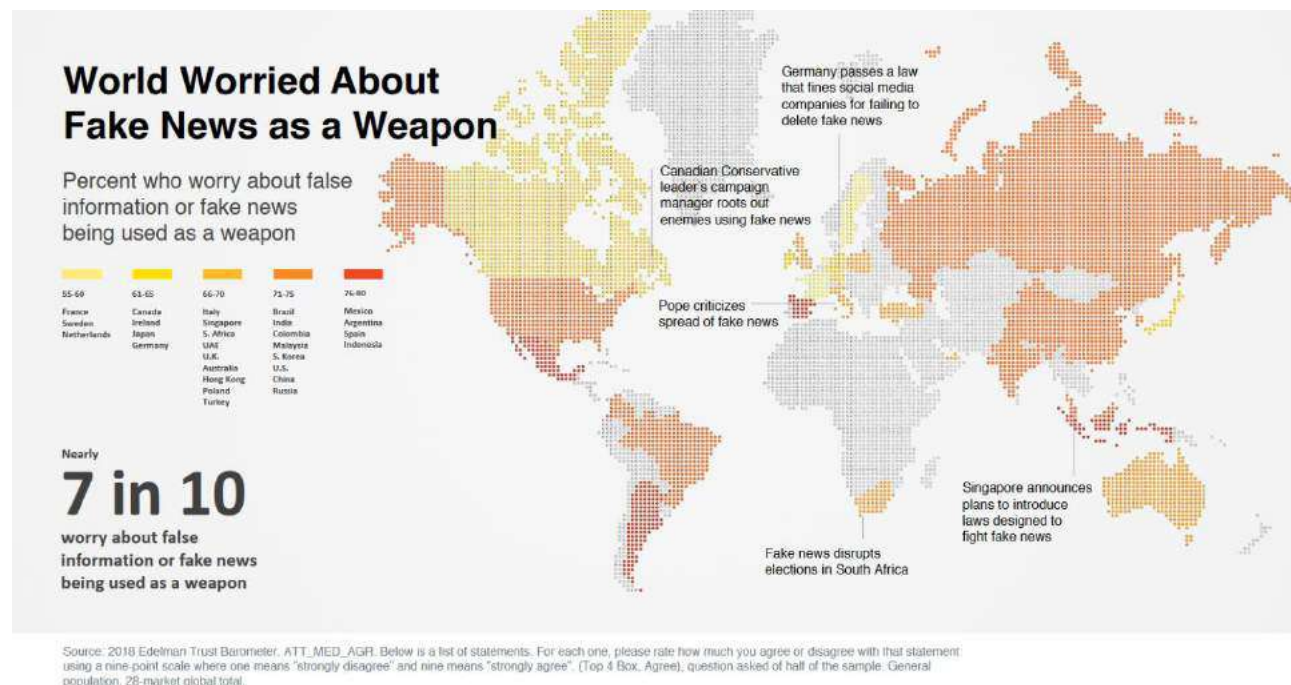
Figure 18. Worrying about false information being used as a weapon, world-wide, 2018, from [75]

In the perception of the people, indeed, media should meet specific expectations regarding their trust-building mandate, mainly to guard the information quality, to educate people on important issues and to inform good life decision.

In the Edelman Trust Barometer Global Report 2019 [76], with more than 33000 respondents world-wide, it was highlighted that people's general trust in media increased with respect to 2018:

Table 6. Trust in the media, selected European countries, 2018-2019, from [76]

Trust in the media		
State	**2018**	**2019**
France	33%	36%
Germany	42%	44%
Ireland	33%	35%
Italy	45%	45%
Netherlands	55%	56%
Poland	34%	-
Spain	44%	36%
Sweden	32%	-
UK	32%	37%

Also, a summarisation of the citizen's word-wide trust in the media from 2012 to 2019 was provided, that highlights the fact that the traditional media are the most trusted (65%), together with search engines, while social media are the least trusted (43%):

TRADITIONAL MEDIA, SEARCH MOST TRUSTED

Percent trust in each source for general news and information
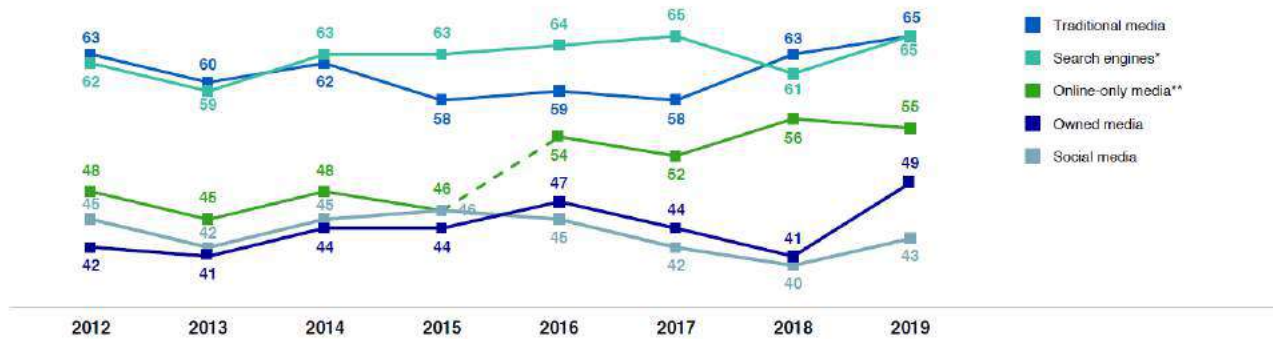
2019 Edelman Trust Barometer. COM_MCL. When looking for general news and information, how much would you trust each type of source for general news and information? Please use a nine-point scale where one means that you "do not trust it at all" and nine means that you "trust it a great deal." 9-point scale; top 4 box, trust. Question asked of half of the sample. General population, 23-market average.

*From 2012-2015, "Online Search Engines" were included as a media type. In 2016, this was changed to "Search Engines."
**From 2012-2015, "Hybrid Media" was included as a media type. In 2016, this was changed to "Online-Only media."

Figure 19. Comparison of trust in different types of media, world-wide, 2012-2019, from [75].

In addition, the overall world-wide news engagement increased in 2019:

Figure 20. New engagement, world-wide, 2018-2019, from [54].

The demographic profile of each news engagement segment was also provided in the report, based on age, gender, education and income:
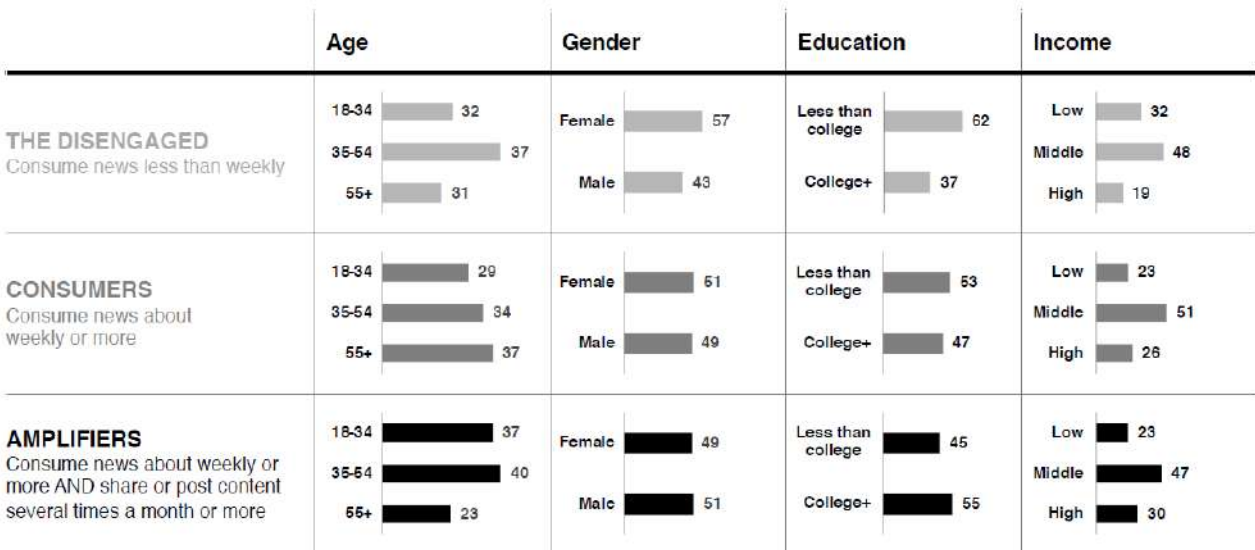
Figure 21. Demographic profile of each news engagement segment, worldwide, 2019, from [75].

5.6.2   Recommendations for citizens' engagement

The spreading of false news online is often seen as an untouchable phenomenon that cannot be corrected, but it is not, and researches who investigated this topic provided useful suggestions about strategies and solution that can address the problem without affecting the society's freedom of speech and fundamental rights.

These solutions appear to be especially effective when implemented in coordination with one another and include [77]: banners and flagging, community reporting and verification, education, layout and formatting, technical verification, third party verification and visibility manipulation.

In addition to this, social cognitive theory(SCT)-guided Intervention has been proposed as a way for deterring the spread of misinformation on social network sites [78]: Chen *et al.* proved that presenting users with a message that highlighted the negative consequences of misinformation effectively reduced the total number of shares for misleading content.

In the light of the motivational affordances for online content sharing (previously described), it is possible to expand solutions found in literature into more specific recommendations for the user engagement to a content validation tool[11], that may be a browser add-on or plug-in, a digital companion, an online news media platform or a search engine itself:

Table 7. Recommendations for a content validation tool designed for citizens

Recommendations for a content validation tool designed for citizens	
Visual indication of the truthfulness of the content	• the online content should be accompanied by a sign, colour, icon, label, banner, or other forms of visual communication instruments that can intuitively tell a message regarding the truthfulness of the content • the red colour seems to be the most appropriate and effective to highlight false content [78] • the use of differentiated icons or labels for different types of mistrusted content (parody, impostor, altered images or videos…) should be investigated • the possibility to also visualise the grade of the verification process ("rating in progress", "preliminary results", "definitive rating"…) should be investigated • the graphic choices should be tuned according to literature, but it would be better to organise questionnaires, surveys, interviews and/or workshops with the users
layout and formatting	• priority should be given to the significant parts of the content regarding truthfulness, for example if a news is published by an impostor website (with URL name very similar to a legitimate news agency) the impostor URL should be made more visible
hypertextuality	• the user should be able to investigate more deeply the truthfulness of the content and understand how the level of rating was evaluated • for example, the graphical warning may be accompanied by an "info" button linking to the visualisation of the chain of searches or other tasks carried out automatically by the tool, or to related news from news agencies or third party verifiers
community reporting, interactivity	• the users may actively contribute to reporting false news and draw attention to it

---

[11] These recommendations are specific to the domain of online content verification, but for user-engagement it is also important to satisfy the general elements affecting user experience and user engagement. Garret *et al.* performed a comprehensive literature review regarding user engagement and the specific elements used in effective website and mobile application design, reasonably valid also for browsers' and social networks' add-ons and digital companions. The design elements most often discussed in relation to user engagement in the reviewed studies were: navigation, graphical representation, organisation, content utility, purpose, simplicity and readability.

	• the possibility to include in interactive functions in the tool and a "share button" for the level of truthfulness of the content should be investigated
visibility manipulation	• in the case of search engines, false content should not be removed from the search results but should be penalised in rankings
education	• the tool may include messages, tutorials, guidelines or serious games designed to educate users on the dangerous consequences of false news sharing and on the best practices to recognise untruthful content

Recommendations stemming from the socio-economical considerations are will be elaborated in the Digital Companion design in D2.3 "Refined Distributed System Architecture" and D4.3 "SocialTruth Digital Companion".

5.7 Human-related aspects in the field of education

This paragraph aims at exploring how SocialTruth can be effective in supporting teachers and students in the field of education.

The aspects investigated are mainly two: the first related to providing support to teachers in the development of web-supported lessons, and the second related to the education of students about online information literacy.

5.7.1 Supporting teachers in the development of web-supported lessons

Wired classrooms and web-supported lessons are becoming common in the European schools and the pedagogical tools available to teachers and students are changing more rapidly than ever before. Not all online tools, or teachers' and students' use of them, are alike. Rather, they represent a broad range of purposes, with different affordances and limitations for literacy practices and teaching [80].

Among the online tools available for teachers, some are designed to simplify lesson planning, others to keep the classroom running smoothly, others to enhance the students' engagement or involve families in learning. Some of them allow the teacher to incorporate online content in their lessons.

Studies reported [81] that teachers find that digital technologies have helped them in teaching their middle school and high school students in many ways. Asked about the impact of the internet and digital tools in their role as educators, teachers said the following bout the overall impact on their teaching and their classroom work:

- internet has a major impact on their ability to access content, resources, and materials for their teaching (92%), to share ideas with other teachers (69%), to interact with parents (67%), and on enabling their interaction with students (57%);
- the internet and other digital tools have added new demands to their lives, by increasing the range of content and skills about which they must be knowledgeable (75%) and by requiring more work on their part to be an effective teacher (41%).

In this context, SocialTruth may be effective in supporting teachers in the exploitation of the afore-mentioned digital teaching tools by providing means of online content verification, to assure that students are engaged with appropriate material, but also for example to avoid copyright issues.

5.7.2 Supporting students about online information literacy

Conventional text-bound teaching in the content areas belies how contemporary youths locate and use information that has relevance for them [82].

As described by Terry Heick [83], reading, writing and arithmetic are the cornerstones of education, since they equip learners with the skills necessary to interact and participate in the world and lay the groundwork for a set of more complex skills: critical thinking, collaboration, nonlinear thinking and others. Though, as more and more of our daily lives rely on the Web - whether it's getting news, connecting with friends and family, or learning about things that interest us - it's time to recognize the need for another

educational building block: web literacy, that is the ability to read, write and participate on the Web. Web literacy touches on a variety of competencies - from composing and coding to understanding why privacy matters online - but it allows students to do one essential thing: meaningfully engage on the Internet.

This issue is connected to the purposes of the SocialTruth project as the spread of false news is also a matter on online information literacy[12]. This is especially the case with people who are going online for the first time. For those individuals, it is hard to distinguish false from real news, and they need to learn how to evaluate news sources, not accept at face value everything they see on social media or digital news sites. Helping people become better consumers of online information is crucial as the world moves towards digital immersion, and education is especially important for young people [84].

Do students know the difference between what's true, biased, and a complete lie? It might be hard to tell the difference at first, but it is becoming the responsibility of schools to help students discern between fact and fiction [85].

Indeed, many initiatives are being brought on all around the world on this subject, for example the Stanford History Education Group prepared a document called "Evaluating Information: The Cornerstone of Civic Online Reasoning" [86] proposing strategies, exercises and sample students work for teaching the ability to judge the credibility of information that floods young people's smartphones, tablets, and computers.

Also, the News Literacy Project [87] is an education program created with input from real journalists (from *New York Times*, the *Wall Street Journal*, and *Buzzfeed News)* that aims at helping young people distinguish real news from fake news, recognise the earmarks of quality journalism and credible information, and judge if articles are accurate and appropriately sourced. It teaches kids to categorize information, make and critique news judgments, detect and dissect viral rumours, interpret and apply their right of speech, and recognize confirmation bias[13] [88].

In this context, SocialTruth may be effective in supporting teachers in their educational role and accompanying students in their everyday online life, by providing evidence about the credibility or untruthfulness of a source, an image, a video, a piece of news, or any other content found online.

---

[12] Online information literacy is part of information literacy, referring to being able to identify, locate, evaluate and effectively use information**.**

[13] The confirmation bias is the tendency to search for or interpret information in a way that confirms one's preconceptions.

5.8 Definitions for a common vocabulary

In order to prevent misunderstanding and confusion, it appears essential to build a common vocabulary regarding the online contents and news from a veracity perspective, and especially to define and categorise different types of news based on their level of veracity.

To this purpose, this appendix is provided to properly define the key-terms of the subject matter: information, news, false news, and their features.

Transversally to this, there is the so-called *fake news*. This locution has a variety of differentiated meanings, depending on the context, and is hard to define. Therefore, although the SocialTruth's Consortium has decided to avoid using it, preferring to use the above mentioned more specific words to describe disinformation and misinformation, a specific section is provided for describing the different definitions of the locution, for completeness.

5.8.1   What is *information*?

Regarding the definition of *news*, a detailed description was provided by Karlova *et al.* in 2013 [61]:

> *Information scientists have long debated the nature of information: what it is, where it comes from, the kinds of actions it affords humans, etc. From its earliest stages, information science has sought to define information, beginning with Shannon and Weaver's idea that information can be quantified as bits of a signal transmitted between one sender and one receiver* [89]*. Later, Taylor* [90] *argued for the need to study 'the conscious within-brain description of the [information] need'. Belkin and Robertson* [90] *notably advocated for a view of information as 'that which is capable of transforming structure' of information inside a user's mind. Dervin and Nilan* [92] *contended that information ought to be viewed 'as something constructed by human beings'.*

> *Tuominen and Savolainen* [93] *articulated a **social constructionist view of the nature of information as a 'communicative construct which is produced in a social context'.** They focused on discursive action as the means by which people construct information. A constructionist view of information is useful when discussing misinformation and disinformation because it emphasizes social context and conversations among people as ways of determining what information is and what can be informative.*

The definition provided by Tuominen and Savolainen seems to be particularly useful, since it connects information with misinformation and disinformation, whose definitions are hereby provided, based on [94]:

> **misinformation** refers to the inadvertent sharing of false or misleading information.
>
> **disinformation** refers to the deliberate creation and sharing of information known to be false or misleading.

It is important to notice that misinformation and disinformation can still be informative, even if affected by concealment, ambivalence, inaccuracy, vagueness, ambiguity, distortion or falsification.

5.8.2   What is *news*?

Regarding the definition of *news*, a very comprehensive and exhaustive description was provided by Edson C. Tandoc *et al.* in 2018 [34]:

> *"News has been defined in a number of ways, ranging from being an account of a recent, interesting, and significant event* [95]*, an account of events that significantly affect people* [96]*, to a dramatic account of something novel or deviant* [97]*. News is often seen as an output of journalism, a field expected to provide* **"independent, reliable, accurate, and comprehensive information"** [98]*. Since the "primary purpose of journalism is to provide citizens with the information they need to be free and self-governing," journalism is expected to report, above all things, the truth. A central element in the professional definition of journalism is adherence to particular standards, such as being objective and accurate. Along with the responsibility of the profession comes power. Thus, journalists have occupied an influential position in society, namely one that can amplify and confer legitimacy to what it reports* [99]*.*
>
> *At the same time, news is socially constructed, and journalists often exercise subjective judgment on which bits of information to include and which to exclude* [100][101]*. Thus, news is vulnerable not only to journalists' own preferences* [102]*, but also to external forces, such as the government, audiences, and advertisers* [103]*. News is also a unique commodity, for while it is sold to audiences, news audiences are subsequently sold to advertisers* [104]*, making it vulnerable to market forces.* ***Still, news is expected to include accurate and real information.*** *A landmark survey of American journalists, for example, differentiated* ***journalists as those involved in the production of reality****, instead of symbolic media content* [105]*. Journalists "make the news" but it does not mean they fake it* [106]*."*

5.8.3   What is *false news*?

As a starting point we can assume that, based on the genuineness of the information transmitted through a news, the following dichotomy exists:

> A (true)[14] **news** is a news whose content is genuine, can be successfully fact-checked, refers to an accurate account of a real event and is accompanied by true contextual information.
>
> A **false news** is a news whose content has no basis in fact, is inauthentic, counterfeit and, by any definition, is a lie.

---

[14] Here the word *true* is indicated in brackets since news is supposedly based on truth, which would make the term *false news* an oxymoron.

Between these two opposites, though, lies a number of shades of greys, and we can recognise a list of parameters which can be used to evaluate the grade of veracity of a news:

Table 8. List of parameters which can be used to evaluate the grade of veracity of a news

PARAMETER	DESCRIPTION
CONTENT	events, statements, ideas and visuals contained in the piece of writing
CONTEXT	circumstances that form the setting for an event, statement, idea, image or video and in terms of which it can be fully understood
CONNECTION	relationship in which different parts of the content are linked or associated with the others (e.g. images supporting text, headlines summarising text…)
SOURCE	entity (someone or something) that supplies information
INTENTION	aim or plan

To our purposes, the categorisation proposed by Claire Wardle [94] appears to be useful, merged with the typology proposed by Tandoc *et.al.* [34]. Based on the genuineness of the content, context, connections, source and on the potential/intention to cause harm, one can distinguish the following types of non-truthful news:

- news focused on humour or exaggeration with no intention to cause harm, such as **satire** (based on actual events) or **parody** (based on non-factual information);
- news containing **false connection**, with headlines, visuals or captions not supporting the content;
- news with **misleading content**, where an issue or an individual are framed with misleading use of information;
- news framed in a **false context**, where genuine content is shared with false contextual information;
- news with **impostor content**, where genuine sources are impersonated with false, made-up sources, and sophistication is used as a mental heuristic to judge its credibility
- news with **manipulated content**, where genuine information or imagery is manipulated to deceive or to create a false narrative

In addition to this, based on the intention of the publisher, we can also distinguish the following words [94][107]:

hoax refers to a false news deliberately fabricated to masquerade as truth and shared with harmful intent.

propaganda refers to news stories which are created by a political entity to influence public perceptions; it is often based on facts, but promotes a particular side or perspective, with the intention to persuade, rather than to inform.

clickbait news refers to online content that uses sensationalized headlines, images and/or captions to tempt a person into clicking to view the original web article at the source publication, typically generating advertisement-based revenue for the publication.

5.8.4   What is *fake news*?

In general, *fake news* or junk news is reported to be a type of yellow journalism or propaganda that consists of deliberate disinformation or hoaxes [33]. Other authors report that a fake news is not (necessarily) a false news. This can be news presented with a hidden agenda, some of which is put forward at the expense of others to support a purpose of the one who broadcasts it. A fake news may simply be misleading, may be presented for the purpose of manipulating, and therefore it may cause public harm [108]. The term is also at times used to cast doubt upon legitimate news from an opposing political standpoint, a tactic known as the lying press [109].

6. Security, Privacy and Ethical considerations

The aspects relating to security, data protection and ethics will be kept in full consideration and all aspects of the activity carried out will be assessed in the perspective of the current provisions and guidelines on the matter.

For the analysis of the critical issues related to ethics, the checklist present in the guidelines for ethics self assessment will be used first (v6.1 feb 4th 2019).

Since use cases will involve human beings in research *([...]any research involving work with humans beings, regardless of its nature or topic. Examples: collection of biological samples, personal data, medical interventions, interviews, observations, tracking or the secondary use of information provided for other purposes, e.g. other research projects, officially collected information, social media sites, etc.)*, the required measures will be taken:

- informed consent form
- information sheet

Research will involve volunteers for social or human sciences research, and it will be necessary to provide:

- details of the recruitment, inclusion and exclusion criteria and informed consent procedures
- copies of ethics approvals

Although from a preliminary analysis is not considered to be the case, it should be noted that if persons unable to give informed consent (including children/minors) and/or vulnerable individuals or groups will be involved in the research activity, further documentation will be needed: details of the age range, assent procedures and parental consent, steps to ensure the welfare of the child or other minor and justification is there for involving minors.

ETHICAL ISSUE RELATED TO PERSONAL DATA PROCESSING

The processing of personal data carried out in the context of use cases will be evaluated in concrete terms: the guidelines are clear to include in this activity processing of personal data, regardless of the method used (e.g. interviews, questionnaires, direct online retrieval, etc.) and the concept that 'Personal data' means information relating to an identified or identifiable natural person. It is also necessary to underline that completely anonymised data does not fall under the data privacy rules.

It should be noted that in case the research activity involves processing of personal data, the following information have to be provided:

1) Details of the technical and organisational measures to safeguard the rights of the research participants;

2) Details of the informed consent procedures;

3) Details of the security measures to prevent unauthorised access to personal data;

4) explain how the processed data are relevant and limited to the purposes of the project;

5) Details of the anonymisation /pseudonymisation techniques (if relevant);

6) Justification of why research data will not be anonymised/ pseudonymised (if relevant);

7) Details of the data transfers (type of data transferred and country to which it is transferred – for both EU and non-EU countries).

Moreover, informed consent forms and information sheets will have to be provided.

Other Risks: it is appropriate to point out the risk deriving from the exit of the United Kingdom from the European Union, which will have direct consequences also on data processing policies, as one of the project partners has its headquarters in London, England.

As regards the aspects related to data security, guidelines and policies will be established to ensure the correct preservation of data and a policy of access to them will be implemented.

7. Requirements definition framework

7.1 Requirements formalization

One of the major goals of this document is to formalize the functional and non-functional requirements for the SocialTruth solution. The set of requirements will provide input to deliverable D2.2 focused on the SocialTruth architecture.

The process of requirements formalization was comprised of three stages:

- Requirements elicitation, based on available sources and close cooperation with end-users,
- Requirements analysis, including their categorization and synthesis (if they are relevant for more than one use case),
- Requirements assessment resulting in assigned priorities based on the common MoSCoW methodology (M – must, S – should, C – could, W – would).

The sources used to create the initial list of requirements during the process of their elicitation, are:

- SocialTruth DoA, where the main functionalities of the SocialTruth solution have been described and general user needs analysed,
- Discussions, webinars, teleconferences carried out with the strong involvement of the project end-users (use-case actors) during the first months of the project,
- Detailed use-cases description provided in section 4 of the current document, and results of their analysis,
- End-user workshop session during the Bydgoszcz Plenary Meeting, 26-27 March 2019.
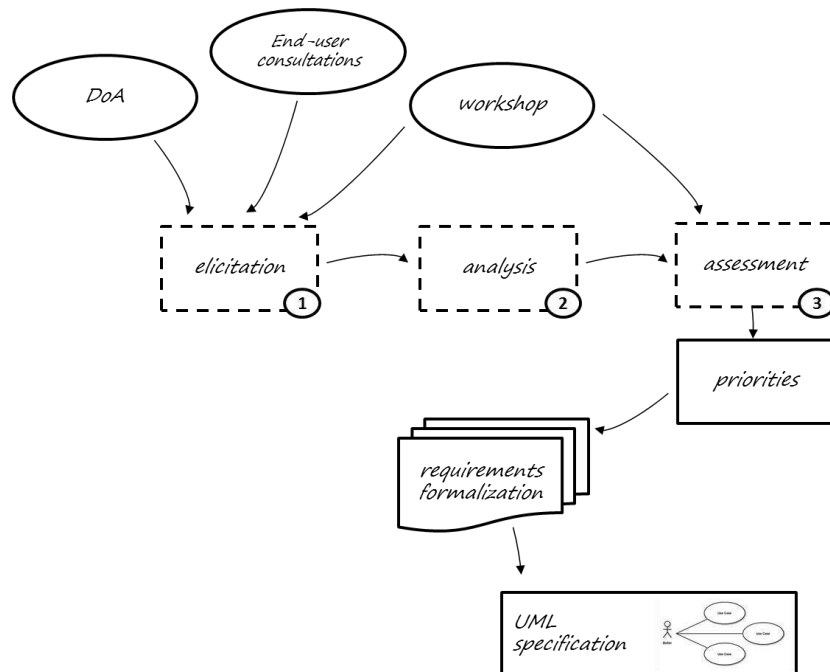
One of the issues touched upon during the March Social Truth plenary meeting in Bydgoszcz, Poland, related to the way the consortium should handle system requirements. The general discussion had two possible, but mutually exclusive conclusions: either the requirements are going to be handled on a per use-case basis, or they should be formed in a more general way, one that would encompass more possible uses. Both options have their pros and cons, but for the most part, the collective, comprehensive approach won the partners hearts. Thus, it was agreed that requirements should be handled in a broader, more general sense.

During the mentioned workshop session, the list of user requirement was assessed by the end-user representatives and other consortium partners to prioritize them. As a result, prioritized user needs with indication of priority scores ranging from 1 to 5 (where 5 is "must have" priority) were collected in the following table:

No	Requirements	Average score	Average score (only end-users)
1	Filter out source/content by credibility index	4,38	4,00
2	Flag untrusted information/source	4,38	5,00
3	Integrate functionalities into the core editorial software dedicated to journalists	3,63	2,00
4	Provide access to the functionalities for a standard Internet user	4,25	4,00
5	Ease requesting the verification of a digital article or social media post	4,38	4,67
6	Provide the web frontend access to the functionalities for a professional users	4,00	4,00
7	Allow the user to launch the verification process	3,50	4,00
8	Ask the user to provide feedback on automated verification	3,88	3,67
9	Ask the professionals and individuals to annotate (justify or explain) reasons of the verification feedback	3,86	3,50
10	Allow user to specify fraudulent criteria for a search results	3,38	2,67
11	Inform the user about possibility of fake news in the search results	3,63	4,33
12	Re-rank the search result based on the fraudulent criteria	4,00	4,00
13	Integrate functionalities into the platform for lesson design	3,75	4,00
14	Near real-time external educational materials verification	3,63	3,67
15	Display credibility index of the educational materials immediately (e.g. when typing web address)	3,63	4,33
16	Provide user (teacher) justification (give reason) why the content is untrusted	4,00	4,33
17	Allow for semantic text searching	3,75	4,00
18	Allow for automated categorization of content based on text analytics	3,29	3,67

19	Allow for identification of a content author's writeprint based on stylometric analysis (at least based on usage of short words, conjunctions, vocabulary richness and complexity, lexical differentiation)	3,14	3,00
20	Allow for mapping authors writing style to another styles from a historical database	3,14	2,67
21	Allow for mapping authors identity to another identities (author names or nicknames) from a historical database based on the stylometric analysis	3,38	2,67
22	Allow for creating new ontologies and taxonomies automatically based on the machine learning techniques	3,29	3,67
23	Allow for identification of writer's emotions based on the text content (deeper granularity than a standard positive/neutral/negative evaluation)	3,29	2,67
24	Allow for identification of writer's intention (estimate a bias)	3,86	3,67
25	Allow for assessment of authenticity of image posted online	4,63	4,33
26	Store analysis performed by different users of the system in the blockchain	3,86	3,33
27	Allow the users to access and check the historical evaluations performed by the system stored in the blockchain	3,75	3,00

Moreover, end-users provided additional requirements during the workshop in Bydgoszcz. They were asked to add requirements that are not initially listed in the form of user story, to focus on real user needs and their perspective. Therefore, additional requirements provided as a result of the workshop are:

- Credibility score for online content could be provided in the 1 to 10 range.
- Assessment of the image and text posted online could be carried out taking into account the context of both text and images, because text and picture can be true considered separately and could be fake news when posted together.
- SocialTruth platform could analyse and provide to the user information on geographical location of the original post and its re-posting sources.

In the next section, this list is split into particular requirements provided in the template and described with an additional information.

7.2 SocialTruth Functional Requirements

Id: F1	Filter out source/content by credibility index				
Status: Proposed	Priority: M	Proponent: UTP	Version: 1.0	Related WP: WP4	Related use-case: Project-wide

Id: F2	**Flag untrusted information/source**				
Status: Proposed	Priority: M	Proponent: UTP	Version: 1.0	Related WP: WP4	Related use-case: Project-wide

Id: F3	**Integrate functionalities into the core editorial software dedicated to journalists**				
Status: Proposed	Priority: C	Proponent: end-users	Version: 2.0	Related WP: WP5	Related use-case: UC1

Id: F4	**Provide access to the functionalities for a standard Internet user**				
Status: Proposed	Priority: M	Proponent: UTP/end-users	Version: 2.0	Related WP: WP4-WP5	Related use-case: Project-wide

Id: F5	**Provide the web frontend access to the functionalities for a professional users**				
Status: Proposed	Priority: S	Proponent: end-users	Version: 2.0	Related WP: WP4-WP5	Related use-case: UC1, UC4

Id: F6	**Allow the user to launch the verification process**				
Status: Proposed	Priority: S	Proponent: UTP	Version: 1.0	Related WP: WP4	Related use-case: Project-wide

Id: F7	**Ask the user to provide feedback on automated verification**				
Status: Proposed	Priority: S	Proponent: UTP	Version: 2.0	Related WP: WP4	Related use-case: Project-wide

Id: F8	**Ask the professionals and individuals to annotate (justify or explain) reasons of the verification feedback**				
Status: Proposed	Priority: S	Proponent: UTP/end-users	Version: 1.0	Related WP: WP4	Related use-case: Project-wide

Id: F9	Allow user to specify fraudulent criteria for a search results					
Status: Proposed	Priority: S	Proponent: UTP	Version: 1.0	Related WP: WP3-WP4	Related use-case: UC3	

Id: F10	Inform the user about possibility of fake news in the search results					
Status: Proposed	Priority: S	Proponent: UTP	Version: 1.0	Related WP: WP3-WP4	Related use-case: UC3	

Id: F11	Re-rank the search result based on the fraudulent criteria					
Status: Proposed	Priority: S	Proponent: UTP	Version: 1.0	Related WP: WP3-WP4	Related use-case: UC3	

Id: F12	Integrate functionalities into the platform for lesson design					
Status: Proposed	Priority: S	Proponent: UTP/end-users	Version: 2.0	Related WP: WP5	Related use-case: UC4	

Id: F13	Display credibility index of the educational materials immediately (e.g. when typing web address)					
Status: Proposed	Priority: S	Proponent: end-users	Version: 1.0	Related WP: WP3-WP4	Related use-case: UC4	

Id: F14	Provide user (teacher) justification (give reason) why the content is untrusted					
Status: Proposed	Priority: S	Proponent: end-users	Version: 3.0	Related WP: WP4	Related use-case: UC4/project-wide	

Id: F15	Allow for semantic text searching					
Status: Proposed	Priority: S	Proponent: UTP	Version: 1.0	Related WP: WP3	Related use-case: Project-wide	

Id: F16	**Allow for automated categorization of content based on text analytics**				
Status: Proposed	Priority: S	Proponent: UTP	Version: 1.0	Related WP: WP3	Related use-case: Project-wide

Id: F17	**Allow for identification of a content author's writeprint based on stylometric analysis (at least based on usage of short words, conjunctions, vocabulary richness and complexity, lexical differentiation)**				
Status: Proposed	Priority: S	Proponent: UTP	Version: 1.0	Related WP: WP3	Related use-case: Project-wide

Id: F18	**Allow for mapping authors writing style to another styles from a historical database**				
Status: Proposed	Priority: S	Proponent: UTP	Version: 1.0	Related WP: WP3	Related use-case: Project-wide

Id: F19	**Allow for mapping authors identity to another identities (author names or nicknames) from a historical database based on the stylometric analysis**				
Status: Proposed	Priority: S	Proponent: UTP	Version: 1.0	Related WP: WP3	Related use-case: Project-wide

Id: F20	**Allow for creating new ontologies and taxonomies automatically based on the machine learning techniques**				
Status: Proposed	Priority: C	Proponent: UTP	Version: 2.0	Related WP: WP3	Related use-case: Project-wide

Id: F21	**Allow for identification of writer's emotions based on the text content (deeper granularity than a standard positive/neutral/negative evaluation)**				
Status: Proposed	Priority: C	Proponent: UTP	Version: 2.0	Related WP: WP3	Related use-case: Project-wide

Id: F22	**Allow for identification of writer's intention (estimate a bias)**				
Status: Proposed	Priority: S	Proponent: UTP	Version: 2.0	Related WP: WP3	Related use-case: Project-wide

Id: F23	Allow for assessment of authenticity of image posted online					
Status: Proposed	Priority: M	Proponent: UTP	Version: 1.0	Related WP: WP3	Related use-case: Project-wide	

Id: F24	Store analysis performed by different users of the system in the blockchain					
Status: Proposed	Priority: M	Proponent: UTP	Version: 1.0	Related WP: WP4	Related use-case: Project-wide	

Id: F25	Allow the users to access and check the historical evaluations performed by the system stored in the blockchain					
Status: Proposed	Priority: M	Proponent: UTP	Version: 1.0	Related WP: WP4	Related use-case: Project-wide	

Id: F26	Credibility score for online content provided in the 1 to 10 range					
Status: Proposed	Priority: C	Proponent: end-users	Version: 1.0	Related WP: WP4	Related use-case: Project-wide	

Id: F27	Credibility assessment and analysis combined for both text and multimedia context (separately could not be fake)					
Status: Proposed	Priority: C	Proponent: end-users	Version: 1.0	Related WP: WP3-WP4	Related use-case: Project-wide	

Id: F28	Information on geographical location of the original post and its re-posting sources					
Status: Proposed	Priority: C	Proponent: end-users	Version: 1.0	Related WP: WP3-WP4	Related use-case: Project-wide	

7.3 Non-Functional Requirements

Id: NF1	Ease requesting the verification of a digital article or social media post (limited number of clicks to launch the process)				
Status: Proposed	Priority: S	Proponent: end-users	Version: 1.0	Related WP: WP2-WP3-WP4	Related use-case: Project-wide

Id: NF2	Near real-time external educational materials verification				
Status: Proposed	Priority: S	Proponent: end-users	Version: 1.0	Related WP: WP3-WP4	Related use-case: UC4

7.4 UML use case diagrams

In this section four UML use-case diagrams related to four project use-cases are presented. Those diagrams present basic functionalities that should be offered by the SocialTruth platform from the perspective of different actors (users).
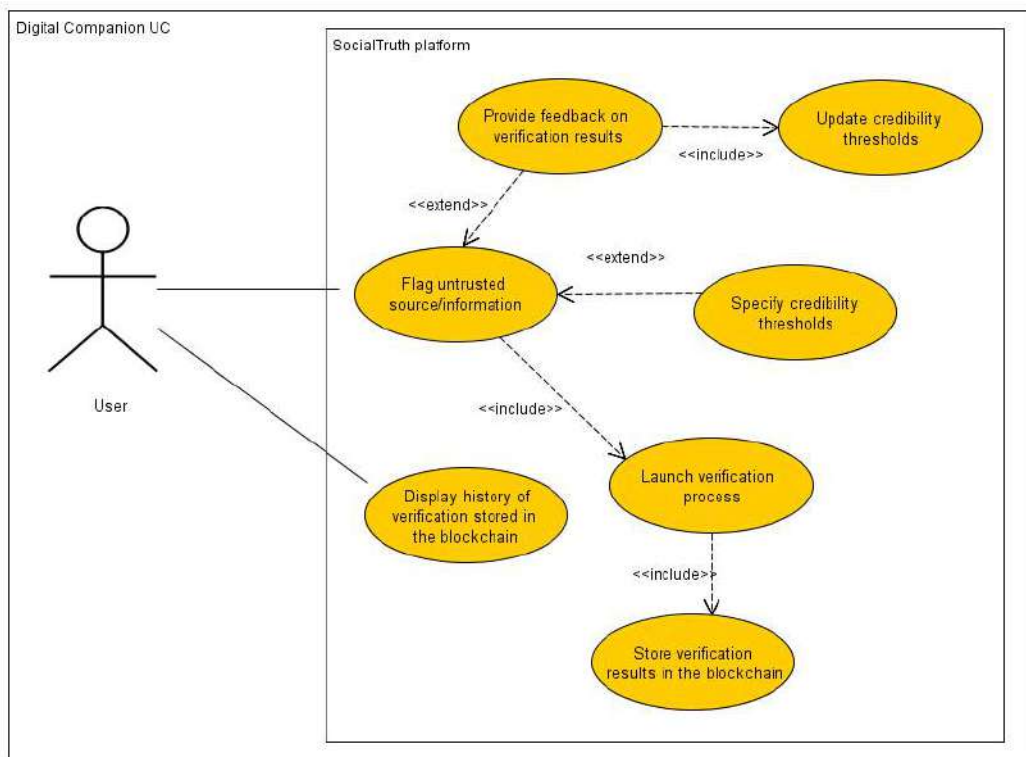

Figure 22: Use-case diagram for UC2: Digital companion for content verification

Description of Digital Companion scenario:

- Actors: user of the SocialTruth platform

- Sub-use-case 1: **Flag untrusted source/information:** user can launch verification process with the minimum interaction with the SocialTruth browser plugin and get the credibility score related to verified online content. Results of verification process are always stored in the blockchain (with relevant meta-information) and can be accessed in the future as the historical verification results (sub-use-case-2). Optionally, user can specify/change criteria for verification process (e.g. by specifying weights for results coming from different verification services or ignoring given services in credibility computation). Optionally, user can provide feedback for verification results, i.e. can fine-tune the thresholds at which content is flagged as a fake.

- Sub-use-case 2: **Display history of verification results stored in the blockchain**: user can check historical records of SocialTruth verification results related to analysed content.

Figure 23: Use-case diagram for UC4: External sources reliability check in the educational domain

Description of Educational domain scenario:

- Actors: user of the SocialTruth platform/teacher designing lessons
- Sub-use-case 1: **Annotate untrusted source/information:** user can launch verification process with the minimum interaction with the SocialTruth browser plugin and get the credibility score related to verified online content to be included in the lesson. Results of verification process are always stored in the blockchain (with relevant meta-information) and can be accessed in the future as the historical verification results (sub-use-case-2). Optionally, user can specify/change criteria for verification process (e.g. by specifying weights for results coming from different verification services or ignoring given services in credibility computation). Optionally, user can

view the justification on why the content is untrusted, i.e. indication of which criteria threshold are exceeded.

- Sub-use-case 2: Display history of verification results stored in the blockchain: user can check historical records of SocialTruth verification results related to analysed content.
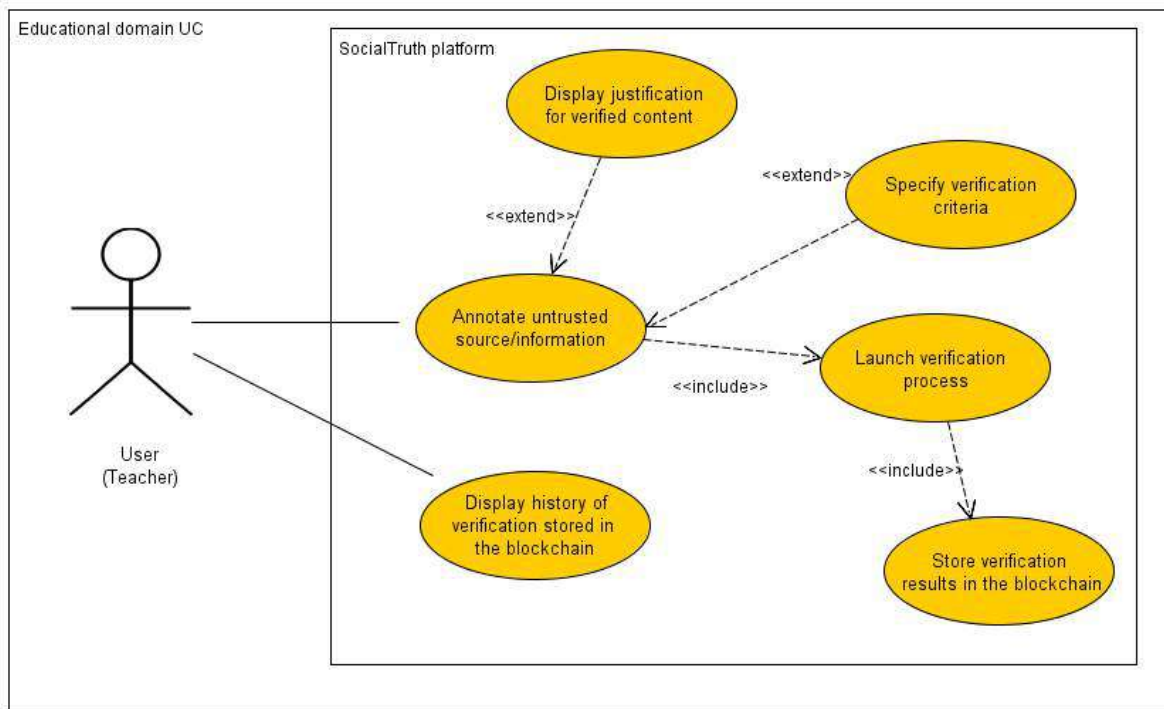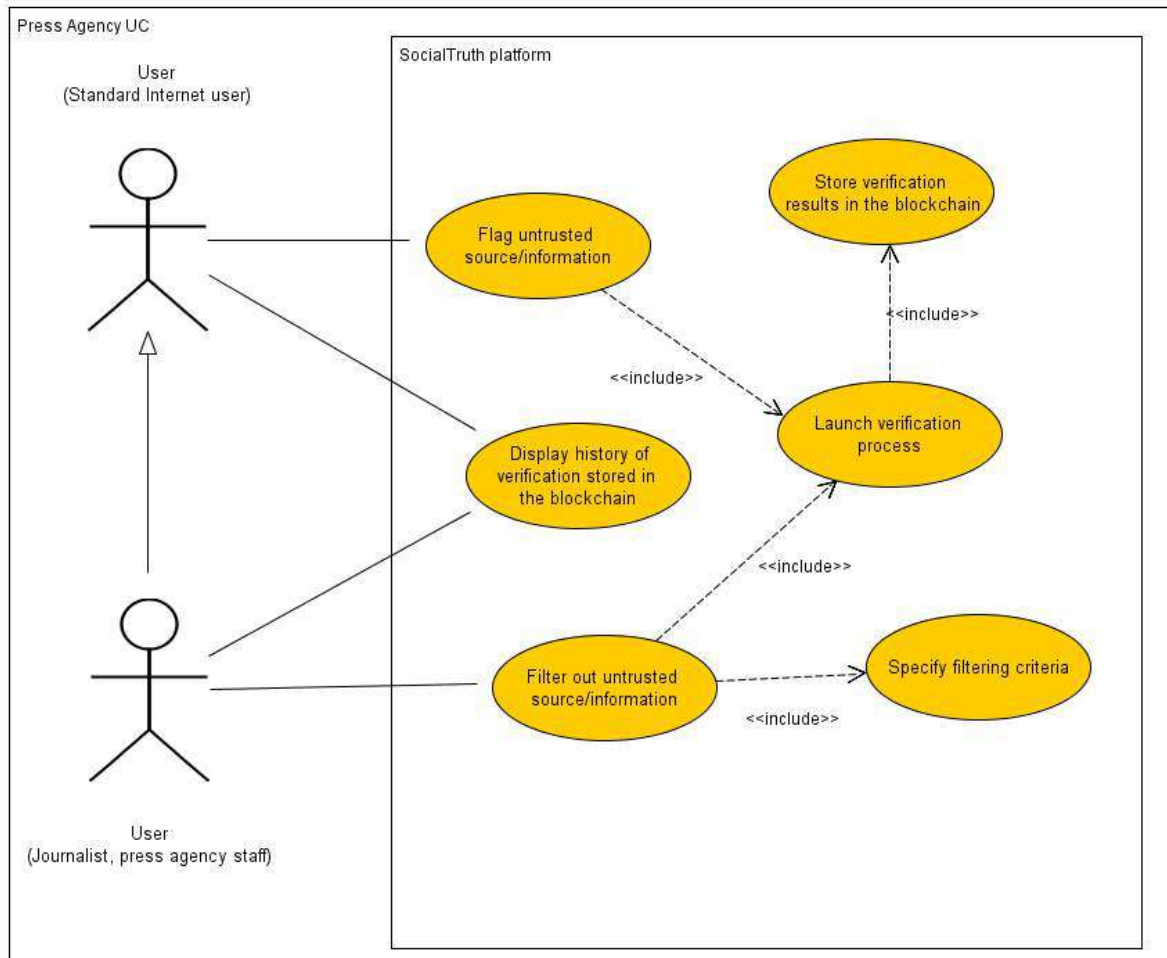

Figure 24: Use-case diagram for UC1: Checking sources in the production process

- Actors: Internet user of the SocialTruth platform/press agency staff (journalist)
- Sub-use-case 1: **Flag untrusted source/information:** Internet user can launch verification process with the minimum interaction with the SocialTruth browser plugin and get the credibility score related to verified online content to be included in the lesson. Results of verification process are always stored in the blockchain (with relevant meta-information) and can be accessed in the future as the historical verification results (sub-use-case-2).
- Sub-use-case 2: **Filter out untrusted source/information**: professional user can filter out (exclude) content that is untrusted after specifying filtering criteria (verification services that will be taken into account).

- Sub-use-case 3: **Display history of verification results stored in the blockchain**: user (both professional and unprofessional) can check historical records of SocialTruth verification results related to analysed content.
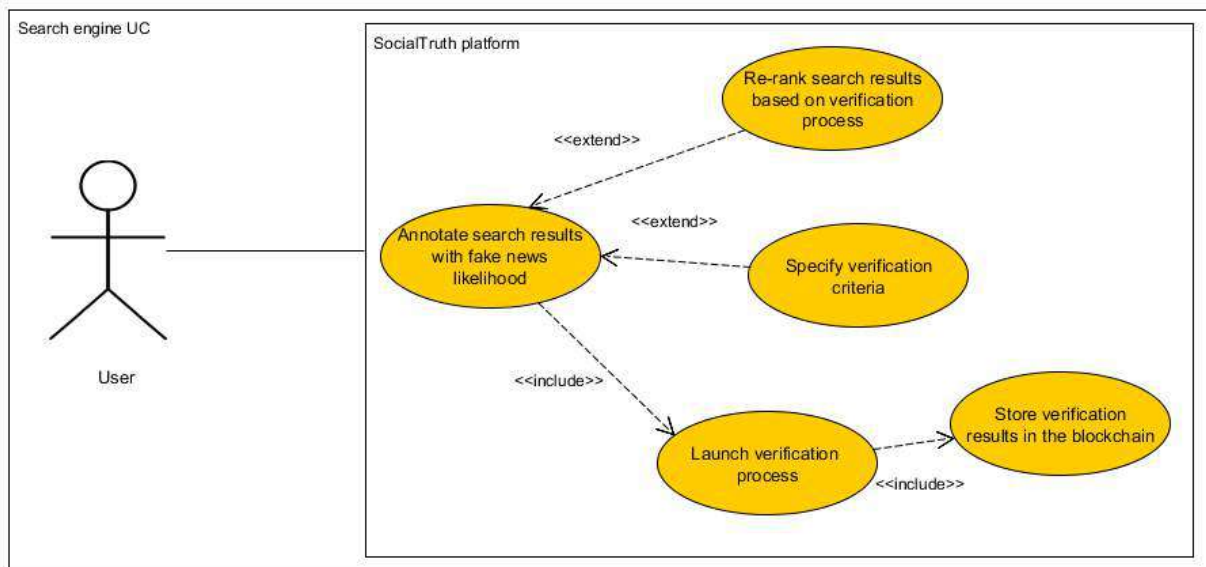
Figure 25: Use-case diagram for UC3: Search engine rankings and advertising prevention

- Actors: Internet user of the SocialTruth platform
- Sub-use-case 1: **Annotate search results with credibility score:** user of QWANT search engine can see fake news likelihood assigned to all search results. Optionally, user can specify criteria to finetune the solution (specify weights for results of different verification services) and launch re-ranking of the original search results with the credibility scores taken into account.

In addition, the general flow of the SocialTruth verification process have been shown in the next UML sequence diagram. This diagram presents actions and dependencies between the SocialTruth architecture layers during the verifaction process.

Figure 26: General sequence diagram for verification process using the SocialTruth platform

8. Trust and decentralisation in SocialTruth

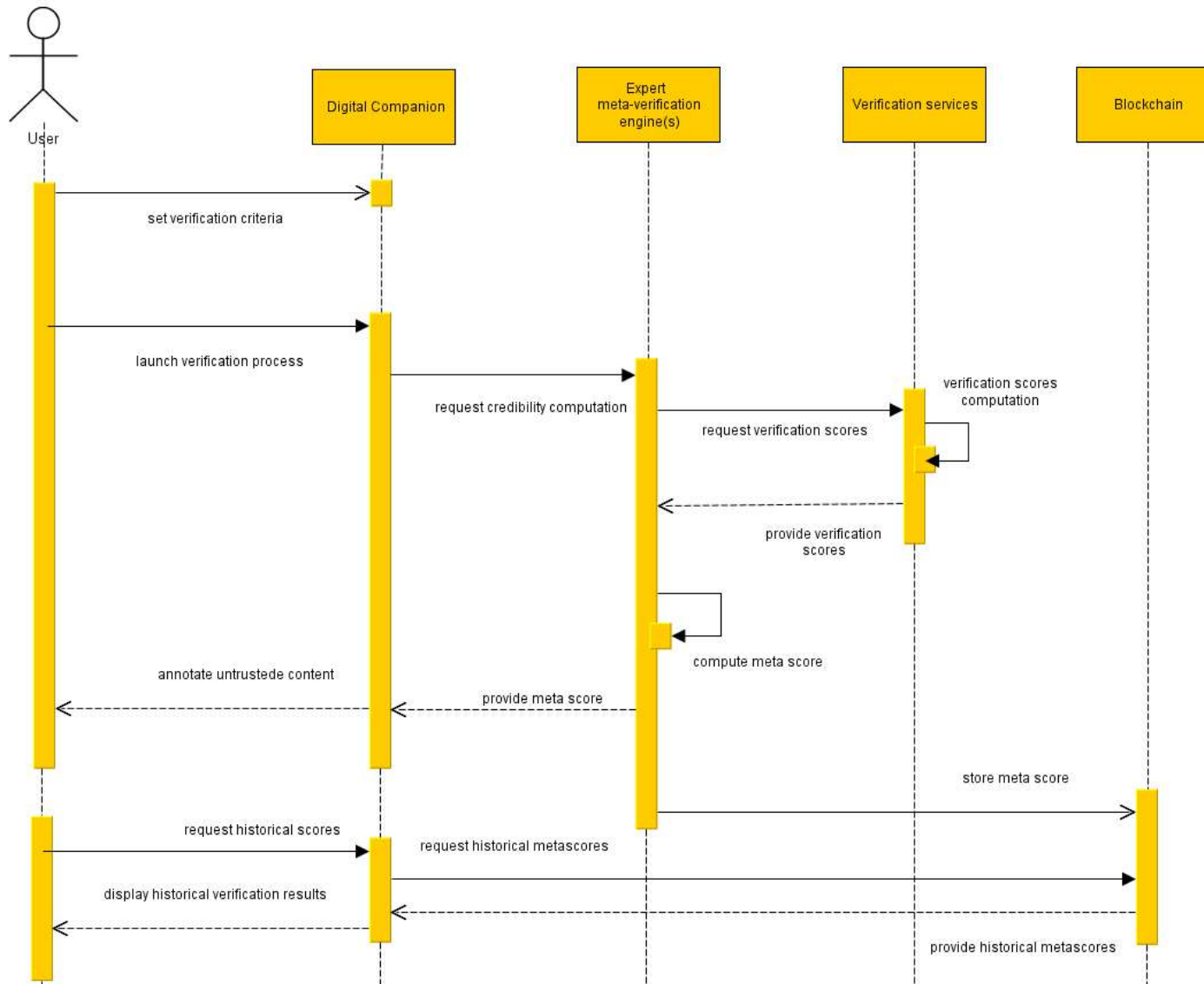A key priority for SocialTruth is the development of an open, democratic, pluralistic and distributed ecosystem that will be supported by a service-driven architecture, which will enable the introduction of third-party verification services, i.e. services that support specific content verification objectives and are offered by external providers. These providers can be SMEs, start-ups, industries, academia and other user communities, beyond the ones that participate in the SocialTruth consortium, as long as they possess or develop content verification technologies.

Expert Meta-Verification Engines (EMVEs) use own algorithms, or machine learning techniques, to combine, weight and fuse results from different verification service providers into single meta-scores. This is also a key concept for the democratisation of content verification, because in the framework of SocialTruth, the end-users will be able to review responses from different EMVEs and decide based on their individual criteria. We could imagine that several organisations could play the role of EMVE providers, be it large news agencies, media industries, public agencies, non-governmental organisations, and others. Some EMVEs could be specialized with respect to their news types or content topics. For example, an EMVE specialized in sport news might provide more accurate results on these specific types of news than on other news types. What is more, some EMVEs could use several or few content verification services offered by third parties. There could be EMVEs that rely solely on an automated meta-verification process, or EMVEs that might also leverage human intervention.

EMVEs should store their assessment outputs in the SocialTruth Blockchain, thus creating an auditable, immutable, secure and trustable trail of content verification results, without the need to have a centralized authority. Inside the SocialTruth blockchain, and for each different article assessment conducted, the EMVEs store their own id, the ids of the verification services they have used to produce the meta-score, together with the article id and the meta-score itself.

During its second half, SocialTruth will be placing a priority on cases where multiple Expert Meta-Verification Engines (EMVEs) participate in its ecosystem. In light of the above, the core SocialTruth concept has been that the SocialTruth Blockchain acts as an auditable aggregator of different article assessment results, and that the end-users shall have the freedom to select (through their Digital Companion) the EMVEs they want or find to be more credible. This core concept can be further extended: the project can also examine more closely the possibility of achieving consensus among various (possibly conflicting) assessments of individual EMVEs.

Discussing this in more detail, the following approaches exist for the implementation of the SocialTruth Blockchain:

- **Approach 1 – SocialTruth Blockchain as auditable aggregator (no conflict resolution):** Following this approach, the SocialTruth Blockchain stores different meta-scores (credibility assessment results), even if they are conflicting. Through the Digital Companion, end-users choose the EMVEs they wish to use, having complete freedom of choice. Statistics regarding fake news detection

success rates could also help them in choosing. Professional users are more probable to pay for the services of high-quality EMVEs, so eventually the number of low-quality EMVEs is expected to deteriorate (the same applies not only for the EMVEs but also for any biased or low-quality verification service providers within the ecosystem). In this sense, even if an adversary or malicious actor enters the SocialTruth ecosystem with multiple EMVEs, it does not pose a critical threat, since end-users are able to configure their Digital Companion to connect to a high-quality EMVE of their choosing (and have an incentive to do so). Still, in cases of extreme differences among high-quality EMVEs, human fact-checkers are expected to intervene for resolving the conflict.

- **Approach 2 – Conflict resolution through consensus in the SocialTruth Blockchain:** In this approach, before writing to the blockchain, the EMVEs attempt to reach a consensus. Following the consensus, only a single meta-score is stored in the blockchain (i.e. for the same article). Following this approach, if an adversary or malicious actor enters the SocialTruth ecosystem with multiple EMVEs, a threat arises, since it could attempt to manipulate the consensus. This implies that, for such a scheme to function properly, we should establish mitigation measures, e.g. by allowing only authorized EMVEs to write into the SocialTruth Blockchain, by complementing this approach with a reputation mechanism for EMVEs, or by investigating an appropriate proof-of-stake algorithm. Each solution or combination of solutions has its pros and cons, which will be deeply investigated in the framework of WP4 activities, especially considering the balance between censorship and consensus manipulation risks. We should also acknowledge that such highly-automated conflict resolution poses several additional challenges that reside at the borders of the SocialTruth scope (in contrast to Approach 1) and in some cases beyond that. It seems, however, that reputation should play a role in such consensus, and that EMVEs should have counterincentives to produce false results on purpose.

Considering the above, SocialTruth will implement the first approach while also investigating the feasibility of a good implementation of the second approach.

In the real world, peer review and feedback play a major role in a system's regulation or self-regulation. Such is the case, for example, of the publication process within the scientific community. In the Socialtruth ecosystem, we may envision that the fact-checking community would periodically and randomly sample the SocialTruth Blockchain to validate a part of the stored assessment results. Such periodic audit procedures, with the cooperation of the fact-checking community and the EU Observatory, would influence the reputation ranking of EMVEs and would reinforce the trust in the SocialTruth blockchain-based ecosystem.

Furthermore, the ability for the end-user to provide feedback on the meta-verification results that seem more relevant, is also an important element that will be considered inside the project, in order to help in acknowledging the most relevant (high-quality) EMVEs and in identifying the less accurate (low-quality) EMVEs, based on the collected feedback responses. Like the aforementioned audit procedures, this user feedback mechanism is also expected to help in distinguishing the high performing from the poorly performing EMVEs, and to reinforce the trust in the SocialTruth blockchain-based ecosystem..

For SocialTruth, the key priority remains to develop the open and pluralistic ecosystem, and subsequently to deal with the identified aforementioned issues, also with the help of the pilot activities. For this purpose, D2.1 is a baseline for the evaluation scenarios that will be developed and executed in WP5. The user feedback will be utilized in the trials, following a ranking approach that will distinguish the features that correspond to the current scope of the system, also aiming to provide policy and technological recommendations regarding evaluation remarks that would constitute significant knowledge to promote further development of machine learning based verification services.

9. Summary and conclusions

D2.1 was the first critical exercise for the SocialTruth consortium. In this we worked closely together starting as early as the Kick-off meeting on December 2019 and closing this in of activities with the second project meeting in Bydgoszcz on March 2019. During this short period we organized subsequent webinars aiming to impose our requirements elicitation methodologies and deliver refined and updated versions of the main SocialTruth concepts and use cases descriptions. This has led to the first batch of requirements that have been prioritized and validated by the consortium in our second plenary meeting. This work is presented in Section 7.

For the next steps, we will build upon the developments presented in this deliverable, to develop the technical specifications of the SocialTruth architecture and the first development iteration of the SocialTruth tools and services.

Still, the delivery of the first application versions will allow us to further query the feedback of the end-users and further improve the initial SocialTruth specifications. In this way the final platform version, will not only deliver a more mature technological solution, but will be more adapted to the actual expectations of our end users. Achieving this goal, will improve the readiness of the SocialTruth to address the requirements of the content verification market and improve the commercialisation potential.

In addition, D2.1 was updated on February 2020 in order to further specify trust and blockchain aspects. These have been included in a new chapter and will form the basis for the end-user pilot evaluation scenarios in WP5. Implementation details from and software architectural standpoint are elaborated in D2.3. Detailed implementation design decisions (e.g. selection of technology and non-functional specs) are planned to be produced as part of WP4 and WP5 deliverables.

10. References

[1] SocialTruth Consortium, "SocialTruth DoA"

[2] Clegg, Dai; Barker, Richard (1994). Case Method Fast-Track: A RAD Approach. Addison-Wesley. ISBN 978-0-201-62432-8.

[3] Anoop K., Gangan M.P., P D., Lajish V.L. (2019) Leveraging Heterogeneous Data for Fake News Detection. In: P D., Jurek-Loughrey A. (eds) Linking and Mining Heterogeneous and Multi-view Data. Unsupervised and Semi-Supervised Learning. Springer, Cham

[4] Boididou, Christina, Middleton, Stuart E., Jin, Zhiwei, Papadopoulos, Symeon, Dang-Nguyen, Duc-Tien, Boato, Giulia and Kompatsiaris, Yiannis (2017) Verifying information with multimedia content on twitter: A comparative study of automated approaches. *Multimedia Tools and Applications*.

[5] Mazumdar, Aniruddha, Jaya Singh, Yosha Singh Tomar and Prabin Kumar Bora. "Universal Image Manipulation Detection using Deep Siamese Convolutional Neural Network." *CoRR* abs/1808.06323 (2018): n. pag.

[6] Bonettini, N. (2017). JPEG-based forensics through convolutional neural networks.

[7] Zhou, P., Han, X., Morariu, V. I., & Davis, L. S. (2018). Learning Rich Features for Image Manipulation Detection. *arXiv preprint arXiv:1805.04953*.

[8] Nakamoto, S. (2008). Bitcoin: A peer to peer electronic cash system.

[9] Chen Z. and Liu B. "Lifelong Machine Learning." Morgan & Claypool Publishers, 2016.

[10] Bing Liu, (2012). Sentiment Analysis and Opinion Mining. Synthesis Lectures on Human Language Technologies, Morgan & Claypool Publishers.

[11] DARPA L2M - Lifelong Learning Machines, URL: https://www.darpa.mil/program/lifelong-learning-machines

[12] D77F 35: Systems Development: Structured Design Methods, SQA 2007

[13] Azoulay, A. (2018). World Trends in Freedom of Expression and Media Development Global Report 2017/2018 (pp. 200).

[14] Baxter, K., Courage, C., & Caine, K. (2015). Chapter 11 - Card Sorting. In K. Baxter, C. Courage & K. Caine (Eds.), Understanding your Users (Second Edition) (pp. 302-337). Boston: Morgan Kaufmann.

[15] BKCASE. (2017). Editorial Board. The Guide to the Systems Engineering Body of Knowledge (SEBoK), v. 1.9.1 R.J. Cloutier (Editor in Chief). Hoboken, NJ: The Trustees of the Stevens Institute of Technology. BKCASE is managed and maintained by the Stevens Institute of Technology Systems Engineering Research Center, the International Council on Systems Engineering, and the Institute of Electrical and Electronics Engineers Computer Society.

[16] Carrizo, D., Dieste, O., & Juristo, N. (2014). Systematizing requirements elicitation technique selection. Information and Software Technology, 56(6), 644-669. doi: https://doi.org/10.1016/j.infsof.2014.01.009

[17] Comision, E. (2018). A multi-dimensional approach to disinformation. Final report of the High Level Expert Group on Fake News and Online Disinformation.

[18] Holan, A. D. (2014). 7 steps to better fact-checking. PolitiFact.

[19] Jang, S. M., Geng, T., Queenie Li, J.-Y., Xia, R., Huang, C.-T., Kim, H., & Tang, J. (2018). A computational approach for examining the roots and spreading patterns of fake news: Evolution tree analysis. Computers in Human Behavior, 84, 103-113. doi: https://doi.org/10.1016/j.chb.2018.02.032

[20] Jang, S. M., & Kim, J. K. (2018). Third person effects of fake news: Fake news regulation and media literacy interventions. Computers in Human Behavior, 80, 295-302. doi: https://doi.org/10.1016/j.chb.2017.11.034

[21] Liu, L., Zhou, Q., Liu, J., & Cao, Z. (2017). Requirements cybernetics: Elicitation based on user behavioral data. Journal of Systems and Software, 124, 187-194. doi: https://doi.org/10.1016/j.jss.2015.12.030

[22] McLeod, R. W. (2017). Human factors in barrier management: Hard truths and challenges. Process Safety and Environmental Protection, 110, 31-42. doi: https://doi.org/10.1016/j.psep.2017.01.012

[23] Melander, L., Dubois, A., Hedvall, K., & Lind, F. (2019). Future goods transport in Sweden 2050: Using a Delphi-based scenario analysis. Technological Forecasting and Social Change, 138, 178-189. doi: https://doi.org/10.1016/j.techfore.2018.08.019

[24] Mishra, D., Aydin, S., Mishra, A., & Ostrovska, S. (2018). Knowledge management in requirement elicitation: Situational methods view. Computer Standards & Interfaces, 56, 49-61. doi: https://doi.org/10.1016/j.csi.2017.09.004

[25] Thomson, R. (2016). What Is Stakeholder Management? Planning Your Stakeholder Communications, from https://www.mindtools.com/pages/article/newPPM_08.htm

[26] Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. Science, 359(6380), 1146-1151. doi: 10.1126/science.aap9559

[27] Yang, J.-B., & Xu, D.-L. (2013). Evidential reasoning rule for evidence combination. Artificial Intelligence, 205, 1-29. doi: https://doi.org/10.1016/j.artint.2013.09.003

[28] Zhao, L., Yin, J., & Song, Y. (2016). An exploration of rumor combating behavior on social media in the context of social crises. Computers in Human Behavior, 58, 25-36. doi: https://doi.org/10.1016/j.chb.2015.11.054

[29] A. Facchi, «La verità come interesse collettivo,» Biblioteca della libertà, LII, 2017.

[30] S. Vosoughi, D. Roy e S. Aral, «The spread of true and false news online,» Science, vol. 359, pp. 1146-1151, 2018.

[31] High Level Group on Media Freedom and Pluralism, «A free and pluralistic media to sustain European democracy,» 2013.

[32] «IFJ International Federation of Journalists,» [Online]. Available: https://www.ifj.org/who/rules-and-policy/principles-on-conduct-of-journalism.html.

[33] Z. Tufekci, «It's the (Democracy-Poisoning) Golden Age of Free Speech,» Wired, January 16, 2018.

[34] E. C. J. Tandoc, Z. Wei Lim e R. Ling, «Defining "fake news", a topology of scholarly definitions,» Digital Journalism, vol. 6, n. 2, pp. 137-153, 2018.

[35] C. J. Glynn, M. E. Huge e L. H. Hoffman, «All the news that's fit to post: A profile on news use on social networking sites,» Computes in Huma Behaviour, vol. 28, pp. 113-119, 2012.

[36] A. Hermida, F. Fletcher, D. Korell e D. Logan, «SHARE, LIKE, RECOMMEND Decoding the social media news consumer,» Journalism Studies, vol. 13, n. 5-6, pp. 815-824, 2012.

[37] B. Brett, «The Psychology of sharing: why do people share online?,» Customer Insight Group - New York Times;, 2018.

[38]C. S. Lee e L. Ma, «News sharing in social media: the effect of gratifications and prior experience,» Computers in Human Behavior, vol. 28, pp. 331-339, 2012.

[39]M. Opgenhaffen, «Multimedia, interactive and hypertextual features in divergent online news platforms: an exploratory study of flemish online news,» First Monday, vol. 16, n. 3, 2011.

[40]B. Franklin, M. Hamer, M. Hanna, M. Kinsey e J. E. Richardson, Key concepts in journalism studies, London: Sage, 2005.

[41]E. a. S. R. Council, «The age of the hypertext,» 2018.

[42]S. Cicconi, Mediapolis : Aspects of Texts, Hypertexts und Multimedial Communication - Hypertextuality 21-43, Berlin-New York: De Gruyter , 1999.

[43]M. Deuze, «Online journalism: modelling the first generation of news media of the world wide web,» First Monday, vol. 6, n. 10, 2001.

[44]K. Aoki, «Taxonomy of Interactivity on the Web,» Boston University.

[45]S. Rafaeli, «Interactivity: From new media to communication,» in Advancing Communication Science: Merging Mass and Interpersonal Process, Newbury Park, CA, Sage, 1988.

[46]S. Rafaeli e F. Sudweeks, «Networked interactivity,» Journal of Computer-Mediated Communication, vol. 2, n. 4, 1997.

[47]Y.-T. Wang, L.-L. Wu, H.-C. Chen e M.-Y. Yeh, «Interactivity of Social Media and Online Consumer Behavior: the Moderating Effects of Opinion Leadership,» in ICIS 2012, 2012.

[48]A. Richard Schrock, «Communicative Affordances of Mobile Media: Portability, Availability, Locatability, and Multimediality,» International Journal of Communication, vol. 9, pp. 1229-1246, 2015.

[49]M. Arnold, «On the phenomenology of technology: The "Janus-faces" of mobile phones,» Information and Organization, vol. 14, n. 5, p. 231–256, 2003.

[50]J. E. Katz e M. Aakhus, Perpetual contact: Mobile communication, private talk, public performance, Cambridge, UK: Cambridge University Press, 2002.

[51]C. Licoppe, «Connected presence: The emergence of a new repertoire for managing social relationships in a changing communication technoscape,» Society and Space, vol. 22, n. 1, p. 135–156, 2004.

[52]B. Light e E. Cassidy, «Strategies for the suspension and prevention of connection: Rendering disconnection as socioeconomic lubricant with Facebook,» New Media & Society, vol. 16, n. 7, p. 1169–1184, 2014.

[53]R. Coyne, The tuning of place: Sociable spaces and pervasive digital media, Cambridge, MA: MIT Press, 2010.

[54]A. de Souza e Silva e J. Frith, «Locative mobile social networks: Mapping communication and location in urban spaces,» Mobilities, vol. 5, n. 4, p. 485–505, 2010.

[55]H. Rheingold, Smart mobs: The next social revolution, Cambridge, MA: Perseus Publishing, 2002.

[56]E. Gordon, «Towards a theory of network locality,» First Monday, vol. 13, n. 10, 2008.

[57]J. Farman, Mobile interface theory: Embodied space and locative media, New York, NY: Routledge, 2012.

[58]we are social, «global digital report 2018,» 2018. [Online]. Available: https://digitalreport.wearesocial.com/.

[59]Reuters Institute for the Study of Journalism, «Digital News Report 2018,» 2018.

[60] S. r. b. t. E. Commission, «Flash Eurobarometer 464 - Fake news and disinformation online,» April 2018. [Online]. Available: https://data.europa.eu/euodp/data/dataset/S2183_464_ENG.

[61] N. A. Karlova e K. E. Fisher, «A social diffusion model of misinformation and disinformation for understanding human information behaviour,» Information Research, vol. 18, n. 1, 2013.

[62] R. Goldman, «Reading fake news, Pakistani Minster directs nuclear threat at Israel,» The New York Times, 2016.

[63] G. Lopez, «Pizzagate, the fake news conspiracy theory that led a gunman to DC's Comet Ping Pong, explained,» Vox, 2016.

[64] C. Kang e A. Goldman, «In Washington Pizzeria Attack, Fake News Brought Real Guns,» The New York Times, 2016.

[65] Y. Adegoke e BBC Africa Eye, «Like. Share. Kill. Nigerian police say false information on Facebook is killing people,» 2018. [Online]. Available: https://www.bbc.co.uk/news/resources/idt-sh/nigeria_fake_news.

[66] V. Goel, S. Raj e P. Ravichandran, «How WhatsApp Leads Mobs to Murder in India,» The New York Times, 18 July 2018.

[67] M. Molteni, «When WhatsApp's Fake News Problem Threatens Public Health,» Wired, 03 September 2018.

[68] J. Suarez Villegas, «Ethical and deontological aspects of online journalism. Their perception by journalists,» Revista Latina de Comunication Social, vol. 69, pp. 91-109, 2015.

[69] B. Kovach e T. Rosenstiel, The Elements of Journalism: What Newspeople Should Know and the Public Should Expect, Three Rivers Press, 3rd edition (2014).

[70] P. B. Brandtzaeg, M. Lüders, J. Spangenberg, L. Rath-Wiggins e A. Følstad, «Emerging Journalistic Verification Practices Concerning Social Media,» Journalism Practice, vol. 10, n. 3, pp. 323-342, 2016.

[71] K. Backholm, J. Ausserhofer, E. Frey, A. Grondhal Larsen, H. Hornmoen, J. Hogvag e G. Reimerth, «Crises, Rumours and Reposts: Journalists' Social Media COntent Gathering and Verification Practices in Breacking News Situations,» Media and COmmunication, vol. 5, n. 2, pp. 67-76, 2017.

[72] S. Schifferes, N. Newman, N. Thurman, D. Corney, A. Goker e C. Martin, «Identifying and verifying news through social media: developing a user-centered tool for professional journalists,» Digital Journalism, vol. 2, n. 3, pp. 406-418, 2014.

[73] R. Schwartz, M. Naaman e R. Teodoro, «Editorial algorithms: using social media to discover and report local news,» in Ninth International AAAI Conference on Web and Social Media, Oxford, UK, 2015.

[74] N. Diakopoulos, M. De Choudhury e M. Naaman, «Finding and assessing soccial media information sources in the context of journalism,» in Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Austin, TX, 2012.

[75] «Edelman - Trust Barometer Global Report 2018,» 2018. [Online]. Available: https://www.edelman.com/sites/g/files/aatuss191/files/2018-10/2018_Edelman_Trust_Barometer_Global_Report_FEB.pdf.

[76] «Edelman Trust Barometer Global Report 2019,» [Online]. Available: https://www.edelman.com/trust-barometer.

[77] L. Ashbaugh, «Click here for ccounter fake news suggestions that will blow your mind,» Global Policy Research Group, University of Maine SPIA, 2017.

[78] X. Chen, S.-C. J. Sin, Y.-L. Theng e L. C. Sian, «Deterring the spread of misinformation on social network sites: a social cognitive theory-guided inteervention,» ASIST, St. Louis, MO, USA, 2015.

[79] G. Foundation, «Assessing the effect of news source ratings on news content,» 2018.

[80] L. J. Handsfield, T. R. Dean e K. M. Cielocha, «Becoming Critical Consumers and Producers of Text: Teaching Literacy with Web 1.0 and Web 2.0,» The Reading Teacher, vol. 63, n. 1, p. 40–50, 2009.

[81] K. Purcell, A. Heaps, J. Buchanan e L. Friedrich, «How Teachers Are Using Technology at Home and in Their Classrooms,» Pew Research Center's Internet & American Life Project, Washington, D.C., 2013.

[82] T. Bean, S. Bean e K. Bean, «Intergenerational conversations and two adolescents' multiple literacies: Implications for redefining content area literacy,» Journal of Adolescent & Adult Literacy, vol. 42, n. 6, p. 438–448, 1999.

[83] T. Heick, «A Definition Of Web Literacy (And How Students Can Benefit),» 2017. [Online]. Available: https://www.teachthought.com/technology/a-definition-of-web-literacy-and-how-students-can-benefit/.

[84] D. M. West, «How to combat fake news and disinformation,» Brookings, 2017.

[85] J. Catapano, «TeachHub.com - Teaching Strategies to Detect Fake News,» [Online]. Available: http://www.teachhub.com/teaching-strategies-detect-fake-news.

[86] S. Wineburg, S. McGrew, J. Breakstone e T. Ortega, «Evaluating Information: The Cornerstone of Civic Online Reasoning. Stanford Digital Repository,» 2016. [Online]. Available: http://purl.stanford.edu/fv751yt5934.

[87] A. Miller, «The news Literary Project,» [Online]. Available: https://newslit.org/.

[88] M. Sullivan e T. Bajarin, «Can you spot fake news before hitting "share"? Kids are learning and so can you,» Fast Company, 2018.

[89] C. E. Shannon e W. Weaver, The mathematical theory of communication, Urbana, IL: University of Illinois Press, 1949.

[90] R. S. Taylor, «The process of asking questions,» American Documentation, vol. 13, n. 4, pp. 391-396, 1962.

[91] N. J. Belkin e S. E. Robertson, «Information science and the phenomenon of information,» Journal of the American Society for Information Science, vol. 27, n. 4, pp. 197-204, 1976.

[92] B. Dervin e M. Nilan, «Information needs and uses,» Annual Review of Information Science and Technology, vol. 21, n. 1, pp. 3-33, 1986.

[93] K. Tuominen e R. Savolainen, «A social constructionist approach to the study of information use as discursive action.,» Information Seeking in Context: Proceedings of the International Conference on Research in Information Needs, Seeking and Use , 1997.

[94] C. Wardle, «Fake news. It's complicated,» 16 February 2017. [Online]. Available: https://firstdraftnews.org/fake-news-complicated/.

[95] J. W. Kershner, The elements of news writing, Pearson Allyn and Bacon, 2005.

[96] B. Richardson, The process of writing news: from information to story, Pearson, 2007.

[97] K. H. Jamieson e K. Campbell, The interplay of influence: news, advertising, politics and the mass mesia, Wadsworth, 1997.

[98] B. Kovach e T. Rosenstiel, The elements of journalism: what newspeople should know and the public should expect - 1st rev.ed., Three Rivers Press, 2007.

[99] M. Schudson, The sociology of news. Edited by Jeffrey Alexander, Contemporary Societies, W.W. Norton & Cmpany, Inc, 2003.

[100]   E. S. Herman e N. Chomsky, Manufacturing consent: the political economy of the mass media, Knopf Doubleday Publishing Group, 2002.

[101]   G. Tuchman, Making news: a study in the construction of reality, Free Press, 1978.

[102]   D. M. White, «The "Gatekeeper". A case study in the selection of news,» Journalism Quarterly, vol. 27, pp. 383-390, 1950.

[103]   P. J. Shoemaker e S. D. Reese, Mediating the message in the 21st Century: a media sociology perspective - 3rd ed., Routeledge, 2013.

[104]   J. McManus, «What kind of commodity is news,» Communication Resarch, vol. 19, n. 6, pp. 787-805, 1992.

[105]   J. Johnstone, E. Slawski e W. Bowman, The news people: a sociological portrait of American journalists and their work, University of Illinois Press, 1976.

[106]   M. Schudson, «The sociology of news production,» Media, Culture & Society, vol. 11, n. 3, pp. 263-282, 1989.

[107]   V. Beal, «Webopedia,» [Online]. Available: https://www.webopedia.com/TERM/C/clickbait.html.

[108]   C. Cuvelliez, «Fake news ≠ false news,» LA Tribune, October 27, 2018.

[109]   P. Mihailidis e S. Viotty, «Spreadable Spectacle in Digital Culture: Civic Expression, Fake News, and the Role of Media Literacies in "Post-Fact" Society,» American Behavioral Scientist, vol. 61, n. 4, pp. 441-454 , 2017.

[110]   P. Zhang, «Motivational Affordances: Fundamental Reasons for ICT Design and Use,» 2008.