Review Article (Invited Authors Only)

new media & society
1–26

DOI: 10.1177/1461444820959296
journals.sagepub.com/home/nms
SAGE

A systematic literature review on disinformation: Toward a unified taxonomical framework

Eleni Kapantai, Androniki Christopoulou, Christos Berberidis and Vassilios Peristeras
International Hellenic University, Greece

Abstract

The scale, volume, and distribution speed of disinformation raise concerns in governments, businesses, and citizens. To respond effectively to this problem, we first need to disambiguate, understand, and clearly define the phenomenon. Our online information landscape is characterized by a variety of different types of false information. There is no commonly agreed typology framework, specific categorization criteria, and explicit definitions as a basis to assist the further investigation of the area. Our work is focused on filling this need. Our contribution is twofold. First, we collect the various implicit and explicit disinformation typologies proposed by scholars. We consolidate the findings following certain design principles to articulate an all-inclusive disinformation typology. Second, we propose three independent dimensions with controlled values per dimension as categorization criteria for all types of disinformation. The taxonomy can promote and support further multidisciplinary research to analyze the special characteristics of the identified disinformation types.

Keywords

Disinformation, fact-checking, fake news, false information, information disorder, taxonomy

Corresponding author:
Eleni Kapantai, School of Science & Technology, International Hellenic University, 14th km Thessaloniki, Moudania 57001, Thermi, Greece.
Email: e.kapantai@ihu.edu.gr

Introduction

Spreading false or inaccurate information is a phenomenon almost as old as human societies. Facts mingle with half-truths or untruths create "factitious informational blends" (Rojecki and Meraz, 2016). What is different today is the speed and the global reach this information disorder can attain (Niklewicz, 2017), coupled with the scale, complexity, and communication abundance (Blumler, 2015). Digital media and especially social media enable people to produce and rapidly spread incorrect information through decentralized and distributed networks (Benkler et al., 2018). In many cases, motives are malicious to promote preset beliefs with potentially harmful societal impact. This new, hyper-dynamic environment seems to introduce a new era in information flows and political communication that, according to Bennett and Pfetsch (2018), demands a reformulation of research frameworks, considering conceptual influences from social media and digital networks.

In the literature, there is a plethora of terms and concepts that are used to refer to false, untrue, or half-true information such as "fake news" (Lazer et al., 2018; Zhou and Zafarani, 2018), "false news" (Vosoughi et al., 2018), "digital misinformation" (World Economic Forum, 2018), "disinformation" (Amazeen and Bucy, 2019; HLEG, 2018; Wardle and Derekshan, 2017), "rumors" (Shao et al., 2018), and so on. The director of the Poynter Institute's International Fact-Checking Network blames media for the misuse of the term and the resulting ambiguity and confusion (Wendling, 2018). Especially the term "fake news" acquired global prominence in 2016, during the US presidential elections and the UK "Brexit" referendum. It was widely (ab)used in this political context to characterize almost any content in conflict with a particular party's views or agenda. Today, a search in Google with the term "fake news" returns approximately 80 million results. Likewise, a search for "false news" returns two million results, for "misinformation" about 35 million and for "disinformation" 13 million, verifying the popularity and the alternative vocabulary used. Google Trends shows a sharp surge of interest around "fake news" in November 2016 (Figure 1).

In our work, we focus on the term "disinformation," which, according to (HLEG, 2018), "includes all forms of false, inaccurate, or misleading information designed, presented and promoted to intentionally cause public harm or for profit."

Realizing the significant effect of false information on a global scale, academia, international, and other organizations try to first understand and then act against the phenomenon. This action takes various forms, including the launch of major counter-disinformation initiatives (European Commission, 2018a; Renda, 2018), articulating theoretical and computational approaches, preparing educational material ("Bad News Game," 2017), developing fact-checking platforms (InVID Project, 2017; Politifact, 2007; Snopes, 1994), and agreeing on a common code of principle for fact-checkers (IFCN, 2017). The European Commission works intensively since 2015 to ensure the protection of European values against the high exposure of citizens to this threat, introducing initiatives such as the High-Level Group of Experts, a public consultation and a Eurobarometer survey, the self-regulatory Code of Practices for the big social platforms (European Commission, 2018b), and so on.

In this article, we perform a thorough and systematic study of the literature to identify the overlapping terminology and typologies used. As a starting point, we adopt the

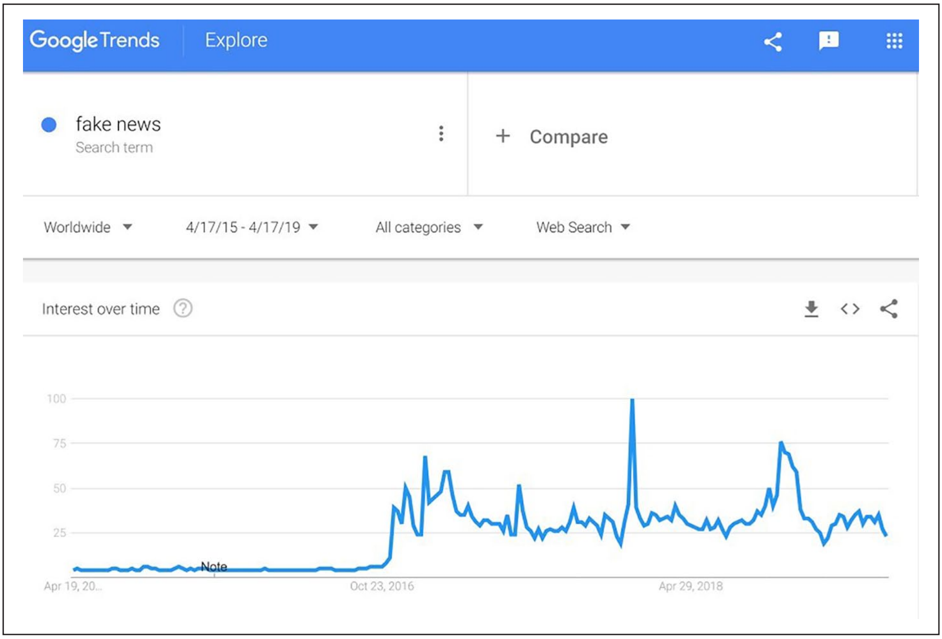

Figure 1. Frequency of "fake news" search term for the 2015–2019 time period.

definition of disinformation by HLEG (2018). We propose a conceptual framework for disinformation based on a typology and classification criteria.

The impact of disinformation

Entering a new era of information warfare, online platforms are weaponized to run targeted campaigns with false information (Zannettou et al., 2019). The consequences of disinformation can be devastating for every aspect of life.

In politics, disinformation has severe repercussions, ranging from legitimate propaganda to election manipulation. A Buzzfeed News analysis (Silverman, 2016) found that during the US presidential campaign, fake news election stories on Facebook outperformed those of news agencies. Similarly, research studies in Italy (Serhan, 2018), Nigeria (Kazeem, 2018), and Israel (Yaron, 2018) questioned integrity of elections, while Kušen and Strembeck (2018) revealed an alerting proliferation of misinformation during the 2016 Australian presidential election.

Concerning societal challenges, the spread of uncertainty, fear, and racism are only some of the consequences of disinformation. Studies in Germany (Müller and Schwarz, 2017, 2018) and the United States (Bursztyn et al., 2018) link content disseminated via social networks with incidents of hate crimes against ethnic minorities. In the UK, people wrongly associate European immigration with the decrease in the quality of healthcare services and increases in crime and unemployment rates (King's College and Ipsos

MORI, 2018). In terms of terrorism and homeland security (Aisch et al., 2016; Starbird et al., 2014), the infamous "pizzagate" story shows how disinformation can threaten not only democracy but human lives. In April 2020, the *Trends Alert* report (CTED, 2020) related COVID-19 conspiracy theories to terrorists' attempts to radicalize individuals and incite violence. One of these theories claims that "infected" immigrants were "imported" to decimate white populations (Wallner and White, 2020).

Pseudoscience can tremendously affect people's lives, provoking easily preventable disasters. In medicine and healthcare, extensively studied topics involving disinformation are vaccination, cancer, nutrition, and smoking (Albarracin et al., 2018; Jolley and Douglas, 2014; Syed-Abdul et al., 2013; Wang et al., 2019). Recently, during the COVID-19 explosion, the idea that death rates are being inflated and therefore there is no reason to observe lockdown regulations or other social distancing measures could help to further spread the epidemic (Lynas, 2020). Disinformation can also have a negative impact in environmental policies; Ward (2018) and Hotten (2015) are typical examples.

From an economic perspective, disinformation poses concern on both public economic growth and individuals' benefits. According to Reuters, conspiracy theories linking 5G (fifth-generation) technology to the spread of COVID-19 have resulted in over 140 arson attacks and assaults (Chee, 2020). Other studies investigate the close relationship between widely spread financial news, rumors, and stock price changes (Bollen et al., 2011). Disinformation is also a major threat for business owners and citizens. Fake reviews are compromising the trustworthiness of the former and affecting the consumer purchase process (Valant, 2015).

The dissemination of disinformation

World Economic Forum (2013) identified the rapid distribution of disinformation through social media, as upcoming danger and one of the 10 most important trends in society. The report emphasized on the intentional nature and the difficulty of correcting disinformation, especially when it occurs within trusted networks (Arnaboldi et al., 2017; World Economic Forum, 2018).

However, disinformation is not primarily a technology-driven phenomenon. The dissemination of false information is also driven by unclear socio-psychological factors. Chadwick et al. (2018) report that those who shared tabloid news stories were more likely to share exaggerated or fabricated news. Cognitive psychologists have shown that in fact humans are only 4% better than chance (50%) to distinguish fake from real (Bond and DePaulo, 2006). In Jang and Kim (2018), researchers found that people see members of the opposite party as more vulnerable to false information than members of their party. It is also worth to mention that people accept more easily information that reflects and reinforces their prior beliefs (confirmation bias). This also known as echo-chambers (Dutton et al., 2017; Flynn et al., 2017). In addition to this popular cognition, Pennycook and Rand (2019) suggest that people fall for fake news because they fail to think. Other factors that play a role in deceiving the information consumer are emotions and repetition (Pennycook et al., 2018). Ghanem et al. (2019) showed that each type of false information has different emotional patterns. In their bestseller "Factfulness," Rosling et al. (2018) identify 10 "instincts," such as fear, urgency, and negativity, that lead people in believing false information and developing a distorted view of the world.

Structure of the paper

The remainder of the article is organized as follows. "Problem definition, scope, and methodology design" section presents the problem definition, the scope of this work, and the methodology we follow. In "Systematic literature review" section, we present the results of a systematic literature review (SLR). In "Disinformation taxonomy and categorization criteria" section, we create our disinformation typology, we identify the categorization criteria, and we link them together in a unified framework. Finally, in "Conclusion and future work" section, we present our conclusions and ideas for future research.

Problem definition, scope, and methodology design

Problem definition and scope

The term "fake news" refers to a range of information types, from low-impact, honest mistakes and satire content to high-impact manipulative techniques and malicious fabrications (HLEG, 2018). There are various definitions (e.g. Egelhofer and Lecheler, 2019) from where we conclude the absence of a universal agreement on the terminology used and the different types of false information. The definition proposed by Allcott and Gentzkow (2017) has been used in many recent studies as a navigator (Conroy et al., 2015; Potthast et al., 2018; Ruchansky et al., 2017; Shu et al., 2017; Wang et al., 2018). However, we deliberately avoid here the use of the term "fake news" as overloaded (Wardle and Derekshan, 2017) and inadequate to describe the complexity of the problem. Instead, we prefer the term "false information" as the broader concept that encompasses a wide spectrum of subtypes.

"Fake news" assumed to be inappropriate not only from a conceptual aspect but also from an etymological view. According to *the Merriam-Webster Dictionary*, the word "fake" has to do with origins and authenticity, something that is not genuine, imitation, or counterfeit, whereas "news" is defined as newly received or noteworthy information, especially about recent events. There are many cases of false information where there might be some level of facticity or examples describing past events as present, thus contradicting with the definitions of "fake" and "news." Moreover, the scope of this discussion goes beyond the "news" field. All these introduce unique attributes that should be carefully examined.

Around this terminology issue, there is a debate to broaden the discussion to include not only the analysis of the content itself but also the motivations and actions of its creators (Newman et al., 2018). Various terms have been used as hypernym alternatives, including "information pollution" (Meel and Vishwakarma, 2019; Wardle and Derekshan, 2017) and "information disorder" (Wardle and Derekshan, 2017). The following concepts found in definitions deserve our attention: the types, the elements, and the phases of false information. The three types are "misinformation," "disinformation" (HLEG, 2018), and "mal-information" (Ireton and Posetti, 2018). Elements and phases relate to dissemination mechanisms of false information, thus considered to be out of the scope here.

Having extensively studied the bibliography proposing taxonomies and typologies of false information, we identified a list of terms, often used interchangeably to describe specific types of disinformation content (Meel and Vishwakarma, 2019). Each study introduces ad hoc definitions, leading to conflicts or overlaps. For example, Amazeen and Bucy (2019), Dupuis and Williams (2019), and HLEG (2018) consider disinformation as an umbrella term in their studies, whereas Wardle and Derekshan (2017) examine it as a narrower term, adopting "information disorder" as hypernym. The lack of a unified categorization framework and vocabulary creates a fragmented news ecosystem which motivated us to compare and combine existing approaches and draft a typology. In this article, from the three above-mentioned false information types, we focus on "disinformation."

In the classification process, the categorization criteria play a central role. In several studies, some general criteria are mentioned or implied; however, in most cases, they were not explicitly attributed to specific types of false information in a coherent manner. Among the challenges, we met, was the use of different terms for describing ultimately the same types or criteria. Moreover, some taxonomies suggested typologies of disinformation with concepts that are at different granularity level. Thus, broader category types may be found at the same level with narrower concepts. Our goal toward a common effort to avoid concept fragmentation has been to define a logical, consistent, and structured way to list the types of false information.

For a complex problem like this, it is essential for scholars and professionals of different disciplines to reach a common understanding, not only on the high-level concepts but also, if possible, at the lower level of more specific terms and subcategories.

Providing a coherent and fine-grained typology could be also a contribution to readers from an educational aspect. Online information may affect people's decisions; thus, having a global perspective around the problem could contribute to avoid profound effects in real-life domains.

Our findings could also provide valuable insights in fields such as Artificial Intelligence (AI), where a systematic and consistent encoding of real-world entities and concepts is of crucial importance. The better defined is a type of disinformation, the better is the information given into a fact-checking or fake news detection system, and as result, the most accurate and comprehensible are the results produced. Today, there are many "fake news" datasets available (e.g. "Liar, Liar Pants on Fire" dataset, Wang, 2017; "Fake News Corpus"[1]), which are used to research and develop detection models, having entirely different labeling schemes. Computational models created using different conceptual schemes are not directly comparable in terms of their performance, challenging the definition of the state of the art in the field and ultimately having a negative effect to the advancement of research.

Research and methodology design

Our approach consists of two parts. Initially, we collect all types of false information in the literature, and after applying some logical preprocessing, we introduce our own typology of disinformation types coupled by a glossary. Then, we propose a novel,

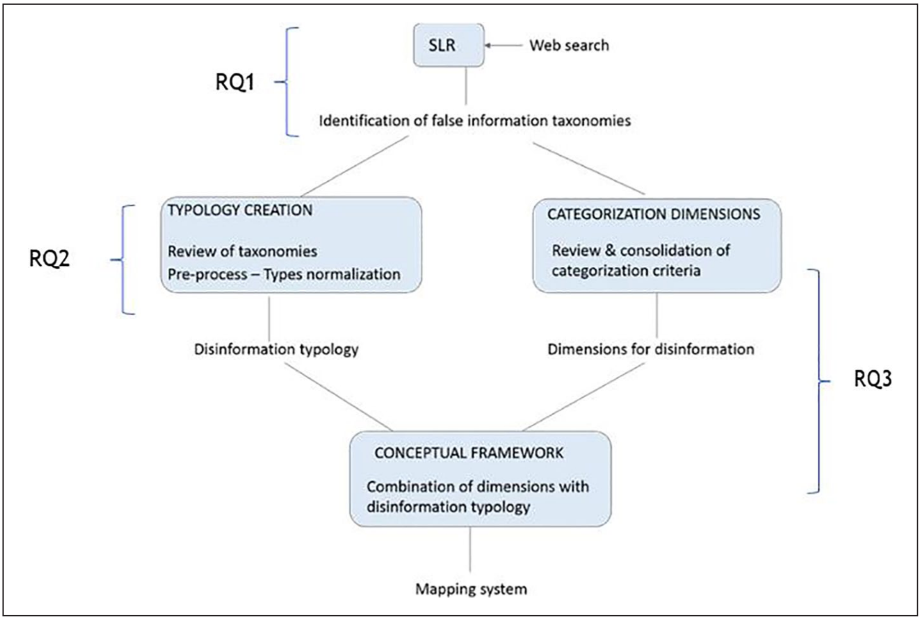

Figure 2. Our research process.

three-dimensional conceptual classification framework, based on categorization criteria found in the existing taxonomies. We define the following research questions:

RQ1: What are the existing taxonomies or typologies for false information categorization?

RQ2: Can we consolidate the taxonomies in an overarching schema and suggest a holistic typology?

RQ3: What are the categorization criteria for the existing taxonomies and which dimensions do we introduce with our typology?

Figure 2 shows an overview of our research process.

Systematic literature review

To comprehensively address RQ1, we conducted an SLR based on Kitchenham's (2007) methodological guidelines. For this research work, we considered papers published within a 4-year period (2015–2019).

The procedure we applied was the following:

1. Selection of our sources (digital libraries),
2. Definition of search terms,

3. Application of each search term on selected sources,
4. Selection of primary studies by use of inclusion and exclusion criteria on search results.

Literature review conduct and results

An automatic searching was based on the following six primary sources of scientific databases to identify relevant publications:

- ACM Digital Library
- IEEE Xplore Digital Library
- Science Direct
- SpringerLink
- Google Scholar
- Scopus

Based on our research questions, we run some pilot searches to obtain an initial list of studies. Those were then used as a basis for the systematic review to define the search terms that best fit our research questions. The search terms along with synonyms used appear below:

1. "fake news,"
2. "false news,"
3. "false information,"
4. "disinformation,"
5. "misinformation,"
6. Taxonomy OR typology OR classification,
7. Categories OR categorization,
8. Types of fake news/false news/false information/disinformation.

Inclusion and exclusion criteria

The following inclusion and exclusion criteria were defined to include papers in the next phases of our research:

CR1: We excluded sources that addressed the disinformation problem solely from a computational perspective, proposing technical approaches based on, for example, machine learning and statistical models to automatically classify news articles into predefined categories, such as fake or real (e.g. Woloszyn and Nejdl, 2018).

CR2: We excluded publications that mention types of false information without any attempt to provide systematic classification or even explanations of the proposed types. This refers to sources where either (a) the disinformation phenomenon is not a central concept (political analysis which just happens to mention terms such as "propaganda" or "hyperpartisan," medical articles mentioning "fake news" in general, etc.), or (b) they mention types of false information outside a general framework or classification model and therefore they are non-exhaustive or indicative (e.g. Campan et al., 2017; Guo et al., 2019; Pierri and Ceri, 2019; Rashkin et al., 2017; Zhou and Zafarani, 2018). Note here that although we exclude these sources as they do not meet our criteria in order to address RQ1, we do consider them for eligibility in terms of RQ2.

CR3: We included only the papers written in English.

SLR results

Our search results, including the citations from all libraries, identified eight primary studies where taxonomical frameworks were proposed (Table 1/[1]–[8]). Considering that false information has not only attracted the interest of the academic community but also of experts in various fields such as communication and journalism, as well as authorities and institutions, we decided to conduct additional research on the web, applying the same query into popular search engines. Therefore, sources that did not belong to the main scientific libraries (Google Scholar, Scopus, etc.) were examined, including national research studies, university initiatives, and international organizations reports. In this step, we identified 15 more references, two of which met our criteria (Table 1/[9] and [10]). Finally, these 10 references were assessed for eligibility in RQ2. In Figure 3, we illustrate the process of our initial search conducted in the libraries. Figure 4 presents in detail the selection process of both records found through database searching and records identified by other sources.

Data extraction

Our first goal was to identify existing taxonomies and typologies of false information (RQ1). For addressing RQ2, we aggregated the identified taxonomies, in a single table (Table 1), where each column corresponds to a reference. We then list the suggested types of false information identified and proposed per taxonomy.

Disinformation taxonomy and categorization criteria

Creation of disinformation typology

To address RQ2 and produce a typology, we had to examine the taxonomies included in Table 1 to gather and consolidate all types of false information listed there.

Review of selected taxonomies. We reviewed the taxonomies considering the more granular level of their proposed types. We observed many commonalities but also differences at both the taxonomy and type levels. Finally, five of the taxonomies were rejected for the following reasons:

1. Tambini (2017) proposes too generic categories resulting in overlaps. The proposed types describe a variety of sociopolitical phenomena, for example, “falsehood to affect election results,” “news that challenges orthodox authority,” suggesting descriptive types.

Table 1. False information taxonomies and typologies.

[1] Zannettou et al. (2019)	[2] Tambini (2017)	[3] Kumar and Shah (2018)	[4] Wardle and Derekshan (2017)	[5] Parikh and Atrey (2018)	[6] Tandoc et al. (2017)	[7] Molina et al. (2019)	[8] Lemieux and Smith (2018)	[9] Pamment et al. (2018)	[10] House of Commons (2018)
Fabricated content	Falsehood to affect election results	Misinformation	Satire	Visual based	News satire	False News	Disinformation	Fabrication	Fabrication
Propaganda	Falsehood for profit gain	Disinformation	False connection	User based	News parody	Polarized Content	Hoax	Manipulation	Manipulated content
Imposter	Bad journalism	Opinion based	Misleading content	Post based	Fabrication	Satire	Bias in Fact selection	Misappropriation	Imposter content
Conspiracy theories	Parody	Fact based	False context	Network based	Manipulation	Misreporting	Rumors	Propaganda	Misleading content
Hoaxes	Ideologically opposed news		Imposter content	Knowledge based	Advertising	Commentary	Hyperbole	Satire	False context
Biased or one-sided	News that challenges orthodox authority		Manipulated content	Stance based		Persuasive Information	Misinformation	Parody	Satire
Fallacy			Fabricated content					Advertising	Deep fakes
Rumors			Leaks						
Clickbait			Harassment						
Satire			Hate speech						

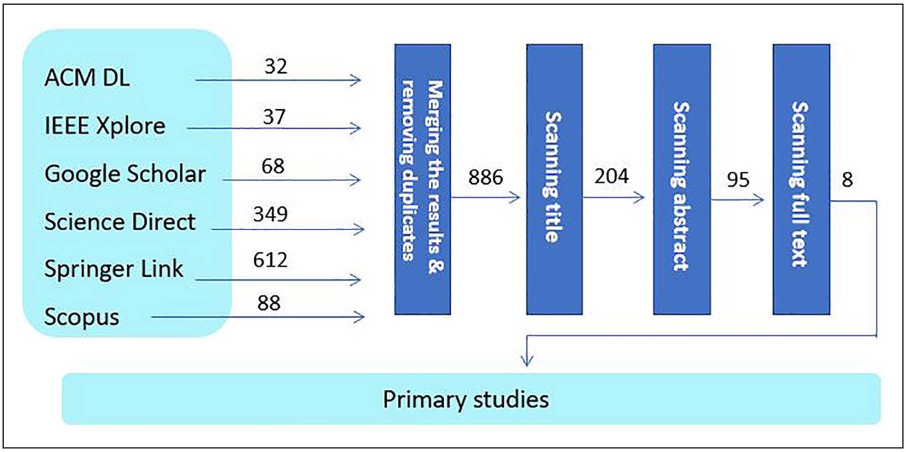

Figure 3. Primary studies selection.

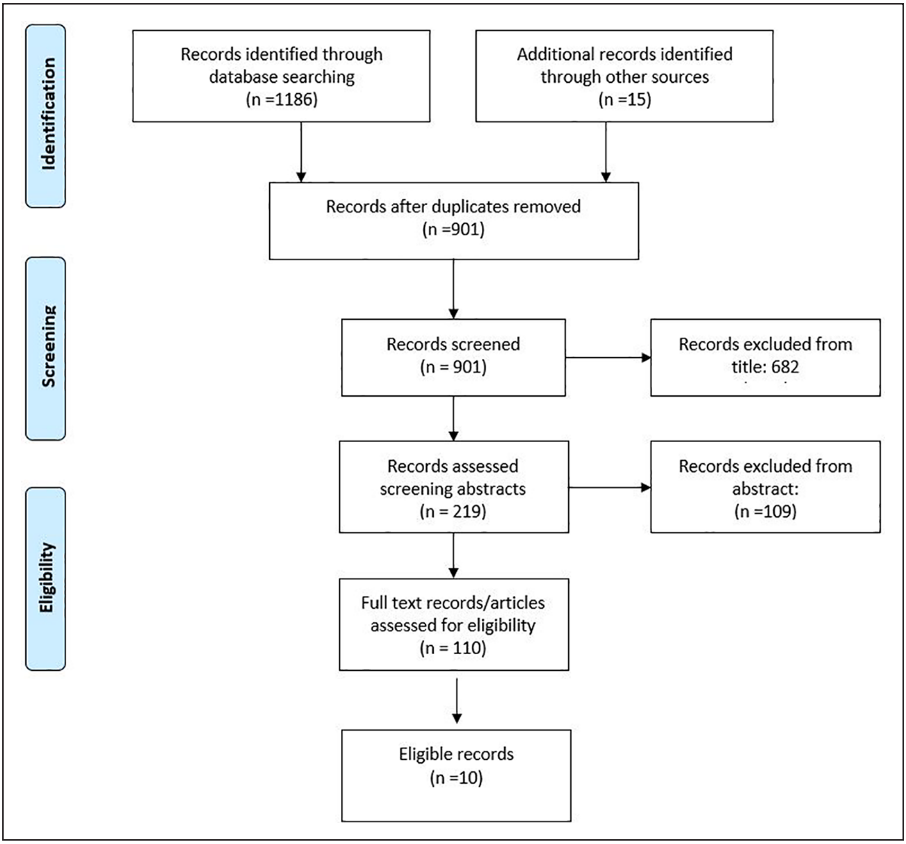

Figure 4. PRISMA flow diagram—Primary and additional records selection.

2. Kumar and Shah (2018) approach the problem from a detection perspective, introducing four general categories, that is, opinion based, fact based, misinformation, disinformation, without specializing on normalized subtypes of false information ecosystem (e.g. satire, parody, and clickbait). They have a rather narrow focus in specific domains and they place the terms disinformation/misinformation at the lowest level, whereas usually these are presented as umbrella terms.
3. Parikh and Atrey (2018) define fake news categories based on technical properties or the format of the news item, such as visual based (e.g. photoshopped images), user based (e.g. fake accounts), style based, and so on. Although this is useful for the construction of automatic detection tools, it introduces a technical perspective which makes impossible the consolidation with the other taxonomies.
4. Molina et al. (2019) discern fake news types based on four operational indicators, that is, message, source, structure, and network. They go beyond content-based approaches, concepts, and definitions focusing on the dissemination of online information and provide an analysis in terms of detection solutions.
5. Lemieux and Smith (2018) place broad categories such as disinformation and misinformation in the same level as more granular types such as hoax and rumors. They also consider mal-information as the umbrella term placed at the same level as disinformation and misinformation.

Extraction of false information types. Next, we focused on the distinct categories proposed by the remaining taxonomies. Our objective was to draft a catalog of clean and normalized terms with definitions. After thorough analysis and removal of repetitions, we list 19 different terms derived from the selected taxonomies (Table 2).

However, considering the wide variety of false information types that can be found on the web and social media, we expanded our search beyond the scientific literature. Finally, we found 20 additional types of false information (Table 3) in other sources (EAVI, 2018; Kumar and Shah, 2018; Woolley and Howard, 2018).

Data pre-process and disinformation typology. Within a total of 39 terms listed in Tables 2 and 3, we detected types that could distract us from a comprehensive categorization process. For this, we employed a two-step processing approach based on a set of logical rules illustrated in Table 9 of the Appendix and explained below. The logical rules we applied during the first stage of processing include the following:

- *Rule A*: Removal of types or definitions that are either generic and confusing (junk news) or too technical (deep fakes).
- *Rule B*: Removal of duplicates by synonym detection avoiding repetitions and overlaps.
- *Rule C*: Removal of terms that were incorrectly categorized as types of disinformation (e.g. lie or illegal content, such as "defamation").
- *Rule D*: Integration of terms and creation of normalized hypergroups.

After applying the above rules, 24 terms were rejected. The remaining 15 describe uniquely and adequately any instance of false information (see Table 4).

Table 2. Unique false information types in the literature.

Clickbait	False context	Misappropriation	Satire
Conspiracy theories	False connection	Misleading content	Advertising
Deep fakes	Biased/one-sided	Parody	Rumors
Fabrication	Imposter	Highly partisan news sites	Manipulation
Fallacy	Hoax	Propaganda	

Table 3. Additional unique false information types from other sources.

Bogus	Error	Harassment	Pseudoscience
Bullying	Fake reviews	Leaks	Urban legend
Defamation	False balance	Lie	Trolling
Disinformatzya	Forgeries	Lying by omission	Typosquatting
Doxing	Hate speech	Manufactured amplification junk news	

Table 4. False information types after the first step of preprocessing.

Clickbait	Hoax	Propaganda	Pseudoscience
Conspiracy theories	Biased/one-sided	Rumors	Trolling
Fabrication	Imposter	Satire	Urban legend
Fallacy	Parody	Fake reviews	

Disinformation typology refinement. As a second and final step of the processing phase, we further refined the identified types to include only those that refer to disinformation. Using our disinformation definition (HLEG, 2018), we exclude satire, parody, and other comedic sources (e.g. memes) because they do not satisfy the "intent to harm" condition of our working definition (HLEG, 2018) but they intent to entertain. We also exclude illegal content like hate speech and defamation as they fall into the mal-information category.

One of our biggest challenges regarding this step of our research was that not all types have the same level of deceptiveness or harmful impact, and thus, some of them could not be strictly considered as disinformation. For example, "fabrication" is more severe than "hyperpartisan" or "clickbait," creating a lot of discussions around the latter. In order to address this, we decided to thoroughly study, process, and consolidate reports found in the existing literature before we classify them as disinformation. HLEG (2018) places clickbait in the low-end spectrum of disinformation. However, the European Consumer Organization (BEUC) commented negatively the report finding unacceptable the absence of any reference " . . . to one of the major potential sources of disinformation—clickbaiting" (HLEG, 2018). According to Pamment et al. (2018), the problem is not just the use of sensational headlines to attract readers but the fact that it has evolved to something with greater impact. Chen et al. (2015) and Faris et al. (2017) consider it particularly harmful because "these stories tethered to something true but exaggerate it or misconstrue it to the point of

Table 5. Normalized disinformation types.

Disinformation typology	
Fabricated	Clickbait
Imposter	Misleading connection
Conspiracy theories	Fake reviews
Hoaxes	Trolling
Biased or one-sided	Pseudoscience
Rumors	

unrecognizability." Blom and Hansen (2015) conclude that clickbait is perhaps closer to manipulation than stimulation. Regarding the term "hyperartisan," there are several definitions in the literature that connect the term with the cases where one side is overly promoted while others are severely understated, although this term has been coined mostly with political parties. Zannettou et al. (2019) propose the more general term "biased or one-sided," which we adopt to cover all cases of extremely imbalanced reporting.

Taking the above into consideration, we finalized our first step, creating a disinformation typology. Table 5 contains the final 11 normalized types of disinformation. We also developed a glossary of definitions to support our typology (Appendix, Table 8). Figure 5 depicts the steps described above.

A unified framework for disinformation

The second part of our work focuses on the categorization criteria of our typology (RQ3).

Identification of categorization criteria. After reviewing the existing taxonomies, we identify and extract the categorization criteria from each study to select relevant and recurring, referred here as "*dimensions*." Our goal is to map them to the types proposed by our taxonomy and assign appropriate values. For the selection of dimensions, we consider three design principles:

1. *Orthogonality*. No subtype is a member of more than one group.
2. *Flexibility*. It is an essential property of dynamic taxonomy design. It ensures the integrity of taxonomy's design, allowing for future additions.
3. *Simplicity*. For our model to be compact and easily applicable, we need as few dimensions as possible, while maintaining the ability to cover all available types of disinformation.

In some models, the categorization criteria were not explicitly described but rather implicitly used by the authors, so it was not always possible to find the underlying logic. The criteria we finally extracted are summarized in Table 6.

Review of the categorization criteria—suggestion of dimensions. Before we articulate our proposed dimensions, we studied the emerged categorization criteria, challenging them to identify inaccuracies or inadequacy.

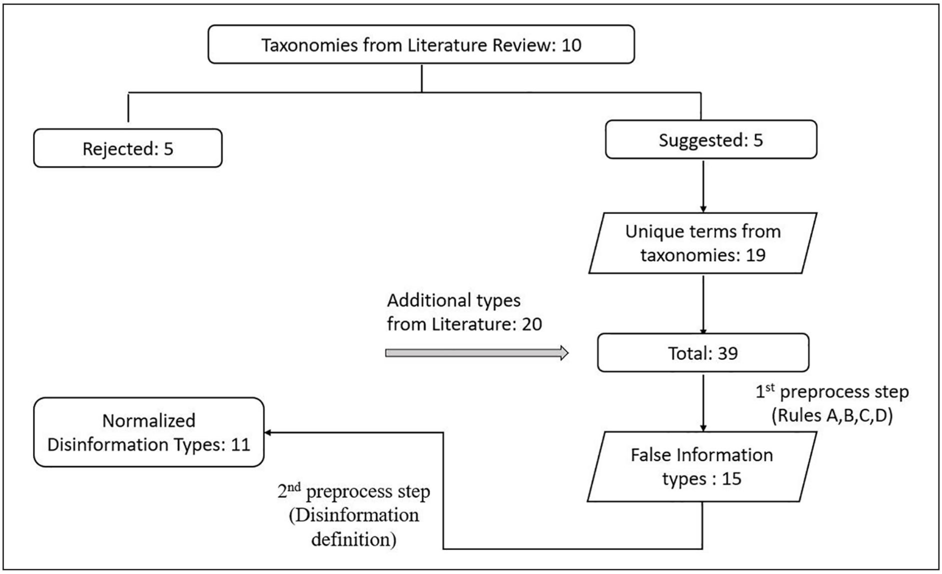

Figure 5. Preprocess analysis—Types of false information.

Table 6. Extracted and suggested dimensions with value lists.

Extracted categorization criteria	Suggested dimensions	Values
1. Facticity–Intention to deceive (Tandoc et al., 2017) 2. Facticity–Intention to deceive/mislead–Informative/ Entertaining character (Pamment et al., 2018) 4. Knowledge–Intention to deceive/mislead (Kumar and Shah, 2018) 5. Severity (Zannettou et al., 2019) 6. Falseness–Intention to harm (Wardle and Derekshan, 2017)	Motivation Facticity Verifiability	Financial–Ideological–Psychological–Unclear Mostly True–Mostly False–False Yes–No

The concepts of "facticity," "knowledge," and "falseness" are extensively used in the literature when examining the factual basis of disinformation. Facticity is defined as the degree to which news and content rely on facts (Tandoc et al., 2017). That degree may vary from entirely false (fabricated) to a mixture of facts and false context or narratives or distortion of information or images (HLEG, 2018; Tambini, 2017). In some cases, facticity is identified to accuracy (House of Commons, 2018; Tambini, 2017). We adopt *facticity* as a more appropriate term to describe this concept.

The informative or entertaining character of false information does not fall into disinformation category. Humorous content, for example, may include misleading elements (claims, videos, etc.) but the creator does not intent to harm or deceive the receiver.

Intention to deceive/mislead cannot be assessed as potential dimension as, by definition, all kind of disinformation types is created to harm or mislead the receiver of the message. During our research, we also encountered authenticity as another interesting criterion. Allcott and Gentzkow (2017) used authenticity as a potential dimension to evaluate the extent to which information can be verified. As authenticity refers to the content origins and genuineness, we introduced *verifiability* as a more appropriate term to label this dimension.

We anticipated that none of the proposed taxonomical frameworks includes all criteria and dimensions and our research verified this assumption. The models focus on the quality of the content, ignoring the creators' motivation and/or the impact that has on recipients. However, as the impact is linked to the consequences of the disinformation dissemination and not with the content itself, we considered it inappropriate for our objective. This motivated us to develop a more comprehensive classification system, incorporating motivation as an additional dimension. Although motivation and intention are similar terms that are often used interchangeably, it is worth noting that motivation refers to the driving force behind an act while intention refers to the objective. Thus, the suggested dimensions in our model include *facticity, verifiability*, and *motivation* (Table 6).

Having identified the three dimensions as the basis of our framework, we further analyze them by defining their *value range*, presented in the following section and Table 6:

- As, by definition, disinformation comes with a particular intent, qualitative subtypes were defined, including financial, ideological, or psychological purposes as separate values for the *Motivation* dimension. Other reasons for producing "polluted" messages could be political, social (Wardle and Derekshan, 2017), advertising, or humorous reasons (Tandoc et al., 2017). To stay compliant to the simplicity principle and based on their definitions, we merged the first two types into the "ideological" category. Advertisement and humor were rejected because they are related to misinformation and not disinformation. Finally, since sometimes primary motives are difficult to discern, we decided to include "unclear" as a fourth possible value for motivation.
- Facticity can be assessed using a quantitative scale, as proposed by one of the most reputed fact-checking communities, Politifact (Holan, 2018). We ended up with three possible values as defined below:
    - *Mostly true* – The statement or parts of it are accurate and contains some facts but needs clarification or additional information.
    - *Mostly false* – The statement contains an element of truth but ignores critical facts that would give a different impression.
    - *False* – The statement is not accurate.
- For the verifiability dimension, we proposed Yes/No as a simple, binary reply to the question, "Is the message easily verifiable?"

Mapping of disinformation typology to a three-dimensional framework. In the last step of our work, we combined the results into a common unified framework supported by our glossary (Appendix, Table 8). The suggested types of disinformation were mapped to the selected dimensions, as shown in Table 7.

Table 7. A unified typology framework for disinformation.

Dimensions/ measurement	Motive				Facticity			Verifiability	
	Profit	Ideological	Psychological	Unclear	Mostly true	Mostly false	False	Yes	Not
Clickbait	✓		✓		✓			✓	
Conspiracy theories		✓	✓		✓			✓	
Fabrication	✓				✓			✓	
Misleading connection	✓				✓			✓	
Hoax	✓				✓			✓	
Biased or one-sided	✓				✓			✓	
Imposter	✓				✓			✓	
Pseudoscience	✓		✓		✓			✓	
Rumors	✓				✓			✓	
Fake reviews	✓				✓			✓	
Trolling	✓				✓			✓	

Conclusion and future work

This work aims to contribute with novel insights into the fast-growing world of false information and disinformation, in a systematic and structured way. Triggered by the absence of a commonly agreed domain language, our objective was to identify and clearly define the various underlying content types in the information disorder ecosystem and organize them. We emphasize on the importance of clear and commonly accepted definitions since different disinformation types might require different theoretical analysis. A shared understanding of definitions is essential to avoid the creation of fragmented islands of counter-disinformation policies and agendas.

Diving into this complex and broad field, we met some strong challenges. First, despite the plethora of scientific studies on the field, we found that most of them introduce isolated and ad hoc approaches, resulting to a fragmentation problem. Another challenge we faced stems from the new wave of Big Data, AI, and Natural Language Processing tools, producing a large volume of research work. In most cases, the rationale and the conceptual model is not adequately explained, because the main goal in this type of research remains to propose efficient (accurate) algorithmic approaches.

Acknowledging the dynamic nature of the domain, we expect that additional types of disinformation will appear. For this reason, it is in our plans to validate the framework after, for example, 2 years to identify candidate new entries. For the remaining part of our model, which refers to the dimensions and their values, we believe our model is more future-proof, without excluding a possible revision. This temporal endurance is supported by our design principles, as well as from the fact that the proposed dimensions do not exhibit dynamic characteristics like the disinformation types.

Another aspect, we realized, that deserves attention is the need for multidisciplinary approaches in understanding and designing actions and tools to fight disinformation. Although the field has strong links with the political communication theory, we believe that modern disinformation exhibits characteristics that call for the exploitation of additional analytical tools. Disinformation is thriving in digital communities characterized by unique features not easily comparable with the past. As already identified by scholars, the scope, volume, speed, and the new communities already justify the revision of existing tools. Moreover, disinformation includes also types that go beyond the world of politics like fake reviews and pseudoscience. Last, the recent impressive progress in technologies like Machine Learning promise the development of (semi-) automated fact-checking tools. This is yet another call for multidisciplinary research on the field.

Authors' note

All authors have agreed to the submission, and the article is not currently being considered for publication by any other print or electronic journal.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement no. 770302, project title: Co-Inform: Co-Creating Misinformation-Resilient Societies.

ORCID iDs

Eleni Kapantai https://orcid.org/0000-0001-9943-6389
Christos Berberidis https://orcid.org/0000-0002-3938-7076

Note

1. https://github.com/several27/FakeNewsCorpus

References

Albarracin D, Romer D, Jones C, et al. (2018) Misleading claims about tobacco products in YouTube videos: experimental effects of misinformation on unhealthy attitudes. *Journal of Medical Internet Research* 20(6): e229.

Allcott H and Gentzkow M (2017) Social media and fake news in the 2016 election. *Journal of Economic Perspectives* 31(2): 211–236.

Amazeen M and Bucy E (2019) Conferring resistance to digital disinformation: the inoculating influence of procedural news knowledge. *Journal of Broadcasting & Electronic Media* 63(3): 415–432.

Arnaboldi V, Conti M, Passarella A, et al. (2017) Online Social Networks and information diffusion: the role of ego networks. *Online Social Networks and Media* 1: 44–55.

Aisch G, Huang J and Kang C (2016) Dissecting the #PizzaGate conspiracy theories. *Nytimes.com*. Available at: https://www.nytimes.com/interactive/2016/12/10/business/media/pizzagate.html (accessed 12 September 2020).

Bad News Game (2017) Available at: https://getbadnews.com/#intro (accessed 5 April 2019).

Benkler Y, Faris R and Roberts H (2018) *Network Propaganda*. New York: Oxford University Press.

Bennett W and Pfetsch B (2018) Rethinking political communication in a time of disrupted public spheres. *Journal of Communication* 68(2): 243–253.

Blom J and Hansen K (2015) Click bait: forward-reference as lure in online news headlines. *Journal of Pragmatics* 76: 87–100.

Blumler J (2015) Core theories of political communication: foundational and freshly minted. *Communication Theory* 25(4): 426–438.

Bollen J, Mao H and Zeng X (2011) Twitter mood predicts the stock market. *Journal of Computational Science* 2(1): 1–8.

Bond C and DePaulo B (2006) Accuracy of deception judgments. *Personality and Social Psychology Review* 10(3): 214–234.

Bursztyn L, Petrova M, Enikolopov R, et al. (2018) Social media and xenophobia: evidence from Russia. *The American Economic Association's RCT Registry*. Available at: https://doi.org/10.1257/rct.3066-1.0 (accessed 5 April 2019).

Campan A, Cuzzocrea A and Truta T (2017) Fighting fake news spread in online social networks: actual trends and future research directions. In: *2017 IEEE international conference on big data (big data)*, Boston, MA, 11–14 December, pp. 4453–4457. New York: IEEE.

Chadwick A, Vaccari C and O'Loughlin B (2018) Do tabloids poison the well of social media? Explaining democratically dysfunctional news sharing. *New Media & Society* 20(11): 4255–4274.

Chee F (2020) Combat 5G COVID-19 fake news, urges Europe. *US*. Available at: https://www.reuters.com/article/us-eu-telecoms-5g/combat-5g-covid-19-fake-news-urges-europe-idUSKBN2392N8 (accessed 30 July 2020).

Chen Y, Conroy N and Rubin V (2015) Misleading online content. In: *WMDD'15 proceedings of the 2015 ACM on workshop on multimodal deception detection*, Seattle, WA, 13 November.

Conroy N, Rubin V and Chen Y (2015) Automatic deception detection: methods for finding fake news. *Proceedings of the Association for Information Science and Technology* 52(1): 1–4.

CTED (2020) Member states concerned by the growing and increasingly transnational threat of extreme right-wing terrorism (2020) Trends Alert, Counter-Terrorism Committee Executive Directorate. Available at: https://www.un.org/sc/ctc/wp-content/uploads/2020/07/CTED_Trends_Alert_Extreme_Right-Wing_Terrorism_JULY.pdf (accessed 30 July 2020).

Dupuis M and Williams A (2019) The spread of disinformation on the web: an examination of memes on social networking. In: *2019 IEEE SmartWorld, ubiquitous intelligence & computing, advanced & trusted computing, scalable computing & communications, cloud & big data computing, Internet of people and smart city innovation*, Leicester, 19–23 August, pp. 1412–1418. New York: IEEE.

Dutton W, Reisdorf B, Dubois E, et al. (2017) *Social shaping of the politics of internet search and networking: moving beyond filter bubbles, echo chambers, and fake news*. Quello Center Working Paper No. 2944191. DOI: 10.2139/ssrn.2944191.

EAVI (2018) Infographic: beyond fake news—10 types of misleading news—eleven languages. Available at: https://eavi.eu/beyond-fake-news-10-types-misleading-info/ (accessed 5 April 2019).

Egelhofer J and Lecheler S (2019) Fake news as a two-dimensional phenomenon: a framework and research agenda. *Annals of the International Communication Association* 43(2): 97–116.

European Commission (2018a) Fake news and online disinformation. *Digital Single Market-Policy*. Available at: https://ec.europa.eu/digital-single-market/en/fake-news-disinformation

European Commission (2018b) *Tackling Online Disinformation: A European Approach*. Brussels. Available at: https://ec.europa.eu/digital-single-market/en/news/report-implementation-communication-tackling-online-disinformation-european-approach

Faris R, Roberts H, Etling B, et al. (2017) *Partisanship, Propaganda, and Disinformation: Online Media and the 2016 U.S. Presidential Election*. Cambridge, MA: *Berkman Klein Center Research Publication, Harvard University*.

Flynn D, Nyhan B and Reifler J (2017) The nature and origins of misperceptions: understanding false and unsupported beliefs about politics. *Political Psychology* 38: 127–150.

Forstorp PA (2005) The construction of pseudo-science: science patrolling and knowledge policing by academic prefects and weeders. *VEST: Journal of Science & Technology Studies* 18(3–4): 59–60.

Ghanem B, Rosso P and Rangel F (2019) An emotional analysis of false information in social media and news articles. Available at: https://arxiv.org/pdf/1908.09951.pdf [cs.CL] 26 August 2019 (accessed 4 December 2019).

Guacho GB, Abdali S, Shah N, et al. (2018) Semi-supervised content-based detection of misinformation via tensor embeddings. In: *2018 IEEE/ACM international conference on advances in social networks analysis and mining (ASONAM)*, Barcelona, 28–31 August, pp. 322–325. New York: IEEE.

Guo B, Ding Y, Yao L, et al. (2019) The future of misinformation detection: new perspectives and trends. Available at: https://arxiv.org/pdf/1909.03654.pdf (accessed 8 December 2019).

HLEG (2018) A multi-dimensional approach to disinformation: report of the independent high-level group (HLEG) on fake news and online disinformation. *European Commission. Publications Office of the European Union*. Available at: https://blog.wan-ifra.org/sites/default/files/field_blog_entry_file/HLEGReportonFakeNewsandOnlineDisinformation.pdf

Holan A (2018) The principles of the truth-o-meter: how we fact-check. Available at: https://www.politifact.com/truth-o-meter/article/2018/feb/12/principles-truth-o-meter-politifacts-methodology-i/ (accessed 5 April 2019).

Hotten R (2015) Volkswagen: the scandal explained. Available at: https://www.bbc.com/news/business-34324772 (accessed 5 April 2019).

House of Commons (2018) Disinformation and "fake news": Interim report, fifth report of session 2017-2019. *UK Parliament*. Available at: https://publications.parliament.uk/pa/cm201719/cmselect/cmcumeds/363/363.pdf (accessed 5 April 2019).

IFCN (2017) IFCN code of principles report 2018. *International Fact-Checking Network*. Available at: https://ifcncodeofprinciples.poynter.org/storage/docs/PUBLIC_VERSION-CODE_OF_PRINCIPLES_REPORT_YEAR_1_REV_AM.pdf?v=1538242914

InVID Project (2017) Web Lyzard technology. Available at: https://invid.weblyzard.com/ (accessed 10 May 2019).

Ireton C and Posetti J (2018) *Journalism, "Fake News" & Disinformation*. Paris: UNESCO.

Jang S and Kim J (2018) Third person effects of fake news: fake news regulation and media literacy interventions. *Computers in Human Behavior* 80: 295–302.

Jolley D and Douglas K (2014) The effects of anti-vaccine conspiracy theories on vaccination intentions. *PLoS ONE* 9(2): e89177.

Kazeem Y (2018) Nigerian media houses are forming a coalition to combat fake news ahead of next year's elections. *Quartz Africa*. Available at: https://qz.com/africa/1478737/fake-news-media-collaborate-ahead-of-nigeria-2019-election/ (accessed 5 April 2019).

King's College and Ipsos MORI (2018) Brexit misperceptions. Report, London. Available at: https://ukandeu.ac.uk/wp-content/uploads/2018/10/Brexit-misperceptions.pdf

Kitchenham BA (2007) Guidelines for performing systematic literature reviews in software engineering version 2.3. *EBSE Technical Report, Keele University and University of Durham, Durham, 9 July*.

Kumar S and Shah N (2018) False information on web and social media: a survey. *arXiv preprint* arXiv:1804.08559.

Kušen E and Strembeck M (2018) Politics, sentiments, and misinformation: an analysis of the Twitter discussion on the 2016 Austrian Presidential Elections. *Online Social Networks and Media—Journal* 5: 37–50.

Lazer D, Baum M, Benkler Y, et al. (2018) The science of fake news. *Science* 359(6380): 1094–1096.

Lemieux V and Smith T (2018) Leveraging archival theory to develop a taxonomy of online disinformation. In: *2018 IEEE international conference on big data (big data)*, Los Angeles, CA, 9–12 December.

Lynas M (2020) COVID: top 10 current conspiracy theories—alliance for science. *Alliance for Science*. Available at: https://allianceforscience.cornell.edu/blog/2020/04/covid-top-10-current-conspiracy-theories/ (accessed 14 August 2020).

Meel P and Vishwakarma D (2019) Fake news, rumor, information pollution in social media and web: a contemporary survey of state-of-the-arts, challenges and opportunities. *Expert Systems with Applications* 153: 112986.

Molina M, Sundar S, Le T, et al. (2019) "Fake news" is not simply false information: a concept explication and taxonomy of online content. *American Behavioral Scientist*. Epub ahead of print 14 October. DOI: 10.1177/0002764219878224.

Müller K and Schwarz C (2017) Fanning the flames of hate: social media and hate crime. *SSRN Electronic Journal*. Available at: http://dx.doi.org/10.2139/ssrn.3082972

Müller K and Schwarz C (2018) Making America hate again? Twitter and hate crime under Trump. *SSRN Electronic Journal*. Available at: https://ssrn.com/abstract=3082972

Newman N, Fletcher R, Kalogeropoulos A, et al. (2018) Reuters institute digital news report 2018. *Reuters Institute for the Study of Journalism*. Available at: http://media.digitalnewsreport.org/wp-content/uploads/2018/06/digital-news-report-2018.pdf?x89475 (accessed 22 June 2019).

Niklewicz K (2017) Weeding out fake news: an approach to social media regulation. *European View* 16(2): 335–335.

Pamment J, Nothhaft H, Agardh-Twetman H, et al. (2018) Countering information influence activities: the state of the art. *Department of Strategic Communication, Lund University*. Available at: https://www.msb.se/RibData/Filer/pdf/28697.pdf (accessed 5 April 2019).

Parikh SB and Atrey PK (2018) Media-rich fake news detection: a survey. In: *2018 IEEE conference on multimedia information processing and retrieval (MIPR)*, Miami, FL, 10–12 April, pp. 436–441. New York: IEEE.

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

Pennycook G, Cannon T and Rand D (2018) Prior exposure increases perceived accuracy of fake news. *Journal of Experimental Psychology: General* 147(12): 1865–1880.

Peterson W and Gist N (1951) Rumor and public opinion. *American Journal of Sociology* 57(2): 159–167.

Pierri F and Ceri S (2019) False news on social media. *ACM SIGMOD Record* 48(2): 18–27.

Politifact (2007) Available at: https://www.politifact.com/ (accessed 5 April 2019).

Potthast M, Kiesel J, Reinartz K, et al. (2018) A stylometric inquiry into hyperpartisan and fake news. In: *Proceedings of the 56th annual meeting of the association for computational linguistics (Volume 1: Long Papers)*, Melbourne, VIC, Australia, 15–20 July.

Pujahari A and Sisodia D (2019) Clickbait detection using multiple categorisation techniques. *Journal of Information Science* 6(2): 137–153.

Rashkin H, Choi E, Jang J, et al. (2017) Truth of varying shades: analyzing language in fake news and political fact-checking. In: *Proceedings of the conference on empirical methods in natural language processing*, Copenhagen, Denmark, 7–11 September, pp. 2931–2937. Association for Computational Linguistics. DOI: 10.18653/v1/D17-1317.

Rehm G (2018) An infrastructure for empowering internet users to handle fake news and other online media phenomena. *Lecture Notes in Computer Science* 10713: 216–231.

Renda A (2018) The legal framework to address "fake news": possible policy actions at the EU level. European Parliament. CEPS Research Report.

Rojecki A and Meraz S (2016) Rumors and factitious informational blends: the role of the web in speculative politics. *New Media & Society* 18(1): 25–43.

Rosling H, Rosling O and Rönnlund AR (2018) *Factfulness: Ten Reasons We're Wrong about the World—And Why Things Are Better Than You Think*. New York: Flatiron Books.

Rubin V, Chen Y and Conroy N (2015) Deception detection for news: three types of fakes. *Proceedings of the Association for Information Science and Technology* 52(1): 1–4.

Ruchansky N, Seo S and Liu Y (2017) CSI: a hybrid deep model for fake news detection. In: *Proceedings of the 2017 conference on information and knowledge management*, Singapore, 6–10 November, pp. 797–806. New York: ACM.

Serhan Y (2018) Italy scrambles to fight misinformation ahead of its elections. *The Atlantic*, 24. Available at: https://www.theatlantic.com/international/archive/2018/02/europe-fake-news/551972/ (accessed 5 April 2019).

Shao C, Ciampaglia G, Varol O, et al. (2018) The spread of low-credibility content by social bots. *Nature Communications* 9: 4787.

Shu K, Sliva A, Wang S, et al. (2017) Fake news detection on social media: a data mining perspective. *ACM SIGKDD Explorations Newsletter* 19(1): 22–36.

Silverman C (2016) This analysis shows how viral fake election news stories outperformed real news on Facebook. *Buzzfeed News*. Available at: https://www.buzzfeednews.com/article/craigsilverman/viral-fake-election-news-outperformed-real-news-on-facebook (accessed 5 April 2019).

Snopes (1994) Available at: http://www.snopes.com/ (accessed 5 April 2019).

Starbird K, Maddock J, Orand M, et al. (2014) Rumors, false flags, and digital vigilantes: misinformation on twitter after the 2013 Boston marathon bombing. In: *iConference 2014 proceedings*, Berlin, Germany, 4–7 March, pp. 654–662. DOI: 10.9776/14308.

Syed-Abdul S, Fernandez-Luque L, Jian W, et al. (2013) Misleading health-related information promoted through video-based social media: anorexia on YouTube. *Journal of Medical Internet Research* 15(2): e30.

Szpakowski M (2018) FakeNewsCorpus. Available at: https://github.com/several27/FakeNewsCorpus/blob/master/README.md (accessed 5 April 2019).

Tambini D (2017) *Fake News: Public Policy Responses. Media Policy Brief 20. London: Media Policy Project, London School of Economics and Political Science*.

Tandoc E, Lim Z and Ling R (2017) Defining "fake news." *Digital Journalism* 6(2): 137–153.

Valant J (2015) Online consumer reviews: the case of misleading or fake reviews. *European Parliamentary Research Service*. Available at: https://www.europarl.europa.eu/thinktank/en/document.html?reference=EPRS_BRI(2015)571301#:~:text=Some%20European%20consumer%20organisations%20say,market%2C%20which%20can%20reduce%20competition

Vosoughi S, Roy D and Aral S (2018) The spread of true and false news online. *Science* 359(6380): 1146–1151.

Wallner C and White J (2020) The far-right and coronavirus: extreme voices amplified by the global crisis. *RUSI*. Available at: https://rusi.org/commentary/far-right-and-coronavirus-extreme-voices-amplified-global-crisis (accessed 30 July 2020).

Wang W (2017) "Liar, liar pants on fire": a new benchmark dataset for fake news detection. In: *Proceedings of the 55th annual meeting of the association for computational linguistics (Volume 2: Short Papers)*, Vancouver, BC, Canada, 30 July–4 August.

Wang Y, Ma F, Jin Z, et al. (2018) EANN: event adversarial neural networks for multimodal fake news detection. In: *KDD'18: Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining*, London, 19–23 August.

Wang Y, McKee M, Torbica A, et al. (2019) Systematic literature review on the spread of health-related misinformation on social media. *Social Science & Medicine* 240: 112552.

Ward B (2018) Another failure to tackle fake news about climate change—Grantham Research Institute on climate change and the environment. Available at: http://www.lse.ac.uk/GranthamInstitute/news/another-failure-to-tackle-fake-news-about-climate-change/ (accessed 5 April 2019).

Wardle C and Derekshan H (2017) Information disorder: toward an interdisciplinary framework for research and policy making. Council of Europe report DGI 09, 31 October. Brussels: Council of Europe.

Wardle C, Greason G, Kerwin J, et al. (2018) Information disorder: the essential glossary. *First Draft*. Available at: https://firstdraftnews.org/wp-content/uploads/2018/07/infoDisorder_glossary.pdf (accessed 5 April 2019).

Wendling M (2018) The (almost) complete history of "fake news." Available at: https://www.bbc.com/news/blogs-trending-42724320 (accessed 5 April 2019).

Woloszyn V and Nejdl W (2018) DistrustRank. In: *Websci'18: Proceedings of the 10th ACM conference on web science*, Amsterdam, 27–30, May.

Woolley S and Howard P (2018) *Computational Propaganda*. Oxford: Oxford University Press.

World Economic Forum (2013) *Global Risks Report*. 8th ed. Available at: http://www3.weforum.org/docs/WEF_GlobalRisks_Report_2013.pdf

World Economic Forum (2018) *Global Risks Report*. 13th ed. Geneva, pp. 48–50. Available at: http://www3.weforum.org/docs/WEF_GRR18_Report.pdf

Yaron O (2018) How fake news is threatening the upcoming elections in Israel. *HAARETZ*. Available at: https://www.haaretz.com/israel-news/.premium.MAGAZINE-the-online-dangers-threatening-fair-elections-in-israel-1.6455522 (accessed 5 April 2019).

Zannettou S, Sirivianos M, Blackburn J, et al. (2019) The web of false information. *Journal of Data and Information Quality* 11(3): 10.

Zhou X and Zafarani R (2018) Fake news: a survey of research, detection methods, and opportunities. Available at: https://arxiv.org/pdf/1812.00315.pdf (accessed 8 December 2019).

Author biographies

Eleni Kapantai is a PhD candidate student on Machine Learning and Natural Language Processing. She received the B.Sc. degree in Applied Informatics from University of Macedonia, Greece, and the M.Sc. degree in e-Business and Digital Marketing from International Hellenic University, Greece. Her research activity in terms of the bachelor thesis focused on machine learning and prediction in betting markets. She has also conducted research in Natural Language Processing, working on sentiment analysis in respect of her master thesis on topic: "Software for monitoring current trends on Twitter." Currently, she is researching on disinformation detection examining both the theoretical and technical aspects and challenges of the problem.

Androniki Christopoulou is currently a researcher at the International Hellenic University, having a bachelor's degree in Computer Science from Aristotle University and an M.Sc. in e-Business and Digital Marketing. For her thesis topic "A study on the role of Social Media, the Initiatives to tackle disinformation and a Systematic Literature Review of False Information Taxonomies," she received a scholarship from Co-Inform, an EU (European Union)-funded project involving top universities and small and medium-sized enterprises (SMEs) in seven European countries.

Christos Berberidis works as a Lab Teaching Faculty at the International Hellenic University in Greece. His work and research focus on the areas of Machine Learning and Natural Language Processing. He has worked as a postdoc associate and in many national and international R&D projects in public and private organizations for more than 20 years. He has published 18 scientific papers in refereed journals and conferences and his work has received more than 600 citations.

Vassilios Peristeras works for the European Commission, DG Informatics, Brussels, and is also Assistant Professor at the International Hellenic University in Greece. His work, teaching, and research focus on the areas of data-driven organizations, eGovernment, interoperability, open and linked data. He has worked as researcher and consultant in various organizations for more than 20 years and has initiated and coordinated several international R&D projects. He has published over 100 scientific papers and has served as editor, program committee member, and reviewer in more than 60 journals, books, and conferences.

Appendix

Table 8. Disinformation typology.

No.	Type	Definition
1	Fabricated	Stories that completely lack of any factual base, 100% false. The intention is to deceive and cause harm (Wardle and Derekshan, 2017). One of the most severe types (Zannettou et al., 2018) as fabrication adopts the style of news articles so the recipients believe it is legitimate (Tandoc et al., 2017). Could be text but also in visual format (Ireton and Posetti, 2018).
2	Imposter	Genuine sources that are impersonated with false, made-up sources to support basically a false narrative. It is actually very misleading since source or author is considered great criteria of verifying credibility (House Of Commons, 2018; Zannettou et al., 2018; Wardle and Derekshan, 2017). (use of journalists name/logo /branding of mimic URLs)
3	Conspiracy theories	Stories without factual base as there is no established baseline for truth. They usually explain important events as secret plots by government or powerful individuals (Zannettou et al., 2018). Conspiracies are, by definition, difficult to verify as true or false, and they are typically originated by people who believe them to be true (Allcott and Gentzkow, 2017). Evidences that refute the conspiracy are regarded as further proof of the conspiracy (EAVI, 2018). Some conspiracy theories may have damaging ripple-effects.
4	Hoaxes	Relatively complex and large-scale fabrications which may include deceptions that go beyond the scope of fun or scam and cause material loss or harm to the victim (Rubin et al., 2015). They contain facts that are either false or inaccurate and are presented as legitimate facts. This category is also known in the research community either as half-truth or factoid stories (Zannettou et al., 2018) able to convince readers of the validity of a paranoia-fueled story (Rashkin et al., 2017).
5	Biased or one-sided	Stories that are extremely biased toward a person/party/situation/event driving division and polarization. The context of this type of news information is extremely imbalanced (i.e. left or right wing), inflammatory, emotional and often riddled with untruths. They contain either a mixture of true and false or mostly false, thus misleading information designed to confirm a particular ideological view (Zannettou et al., 2018; Potthast et al., 2018).
6	Rumors	Refers to stories whose truthfulness is ambiguous or never confirmed (gossip, innuendo, unverified claims). This kind of false information is widely propagated on online social networks (Peterson and Gist, 1951).
7	Clickbait	Sources that provide generally credible or dubious factual content but deliberately use exaggerated, misleading, and unverified headlines and thumbnails (Rehm, 2018; Szpakowski, 2018) to lure readers open the intended Web page (Ghanem et al., 2019). The goal is to increase their traffic for profit, popularity, or sensationalization (Pujahari and Sisodia, 2019; Zannettou et al., 2018). Once the reader is there, the content rarely satisfies their interest (EAVI, 2018).
8	Misleading connection	Misleading use of information to frame an issue or individual. When headlines, visuals, or captions do not support the content. Separate parts of source information may be factual but are presented using wrong connection (context/content).
9	Fake reviews	Any (positive, neutral, or negative) review that is not an actual consumer's honest and impartial opinion or that does not reflect a consumer's genuine experience of a product, service or business (Valant, 2015).
10	Trolling	The act of deliberately posting offensive or inflammatory content to an online community with the intent of provoking readers or disrupting conversation. Today, the term "troll" is most often used to refer to any person harassing or insulting others online (Wardle et al., 2018).
11	Pseudoscience	Information that misrepresents real scientific studies with dubious or false claims. Often contradicts experts (EAVI, 2018). Promotes metaphysics, naturalistic fallacies, and other (Guacho et al., 2018). The actors hijack scientific legitimacy for profit or fame (Forstrop, 2005).

Table 9. Typology preprocessing.

Unique types from taxonomies	First phase of preprocessing	Second phase of preprocessing (proposed typology)
Clickbait	Clickbait	Clickbait
Conspiracy theories	Conspiracy theories	Conspiracy theories
Deep fakes	Eliminated (Rule A)	
Fabrication	Fabrication	Fabrication
Fallacy	Fallacy	
False connection	Eliminated (Rule D)	Misleading connection
False context	Eliminated (Rule D)	
Hoax	Hoax	Hoax
Biased/one-sided	Biased/one-sided	Biased or one-sided
Imposter	Imposter	Imposter
Manipulation	Eliminated (Rule A)	
Misappropriation	Eliminated (Rule B [Similar to Manipulation])	
Misleading content	Eliminated (Rule D)	
Parody	Parody	
Highly partisan news sites	Eliminated (Rule B [Similar to Hyperpartisan])	
Propaganda	Propaganda	
Rumors	Rumors	Rumors
Satire	Satire	
Advertising	Eliminated (Rule C [advertising is not a false information type, clickbait is])	
Additional types from literature		
Bogus	Eliminated (Rule A)	
Bullying	Eliminated (Rule C)	
Defamation	Eliminated (Rule C)	
Disinformatzya	Eliminated (Rule C)	
Doxing	Eliminated (Rule A)	
Error	Eliminated (Rule A)	
Fake reviews	Fake reviews	Fake reviews
False balance	Eliminated (Rule B)	
Forgeries	Eliminated (Rule C)	
Hate speech	Eliminated (Rule C)	
Harassment	Eliminated (Rule C)	
Junk news	Eliminated (Rule A)	
Leaks	Eliminated (Rule C)	
Lie	Eliminated (Rule C)	
Lying by omission	Eliminated (Rule A)	
Manufactured amplification	Eliminated (Rule A)	
Pseudoscience	Pseudoscience	Pseudoscience
Trolling	Trolling	Trolling
Typosquatting	Eliminated (Rule A)	
Urban legend	Urban legend	