Humanities & Social Sciences Communications

ARTICLE

https://doi.org/10.1057/s41599-020-00702-9 OPEN

A value-driven approach to addressing misinformation in social media

Nadejda Komendantova[1✉], Love Ekenberg[1,2], Mattias Svahn[2], Aron Larsson[2,3], Syed Iftikhar Hussain Shah[4], Myrsini Glinos[2], Vasilis Koulolias[2] & Mats Danielson[1,2]

Misinformation in social media is an actual and contested policy problem given its outreach and the variety of stakeholders involved. In particular, increased social media use makes the spread of misinformation almost universal. Here we demonstrate a framework for evaluating tools for detecting misinformation using a preference elicitation approach, as well as an integrated decision analytic process for evaluating desirable features of systems for combatting misinformation. The framework was tested in three countries (Austria, Greece, and Sweden) with three groups of stakeholders (policymakers, journalists, and citizens). Multi-criteria decision analysis was the methodological basis for the research. The results showed that participants prioritised information regarding the actors behind the distribution of misinformation and tracing the life cycle of misinformative posts. Another important criterion was whether someone intended to delude others, which shows a preference for trust, accountability, and quality in, for instance, journalism. Also, how misinformation travels is important. However, all criteria that involved active contributions to dealing with misinformation were ranked low in importance, which shows that participants may not have felt personally involved enough in the subject or situation. The results also show differences in preferences for tools that are influenced by cultural background and that might be considered in the further development of tools.

[1] International Institute for Applied Systems Analysis (IIASA), Vienna, Austria. [2] Stockholm University, Stockholm, Sweden. [3] Mid Sweden University, Sundsvall, Sweden. [4] International Hellenic University, Thermi, Greece. ✉email: komendan@iiasa.ac.at

Introduction

Misinformation in social media is currently attracting a lot of attention. Misinformation is not a new phenomenon and has probably existed since the dawn of humanity. Structural evidence of scientific research on misinformation can be found Allport and Postman's (1946) basic law of rumour, which demonstrates that the strength of a rumour is dependent on the importance of the subject and individual concerns regarding it, as well as the time and ambiguity of the evidence on the topic. New technical capabilities, such as social media, have naturally made these effects more universal. The 2000s witnessed rapid developments in social media and its increased outreach to everybody with Internet access. This has facilitated the spread of information, including misinformation and rumours, in virtually everything from local neighbourhoods to global concerns (Del Vicario et al., 2016).

Until recently, there has been limited scientific evidence on how to deal with misinformation, but research on the topic has increased rapidly over the past few years. For instance, researchers have suggested various ways of dealing with citizen awareness, such as nudging, as a way of vaccinating social media users against misinformation (Piccolo et al., 2019). Other topics studied include nudging for accuracy in sharing on social media (Pennycook et al., 2020) and the limits of human cognition in dealing with and spreading misinformation. Finally, researchers have examined a variety of approaches for making fact-checking more efficient, such as automatic detection of misinformation and correction of data, while at the same time pointing out the importance of human fact-checkers, as fully automated fact-checking methods are not yet strong enough.[1]

This systemic problem requires stakeholder involvement at different levels, as misinformation is so widespread and constantly changing. Extensive stakeholder involvement is necessary for designing policies, methods, and tools. However, existing approaches to developing online tools tend to follow the traditional path of dissemination of knowledge from science to stakeholders while viewing technology users as passive consumers of finished products rather than active co-creators. This is particularly alarming today when available anti-misinformation products and tools are still new to the mass market and hence malleable, which is rare in the life cycle of a product (Smith and Medin, 1981; Svahn and Lange, 2009). Value-based software engineering (Boehm, 2003) is an emerging approach that aims to develop software tools (e.g., the tool by Aurum and Wohlin, 2007) based on the values and objectives of various stakeholder groups (Biffl et al., 2006), providing an economic categorisation of the value concept based on the monetary exchange between a customer and a provider.

In this study, we investigate two major research questions:

- What are preferences for, perceptions of, and views of the features of tools for dealing with misinformation?
- How do these preferences depend on the cultural backgrounds of stakeholder groups and participants?

Our goal is to study the preferences of various stakeholder groups for features of tools, to study the impact of cultural background on these preferences, and to develop recommendations for considering these preferences in the further development of tools for dealing with misinformation.

The next section provides a background of misinformation and discusses why we need automatic tools in a general setting. Section 'Methodology' describes the integrated methodology used, and Section 'Results' presents the results and a discussion. Finally, Section 'Conclusions' concludes the article.

Background

A variety of definitions exist for misinformation, disinformation, fake news, rumours, and similar terms, and a large number of them emphasise the distinctions between misinformation and disinformation, as well as between disinformation and fake news. A review of the 2016 Presidential election in the United States, for instance, identified six different types of misinformation: authentic material used in the wrong context, imposter news sites designed to look like known brands, fake news sites, fake information, manipulated content, and parody content (Wardle, 2016). Wardle and Derakhshan (2017) suggested that misinformation refers to misleading information created without the intent to harm, whereas disinformation refers to information deliberately fabricated with the intent to impact social groups or societies. Burgoon et al. (2003) discussed misinformation in terms of deceptive language and false context. Farrel et al. (2018) distinguished between disinformation and misinformation, considering both subsets of misinformation: Disinformation largely involves the intent to deceive, whereas misinformation does not need to involve intentional deception. Giglietto et al. (2016) proposed a taxonomy based on perceptions of the source, the story, and the context and decisions of the audience and the propagator. In their taxonomy, there is "pure disinformation" when both the original author and the propagator are aware of the false nature of information but nevertheless decide to share it. There is "misinformation propagated through disinformation" when information is originally produced as true and then shared by a propagator who thinks it is false. There is also "disinformation propagated through misinformation" when information is devised as false by a creator but is perceived as true by a propagator.

Irrespective of such distinctions, both misinformation and disinformation impact the public debate on issues such as health and science (e.g., the anti-vaccine movement), foreign policy (e.g., the wars in Iraq and Ukraine), migration, elections and so on. Recognising this, researchers from a variety of disciplines, including social sciences such as journalism (Ekström et al., 2019) and psychology (Ecker, 2017), have examined misinformation and disinformation. The problems of misinformation and disinformation are usually called "wicked problems" by design scientists, as no single comprehensive solution is capable of fully resolving them and attempts to mitigate them often can make them worse. Some examples of this include the backfire effect (Nyhan and Reifler, 2010), false misinformation warnings (Freeze et al., 2020), and the naiveté of social engineering in technology (Tromble and McGregor, 2019). Misinformation and disinformation are also studied with regard to social psychology (e.g., people's values, beliefs, information literacy, and motivations), regulatory and technical perspectives (social media, detection tools), and the practice of fact-checking. Given the large volume of published work we rely here on Vanenzuala et al. (2019), who conducted a meta-analysis of 650 articles on this topic to identify regulatory, technical, and normative aspects of misinformation.

Cognitive psychologists have investigated the effectiveness of corrections and warnings of misinformation for a long time. Ecker et al. (2010) studied whether the continued influence of misinformation can be reduced by explicit warnings at the outset that people may be misled. They found that a specific warning with detailed information was more efficient than a general warning reminding people that facts are not always properly checked. However, a specific warning can reduce reliance on an outdated source of information but not eliminate it. Pennycook et al. (2018) investigated how fluency via prior exposure contributes to the believability of fake news. They found that tagging fake stories as disputed is not an effective solution because it simply attracts even more attention to the problem. They also found that repeating headlines increases perceptions of their accuracy. Schwarz et al. (2016) found that the myth-versus-fact article format is not efficient to deal with fake news because such articles subtly reinforce the myths through repetition and further increase the spread and acceptance of misinformation. Unfortunately, such articles make misinformation even more easily accessible by repeating it and illustrating it with pictures. This increases the probability that misinformation that the communicator wanted to debunk will continue to be delivered. They found that it is better to simply provide correct information rather than try to correct wrong information. They also identified five criteria that people use to assess the accuracy of information: acceptance by others, amount of supporting evidence, compatibility with one's own beliefs, general coherence of the statement, and credibility of the information source. Lerman (2016) stated that the interplay between humans' cognitive limits and the social media network structure influences the spread of information. Finally, Chan et al. (2017) found that debunking messages for the correction of misinformation only increases the effects of the misinformation.

Misinformation and tools for mitigating it. Several tools have been developed to counter misinformation, such as Botometer, Foller.me, TinEye, Insigna, Rbutr, Fakespot, NewsGuard, Greek Hoaxes Detector, DejaVu and Social Sensor.

- Botometer detects social bots and classifies online social media user accounts as either bots or human beings. This classification is based on various features of the user account profile, online social network structures, historical patterns of activity, and language and sentiments (Yang et al., 2019; Botometer tool, 2019).
- Foller.me analyses the profiles and tweets of social network users and shows various user characteristics, for example, general information such as name, location, language, join date, and time zone; statistics about tweets (number of tweets, followers, following); and tweet analysis (tweet replies, retweets, tweets with links). The main idea is to understand the detailed profiles of social media users to verify social media content (Sloan and Quan-Haase, 2016; Foller.Me tool, 2019).
- TinEye analyses user-generated content, like photos and videos, as well as detects whether an image, audio content, or video content is fake (Middleton, 2017; Tineye Tool, 2019). Members of the global community, in particular journalists, use this tool and others, such as FotoForensics and Google Reverse Image, to examine user-generated content.
- Rbutr is a machine-learning algorithm applied to community feedback to capture webpages with disputed, rebutted, or contradicted parts elsewhere on the Internet. This tool also provides sector-wise (e.g., health, education, immigrant, climate change) repositories of news and community rebuttal (Mensio and Alani, 2019) and provides warning messages (e.g., "This is potentially malicious") for particular news webpages with a bad reputation.
- Fakespot is a browser plugin that assesses the validity of online reviews based on their URL (Mensio and Alani, 2019; Fakespot Analyzer Tool, 2019).
- NewsGuard is another browser plugin that integrates the opinions of a large pool of journalists and informs users about the reliability of news websites and organisations. It uses nine journalist credibility and transparency criteria that are combined into labels (NewsGuard Tool, 2019).
- Greek Hoaxes Detector is a browser plugin that analyses news articles and assigns labels such as "scam," "hoax" or "fake" (Ellinika Hoaxes Tool, 2019).
- DejaVu is a system for detecting visual misinformation in the form of image manipulation aimed for use by journalists (Matatov et al., 2018).
- Social Sensor is a software that gathers social media data and analyses trends and what influences them (Schifferes et al., 2014).

The aforementioned tools were designed for particular purposes and are limited in several respects, such as the following:

1. Requirements for participation: Some tools were developed based on stakeholder feedback. However, the developers did not involve end users in the process of developing the tools. They also did not collect end users' preferences regarding these tools. When stakeholders were involved, it was frequently only one group of stakeholders or a very narrow circle of professionals who deal with misinformation. This has resulted in a narrow focus on professional intent instead of on how consumers of information can reduce their uptake of misinformation.
2. Technical issues: Almost all of these browser plugins support only Google Chrome.
3. Lack of integration of the views of fact-checkers: Fact-checkers are part of a growing community that plays an essential role in media policies. However, several of these tools were developed without any consideration from this community, which has led to unnecessarily incomplete detection mechanisms.

Consequently, the full potential of fact-checking services has not been fully realised, and the lack of transparency in development and input parameters makes them unclear. This has led to decreased user trust, which is why it seems reasonable to evaluate the functionality of existing fact-checking tools to identify possible gaps. This is best done in a collaborative environment with a high degree of involvement by relevant stakeholders (Horne et al., 2019).

A few studies have focussed on assessing the perceived needs of journalists navigating misinformation. In Schifferes et al. (2014), 22 journalists participated in an interview regarding the functionalities most relevant for tools countering online misinformation. According to this study, journalists emphasise the need to predict breaking news and verify content on social media as true or false. Brandtzaeg et al. (2018) conducted a study with 32 journalists and social media users on perceptions of fact-checking tools, concluding that users must be able to understand the limitations of tools and that tools need to be transparent on all ends, including in terms of funding. To the best of our knowledge, policymakers have not yet been included in such studies, although it is clear that policies desire the delivery of tools for dealing with misinformation.

Participatory governance and value-based software engineering. Several scientific works have discussed the need to understand the typology and features of misinformation (Rossi and Lenzini, 2020; Koulolias et al., 2018). The design and evaluation process we argue for in this article involves two components: (a) co-creation by users and elicitation of user preferences and (b) adequate aggregation and evaluation mechanisms. By "co-creation," we mean a process that is aligned with Peters and Heraud (2015) and Gummesson et al. (2014) as an adaptive and inclusive approach to participatory governance based on the engagement and involvement of various stakeholder groups. Participatory governance, which is embodied in processes that empower citizens to participate in public decision making, has been gaining acceptance as an effective means of tackling democratic deficits and improving public accountability.

Participatory governance and co-production processes require an understanding of human factors such as individual patterns of decision-making processes, as well as cognitive and behavioural biases; institutional structures; perceptions of the risks, benefits, and costs of various policy interventions; as well as a need for compromise-oriented solutions to honour diverse views and a variety of voices.

Participatory governance also requires the involvement of various stakeholders. Stakeholder involvement in decision-making processes and in the development of tools and decision support systems is essential for meeting stakeholder requirements (cf. Komendantova et al., 2014). Furthermore, authors such as Kujala and Väänänem-Vainio-Mattila (2009) have shown that it is essential to consider stakeholders' values regarding the functionalities and features of a tool when designing new software and that tools so designed are more likely to be used by the groups in question.

To achieve this, a number of techniques may be used. Khari and Kumar (2013) tested common approaches experimentally with stakeholders, concluding that value-oriented prioritisation (VOP) met the demands and the environment of the stakeholders better than other techniques. VOP, a so-called preference-based approach that relies on techniques and models from the field of decision analysis, aims to elicit users' values by studying their preferences (see Vetschera, 2006, for an introduction in the context of software engineering). Basic VOP is a scoring-based additive weighting approach in which a stakeholder or prospective user ranks features (or requirements) according to his or her value-in-use (see Azar et al., 2007). If there is more than one user or stakeholder, the VOP process turns into a group decision problem (i.e., gathering preferential data from several stakeholders or prospective users to identify a selection of features that provides maximum value to users while respecting the resources of the development team). However, VOP in itself is not flexible enough to handle ranking statements and aggregate preferences from several stakeholders in an elaborated way. For this purpose, there exist the novel methods from the field of decision analysis described in the following section.

Methodology

The empirical data in this study were collected during a co-creation process with stakeholder groups that used workshops and interviews to extract design components from stakeholder dialogues and findings. The aim was to provide insights into expected requirements for anti-disinformation tools. A specially adapted multi-criteria decision framework (Danielson et al., 2020) was then used to understand the desirability of various system features of a tool for mitigating misinformation.

Workshop setup and participants. The co-creation workshops consisted of stakeholders from three groups (journalists/fact-checkers, policymakers and citizens) in three countries (Austria, Greece and Sweden). The purpose of the workshops was to discuss misinformation and, over several sessions, collect perceptions of misinformation, test and discuss tools that address misinformation, as well as various features of these tools. Furthermore, we explored how information about particular online tools can be transferred to stimulate critical thinking and trust, as the latter is an important parameter in software adoption (Wu et al., 2011).

We used the following sampling and invitation process. After thorough desktop research, a list of organisations was created that identified the most important stakeholders on the topic. A final contact list of various organisations representing our three stakeholder groups (policymakers, citizens, and journalists) was prepared. Hosting pilot team members were assigned to contact the organisations and to update the list accordingly. Subsequently, formal letters of invitation were issued to the target participants. The letter included a brief description of the Co-Inform program and the workshop objectives. It also included the workshop agenda (Appendix II). The team followed up with phone calls to the identified stakeholders and personally explained to them the goals of the project, the workshop methodology, and the importance of their participation. Two days before the event, a reminder e-mail was sent to the list of confirmed participants that provided them with more information about the location of the event.

The formats of the workshops, as well as the sampling and invitation processes, were identical for all three countries to exclude the possibility that the results were influenced by differences in sampling process or format.

The policymaker group consisted of government organisations (Ministry of Finance, Ministry of Education, Ministry of Health), nongovernmental organisations (Solidarity Now, Danish Refugee Council, UNHCR and others), grassroots organisations (domain expert organisations like Velos Youth Center), and municipality services organisations (organisations that provided aid to refugees, like Greek Refugee Council, could help us recruit refugees). The citizen group consisted of people from local communities, people from civil societies, refugees, migrants, as well as academics. The journalist group consisted of people from news agencies, radio, and television.

The first co-creation workshop took place in September 2018 in Tokyo, Japan, and was organised by the International Council for Information Technology in Government Administration and the Organisation for Economic Co-operation and Development (OECD). The 103 participants at the first multi-stakeholder workshop included 11 government chief information officers, 65 high-ranking public officials, 8 journalists, 8 executives of international organisations, 9 executives from the private sector, and 2 policymakers. The purpose of the workshop was to assess the effects of misinformation in society and suggest mitigation strategies for the public sector.

The second co-creation workshop was portioned among the three countries and took place in February–March 2019. The purpose of this workshop was to assess the initial needs of participants around misinformation, their level of trust in news sources, and their perceptions of misinformation and to collect their recommendations on possible interventions and policies. In Vienna, the Co-Inform workshop was organised in cooperation with the Ministry of Economy and Digitalization and included 21 policymaker, journalist, and citizen stakeholders, including representatives of the Austrian Chamber of Labour, the Housing Service of the Municipality of Vienna, and the Austrian Association of Cities and Towns. In Sweden, it included 16 participants, of whom four were journalists, five policymakers (mainly from the Social Democratic Party), and seven citizens (including from Anti-Rumour Sweden). It was hosted by the Botkyrka Multicultural Centre. In Greece, the workshop took place in the community of Serafeio with 31 participants (9 journalists, 9 policymakers, and 13 citizens), including representatives of the Ministries of Finance, Digital Policy, Health, Immigration and Education.

The third co-creation workshop took place in these same countries in November 2019. The major theme of the third Co-Inform workshop was "Which features make people engage with misinformation-combatting tools, and why?" The theme was addressed over a series of five sessions: introduction to the overall workshop process, categorisation theory exercise, assessment of features of the interface of a potential tool, Multi-Criteria Decision Analysis (MCDA) sessions, and repertoire grid-nudging focus group sessions. Altogether 15 participants attended the third Co-Inform workshop in Sweden: 3 journalists, 1 policymaker, and 11 citizens. In Greece, 19 people participated: six citizens, seven journalists, and six policymakers. In Austria, 16 stakeholders attended the workshop: five citizens, six journalists, and five policymakers.

The only difference among the aforementioned workshops was that the participants belonged to three different cultures:

- Workshop 2 (as per our article): We recruited participants from all stakeholder groups (citizens, journalists, policymakers) who were related to organisations that worked with migrants.
- Workshop 3 (as per our article): We recruited participants from all stakeholder groups (citizens, journalists, policymakers) without focussing on any specific domain.

The format, agenda, and master plan of the workshops were as follows. All pilot countries (Greece, Austria and Sweden) followed a common format that included an agenda, templates, survey forms, exercises, and workshop sessions based on a common master plan that was prepared by the responsible Co-Inform project partners in consultation with Co-Inform project technical partners. Each workshop followed the same master plan. In addition, discussions were held in all three Co-Inform pilot countries on common topics as per the master plans of the workshops.

A main objective of the third workshop was to collect input on perceptions of functionalities, user experience features, and system features of tools that deal with misinformation in social media. Four sessions were conducted during each workshop. During the first session, the participants were presented with features. This was followed by a detailed discussion of each feature and the collection of feedback on what should be included or added. The following features were subject to evaluation by the participants:

- Feature 1 (*Awareness*): I am aware of existing misinformation online.
- Feature 2 (*Why and when*): I want to know why a claim has been flagged as misinformative. And I want to know who flagged it and when.
- Feature 3a (*How it spreads and by whom*): I come across something that I find misinformative. I would like to know how this information has spread online and who has shared it.
- Feature 3b (*Life cycle [timeline]*): I want to know the life cycle (timeline) of a misinformative post/article (e.g., when it was first published, how many fact-checkers have debunked it, and when it was shared again).
- Feature 4a (*Sharing over time*): I want to be able to quickly understand how much misinformation people have shared over time through an overall misinformation score.
- Feature 4b (*How misinformative an item is*): I want to be able to quickly understand how much misinformation a news item or tweet may contain through the provision of an overall misinformation score.
- Feature 5a (*Instant feedback on arrival*): When I encounter a tweet from someone else that contains misinformative content, I want to be informed that it is misinformative.
- Feature 5b (*Inform on consistent accounts*): I want the Co-Inform system to inform me of which accounts (within my network) consistently generate/share/create misinformative content.
- Feature 5c (*Self-notification*): I want the Co-Inform tools to notify me whenever I repeatedly share misinformation.
- Feature 6 (*Credibility indicators*): I want to see credibility indicators that I will immediately understand, and I want the credibility indicators to look very familiar, like other indicators online.
- Feature 8 (*Post support or refute*): I want to be able to post links to reputable articles and data that support or refute the story or claim.
- Feature 9 (*Tag veracity*): I want to be able to tag the veracity of an element (tweet, story, image, or sentence/claim) in the current tab I am seeing.
- Feature 10 (*Platform feedback*): I want to be able to receive feedback on what the platform is doing and has done with the tags and evidence I have submitted.

The participants then ranked the features under three criteria, creating three different rankings of the features, one for each criterion. The three criteria were as follows:

- Trust: for trust in this tool
- Critical thinking: for making me think twice before I trust and/or share
- Transparency: for transparency in how the tool makes judgements

Thereafter, they ranked the three criteria with respect to their relative importance based on question on the form: "The top-ranked features under Trust, do they provide more or less value for you compared to the top-ranked features under Critical thinking?" If the answer was "yes," then trust was ranked above critical thinking (i.e., it was deemed to be of more relative importance, because the participant perceived greater value if the top-ranked features for trust were available compared to the top-ranked features for critical thinking).

Elicitation and evaluation. Danielson et al.'s (2020) decision analytic framework was used as the rank-based elicitation and evaluation method. The method was implemented in DecideIT 3.1, which was also used in the workshops. Briefly, DecideIT is capable of operating with incomplete or numerically imprecise input data, such as rankings and interval value statements, in a combined model. To represent the ranking statements, we used a cardinal ranking approach (P-CAR). P-CAR is a calibrated method of creating feasible input in the form of surrogate imprecise value statements, which are derived from rankings provided by stakeholders. The feasible information is represented in the form of linear inequalities (greater than) in combination with interval bounds and a focal point that represents the most feasible surrogate value for a given element given its position in the ranking. This enables conventional multi-attribute value aggregation (Dyer and Sarin, 1979) so the results can be evaluated across multiple stakeholders and criteria. See Danielson and Ekenberg (2019) for details on P-CAR.

The evaluation method originated from earlier work on evaluating decision situations involving numerically imprecise input. To avoid problems with aggregation when handling set membership functions and similar features, higher order distributions for better discrimination between the possible outcomes are introduced. To alleviate the problem of overlapping results, the methodology also contains a new evaluation method based on the resulting belief mass over the output intervals, without introducing further complicating aspects into the decision model. During the process, consideration is given to the entire range of output values, as well as how plausible it is that a specific feature will outrank the remaining ones, thus providing a robustness measure. In this way, DecideIT can evaluate the actual proportion of aggregated values for which a feature is considered more favourable than another, that is, whether there is a significantly larger amount of the feasible information (i.e., in

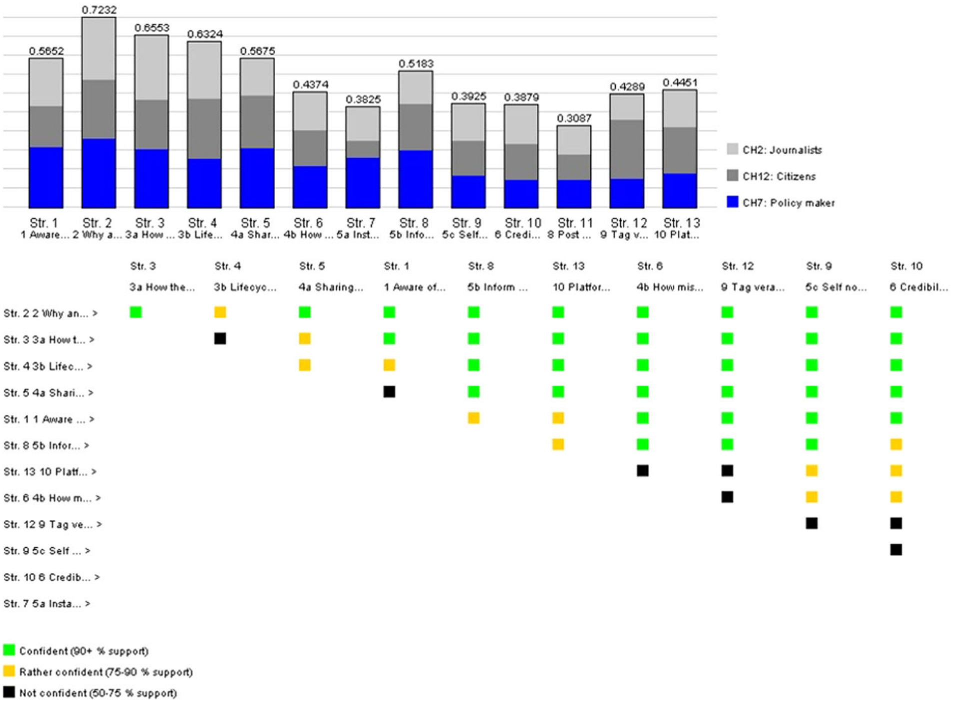


Fig. 1 Results for all groups of stakeholders. Preferences.

Table 1 Major preferences of various stakeholders regarding how to address misinformation in the three case countries (Sweden, Greece and Austria).

Major preference	Citizens	Policymakers	Journalists
The main goal of misinformation (political, societal) should be tracked	+++	+	++
Internet and social media, as well as the mass production of information, are factors that help the spread of fake news	++	+	++
Need for education and relevant tools	+++	++	+++
Provision of references for validation of information	+	+	+
Tools that give indications and extra information on an article are needed	+	+++	+
Tools that help in sharing are needed	+	++	+
Collaboration among all is needed through facilitating tools	+++	+++	+++

the set of rankings provided by the participants where one feature is deemed to provide more value compared to another feature). This can be seen more concretely in Fig. 1, which shows the proportion of feasible information (e.g., Feature 2 is deemed to be more valuable than the rest, Features 3a and 3b are basically equal and more valuable than Feature 4 and the remaining features). See Danielson et al. (2020) for a detailed description of the tool and its underlying theory and Larsson et al. (2018) for details on aggregation across multiple stakeholders/participants.

Results

In line with OECD's Recommendation on Digital Government Strategies, the findings from the first co-creation workshop in Tokyo emphasised the need for open, inclusive, accountable, and transparent processes by national governments and highlighted the fact that digital transformation in the public sector, as well as increasing accessibility of the Internet, has exacerbated various problems related to misinformation. Given the importance of factual information for combatting misinformation in the public arena, governments need to collaborate with stakeholders and invest in innovative ways of dealing with misinformation. A number of specific actions were proposed to deal with this societal challenge. Empowerment of citizens, encouraged engagement, education, moderate legislative action, as well as investment in new technologies are invaluable means of tackling misinformation. For fragmented technological and innovative solutions to succeed in tackling misinformation on a broad scale, they need to be integrated and embedded into a co-creational system of policies. More collaborative and effective management of misinformation needs to be supplemented with informed behaviours among citizens. Creating a trusted environment for citizens with adequate education is necessary as we enter an era in which big technological advances have the potential to disrupt even more than they already have.

The subsequent workshops took place in three different locations on two separate occasions and provided cross-cultural data for comparing the needs of various stakeholder groups related to decision support models. Data were collected, and the needs of citizens, policymakers, and journalists were identified. Table 1 shows that the need for collaboration and facilitation of tools was identified in all three case countries and by all three stakeholder groups. The need for tools to address education and awareness raising was also identified in all three countries and across all three groups of stakeholders. These tools are also required for sharing reliable information. However, an automatic correction

Table 2 Citizen preferences in Vienna, Stockholm and Athens.

Country	Major	Minor	Comparison
Vienna	Credibility indicators Awareness	Instant feedback on arrival Post support or refute How misinformative an item is	Had a lower score on platform feedback, life cycle (timeline), and sharing over time than in Stockholm and Athens
Stockholm	Why and when How it spreads and by whom Life cycle (timeline) Sharing over time Inform on consistent accounts	Tag veracity Platform feedback	Had a lower score on tag veracity than in Vienna and Athens
Athens	Life cycle (timeline) Sharing over time Tag veracity Platform feedback	Awareness Instant feedback on arrival Credibility indicators	Had a lower score on why and when than in Vienna and Stockholm

mechanism for validating information was identified as the least desired option.

These preferences were identified during roundtable discussions in the workshops. Discussions followed the same protocol in all three countries: Topics for discussion were provided, but preferences were not identified in advance. The major preferences were identified based on a review of the transcripts of these discussions and on the frequency of mention of certain topics.

A rather passive attitude. The results for all groups of stakeholders and all three countries showed interest in the spread of misinformation but revealed a rather passive attitude toward dealing with misinformation. The three most popular answers across all groups of participants were the following:

- Feature 2 (*Why and when*): I want to know why a claim has been flagged as misinformative. And I want to know who flagged it and when.
- Feature 3a (*How it spreads and by whom*): I come across something that I find misinformative. I would like to know how this information has spread online and who has shared it.
- Feature 3b (*Life cycle [timeline]*): I want to know the life cycle (timeline) of a misinformative post/article (e.g., when it was first published, how many fact-checkers have debunked it, and when it was shared again).

A general observation was thus that the participants wanted to know who spread the misinformation, why, and how, as well as the timeline of the spread. However, although they wanted to be informed, they did not feel motivated to take further action. This is a rather passive way of dealing with misinformation in general. The participants wanted to be informed when an item was misinformative, but they did not prioritise dealing with the general topic of misinformation, actively reporting and correcting information, or being informed if they themselves were sharing misinformation.[2]

Figure 1 aggregates the results from all stakeholder groups in all three countries. The problem is a multi-stakeholder multi-criteria decision problem that is evaluated as a multi-linear problem given the background information. This means that the weighted averages of the values of the respective features are evaluated balanced by weights derived from the criteria rankings above (i.e., equations of the format $E(F_j) = \Sigma w_i v_{ij}$, where w_i is the weight of criterion i and v_{ij} is the value of feature F_j under criterion i). The value $\mathrm{E}(\mathrm{F}_j)$ is computed by solving successive optimisation problems in the program DecideIT (cf., e.g., Danielson and Ekenberg, 2019). Briefly, the higher the bar for a feature, the better that feature is. The respective portions (blue, light grey and dark grey) show the impact of the respective criteria. Furthermore, the coloured squares show the robustness of the results. Green indicates a significant difference between features, which means that there must be substantial changes in the input data for a feature to change. Orange indicates a difference that is more sensitive to the input data. Black indicates that there is basically no difference between the features.[3] For instance, from Fig. 1, we can see that Feature 2 (*Why and when*) is deemed more valuable than the rest. We can also see from the green square that this result is quite robust.

From the figure, we see that Features 3a and 3b are basically equal (black square) but preferred over Feature 4 (yellow square) and the remaining features. Some answers indicating a more active position, such as Feature 5c (*Self-notification*): "I want the Co-Inform tools to notify me whenever I repeatedly share misinformation" and Feature 6 (*Credibility indicators*): "I want to see credibility indicators that I will immediately understand, and I want the credibility indicators to look very familiar, like other indicators online," were ranked at the bottom.

Citizens. Table 2 shows the main differences in citizens' attitudes in the case countries. Heterogeneity regarding the best technology features can be observed.

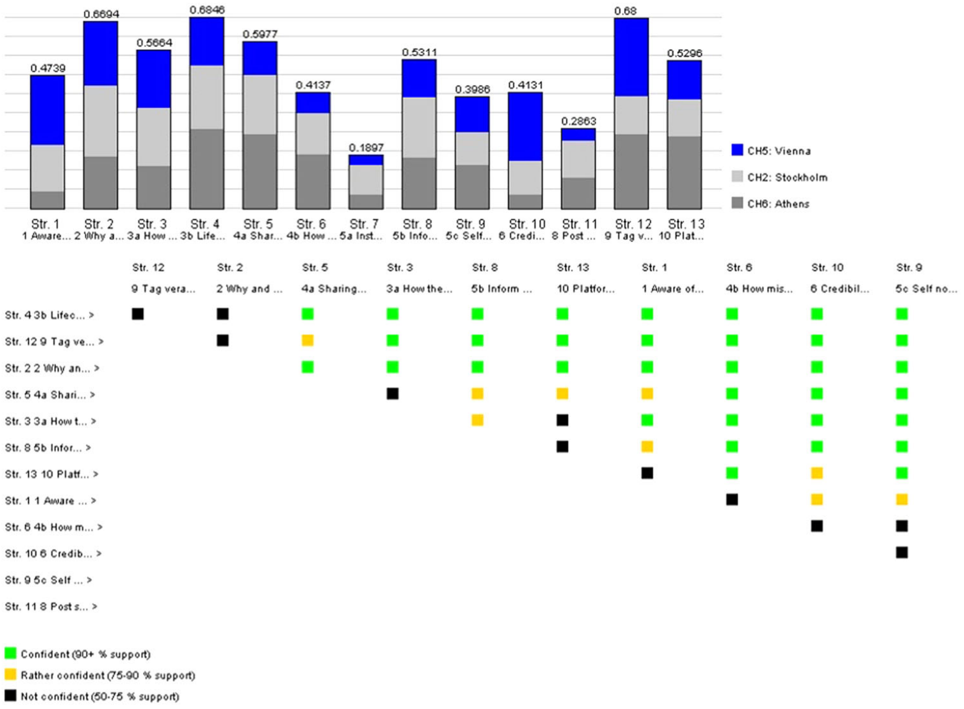
In particular, the most preferred option for the citizens from Vienna was as follows:

- Feature 6 (*Credibility indicators*): I want to see credibility indicators that I will immediately understand, and I want the credibility indicators to look very familiar, like other indicators online.

  This is somewhat surprising, as this feature ranked quite low in the joint results for all three countries and all three groups of stakeholders (see Fig. 1) and was ranked lowest in the Athens workshop. There seemed to be quite a strong polarisation regarding this feature. In contrast, Stockholm and Athens had equal preferences for Features 3b and 4a.

  - Feature 3b (*Life cycle [timeline]*): I want to know the life cycle (timeline) of a misinformative post/article (e.g., when it was first published, how many fact-checkers have debunked it, and when it was shared again).
  - Feature 4a (*Sharing over time*): I want to be able to quickly understand how much misinformation people have shared over time through an overall misinformation score.

Athens and Vienna had the lowest scores on Feature 5a (*Instant feedback on arrival*): "When I encounter a tweet from



	Str. 12 9 Tag vera...	Str. 2 2 Why and ...	Str. 5 4a Sharing...	Str. 3 3a How the...	Str. 8 5b Inform ...	Str. 13 10 Platfor...	Str. 1 1 Aware of...	Str. 6 4b How mis...	Str. 10 6 Credibil...	Str. 9 5c Self no...
Str. 4 3b Lifec... >	⬛	⬛	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩
Str. 12 9 Tag ve... >		⬛	🟨	🟩	🟩	🟩	🟩	🟩	🟩	🟩
Str. 2 2 Why an... >			🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩
Str. 5 4a Shari... >				⬛	🟨	🟨	🟨	🟩	🟩	🟩
Str. 3 3a How t... >					🟨	⬛	🟩	🟩	🟩	🟩
Str. 8 5b Infor... >						⬛	🟨	🟩	🟩	🟩
Str. 13 10 Platf... >							⬛	🟩	🟨	🟩
Str. 1 1 Aware ... >								⬛	🟨	🟨
Str. 6 4b How m... >									⬛	⬛
Str. 10 6 Credib... >										⬛
Str. 9 5c Self ... >										
Str. 11 8 Post s... >										

🟩 Confident (90+ % support)
🟨 Rather confident (75-90 % support)
⬛ Not confident (50-75 % support)

Fig. 2 Citizen preferences in Sweden, Greece and Austria. Preferences.

Table 3 Journalist preferences in Vienna, Stockholm and Athens.

Country	Major	Minor	Comparison
Vienna	How misinformative an item is Awareness Self-notification Credibility indicators	Post support or refute Tag veracity	Vienna had lower preferences than Athens and Stockholm for why and when, how it spreads and by whom, and life cycle and higher preferences for how misinformative an item is, self-notification and credibility indicators
Stockholm	Why and when How it spreads and by whom Life cycle (timeline)	How misinformative the item is Credibility indicators	Stockholm had lower preferences for how misinformative an item is and credibility indicators
Athens	Why and when How it spreads and by whom Life cycle (timeline) Platform feedback	Instant feedback on arrival Self-notification	Athens had lower preferences for instant feedback on arrival and self-notification

someone else that contains misinformative content, I want to be informed that it is misinformative." Fig. 2 aggregates citizens' preferences in the three countries (equally weighted).

We can see that Features 3b and 9 are the most (and equally preferred) features when data from the three countries are aggregated, followed by Feature 2. There was thus no clear distinction between the two highest ranked options.

Journalists. An overview of journalists' preferences in the respective countries is shown in Table 3.

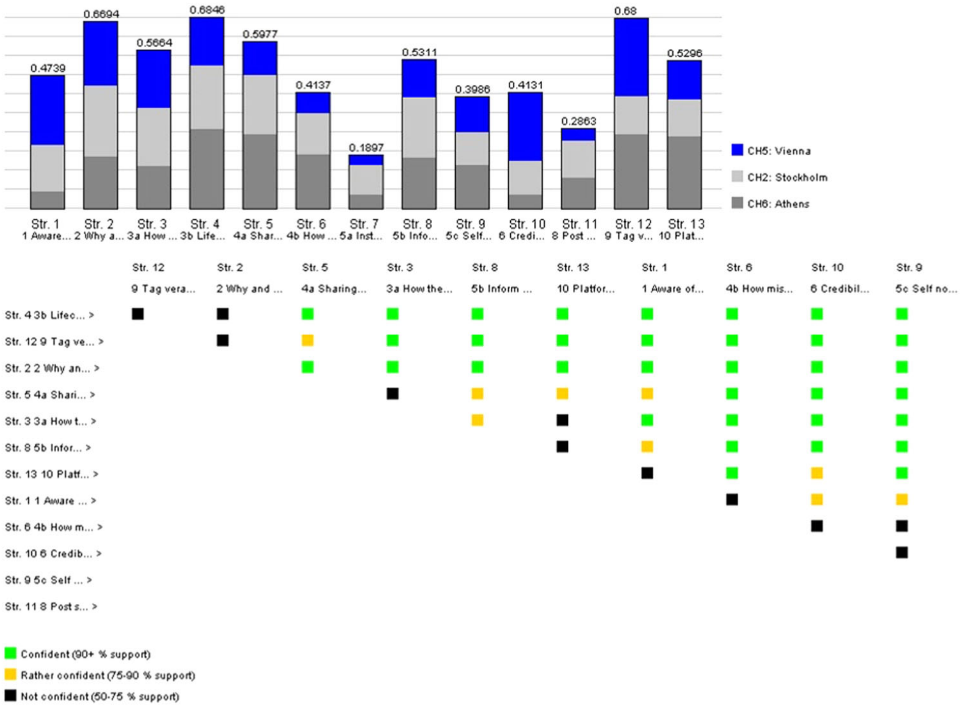
The highest ranked features for the journalist group were as follows:

- Feature 2 (*Why and when*): I want to know why a claim has been flagged as misinformative. And I want to know who flagged it and when.
- Feature 3a (*How it spreads and by whom*): I come across something that I find misinformative. I would like to know how this information has spread online and who has shared it.

These two options were also the best choices in Stockholm and Athens and (consequently) corresponded well with the average for all groups and all three case countries (see Fig. 1). Another difference between journalists in Vienna and journalists in the two other case countries was that the highest ranked features in Vienna were the lowest ranked ones in Athens and Stockholm. Figure 3 aggregates journalists' preferences in the three countries (equally weighted).

Policymakers. There were no results for policymakers in Stockholm because only one policymaker participated and did not make any choices for the analysis. Table 4 shows the two most prioritised features of the policymakers in Greece and Austria.

We can see that the most preferred features were the following:



	Str. 12 9 Tag vera...	Str. 2 2 Why and ...	Str. 5 4a Sharing...	Str. 3 3a How the...	Str. 8 5b Inform ...	Str. 13 10 Platfor...	Str. 1 1 Aware of...	Str. 6 4b How mis...	Str. 10 6 Credibil...	Str. 9 5c Self no...
Str. 4 3b Lifec... >	■	■	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩
Str. 12 9 Tag ve... >		■	🟨	🟩	🟩	🟩	🟩	🟩	🟩	🟩
Str. 2 2 Why an... >			🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩
Str. 5 4a Shari... >				■	🟨	🟨	🟨	🟩	🟩	🟩
Str. 3 3a How t... >					🟨	■	🟩	🟩	🟩	🟩
Str. 8 5b Infor... >						■	🟨	🟩	🟩	🟩
Str. 13 10 Platf... >							■	🟩	🟨	🟩
Str. 1 1 Aware ... >								■	🟨	🟨
Str. 6 4b How m... >									■	■
Str. 10 6 Credib... >										■
Str. 9 5c Self ... >										
Str. 11 8 Post s... >										

🟩 Confident (90+ % support)
🟨 Rather confident (75-90 % support)
■ Not confident (50-75 % support)

Fig. 3 Journalists preferences in Vienna, Stockholm and Athens. Preferences.

Table 4 Policymaker preferences in Vienna and Athens.

Country	Major	Minor
Athens	Life cycle (timeline)	Credibility indicators
	Why and when	
Vienna	Why and when	Post support or refute
	Awareness	Tag veracity
	Credibility indicators	Life cycle (timeline)
	Sharing over time	Platform feedback

- Feature 2 (*Why and when*): I want to know why a claim has been flagged as misinformative. And I want to know who flagged it and when.
- Feature 4a (*Sharing over time*): I want to be able to quickly understand how much misinformation people have shared over time through an overall misinformation score.

Figure 4 aggregates policymakers' preferences in the two countries (equally weighted).

Differences by country. There were some significant differences among the joint stakeholder groups in the different countries. As can be seen from Table 5, the stakeholder groups in Austria preferred credibility indicators. In Greece, the stakeholder groups preferred life cycle (timeline) features. In Sweden, the stakeholder groups preferred the why and when feature, as well as the how it spreads and by whom feature.

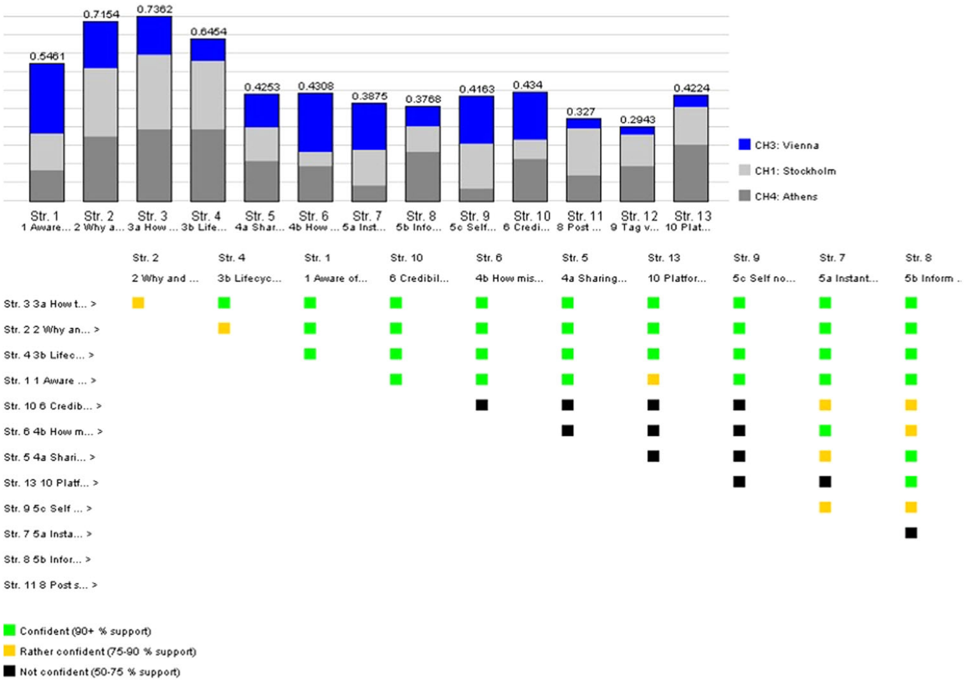
It was beyond the scope of this study to understand why these differences appeared and how cultural background influenced them. However, it would be interesting for further research to identify the impacts of various cultural factors, such as values, history of participation in each country, and others, on preferences for features of tools for mitigating misinformation in various countries. Therefore, we recommend that developers of tools consider differences in preferences regarding tools that are influenced by cultural background.

In sum, there were a variety of preferences for necessary features of tools for mitigating misinformation, and they seemed to be correlated with cultural background rather than with stakeholder group. That is, it seems that cultural context plays a large role in these preferences (and thus the intended specific use of the tools). An important conclusion from this (albeit limited) study is that tools for mitigating misinformation must be flexible enough because it will probably be hard to produce a global (or context-independent) tool that balances all desirable features in such a way as to appeal to a general global public.

Conclusions

There is a need for comprehensive solutions for designing tools for mitigating misinformation from a value-based software engineering perspective. Here, we demonstrated a framework for evaluating tools for detecting misinformation using a preference elicitation approach, as well as an integrated decision analytic process for evaluating desirable features of systems for combatting misinformation. The framework was tested in workshop settings in Athens, Vienna and Stockholm, where a decision analytic methodology was used to address three interdependent factors:

- Elicitation: There are significant difficulties with eliciting preferences, in particular in group decision making and negotiations. We therefore constructed a process based on preference rankings and negotiations, as well as algorithms for aggregating the results.
- Evaluation: In general, decision analytic evaluation methods are inflexible in relation to the complex nature of decision problems, and it is usually difficult to aggregate information. Furthermore, there is little or no constructive feedback from evaluation methods. Our process enables more interactive use of group rankings and possibilities to see the effects of conflicting opinions and how they affect the final results. If the results are not robust (because opinions conflict too much), negotiation can begin again until there is agreement



	Str. 2	Str. 4	Str. 1	Str. 10	Str. 6	Str. 5	Str. 13	Str. 9	Str. 7	Str. 8
	2 Why and ...	3b Lifecyc...	1 Aware of...	6 Credibil...	4b How mis...	4a Sharing...	10 Platfor...	5c Self no...	5a Instant...	5b Inform ..
Str. 3 3a How t... >	🟨	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩
Str. 2 2 Why an... >		🟨	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩
Str. 4 3b Lifec... >			🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩
Str. 1 1 Aware ... >				🟩	🟩	🟩	🟨	🟩	🟩	🟩
Str. 10 6 Credib... >					■	■	■	■	🟨	🟨
Str. 6 4b How m... >						■	■	■	🟩	🟨
Str. 5 4a Shari... >							■	■	🟨	🟩
Str. 13 10 Platf... >								■	■	🟩
Str. 9 5c Self ... >									🟨	🟨
Str. 7 5a Insta... >										■
Str. 8 5b Infor... >										
Str. 11 8 Post s... >										

🟩 Confident (90+ % support)
🟨 Rather confident (75-90 % support)
■ Not confident (50-75 % support)

Fig. 4 Policymakers preferences in Vienna and Athens. Preferences.

Table 5 Preferences of the three stakeholder groups in Austria, Greece and Sweden.

Stakeholder group	Austria	Greece	Sweden
Citizens	Credibility indicators, awareness	Life cycle (timeline), sharing over time, tag veracity, platform feedback	Why and when, how it spreads and by whom, life cycle (timeline), sharing over time, inform on consistent accounts
Journalists	How misinformative an item is, awareness, self-notification, credibility indicators	Why and when, how it spreads and by whom, life cycle (timeline), platform feedback	Why and when, how it spreads and by whom, life cycle (timeline)
Policymakers	Why and when, awareness, credibility indicators, sharing over time	Life cycle (timeline), why and when	

or at least clarification where there are essential conflicts.

- Communication: A key aspect of exploiting decision analytic support is facilitating communication of the result and of the perspectives of the members of a group. Here we suggest a workshop format with stakeholder representatives. This is, however, not a required format, and the same basic ideas can be used in a more distributed manner, where direct interaction can be combined with the use of questionnaires and feedback mechanisms for respondents.

In our case study, the participants prioritised information regarding the actors behind the distribution of misinformation and tracing the life cycle of misinformative posts. The fact that it mattered to participants whether someone intended to delude others indicates the participants' preference for trust, accountability, and quality in, for instance, journalism.

Furthermore, the three most valued features across all participants related to the timing and travel of misinformation (when, spread, life cycle), which indicates the significance the participants attributed to the chain of transmission by which a story reached the user. How misinformation travels thus seems to be important for the participants' assessment of the veracity of a claim. With Allport and Postman's (1946) basic law of rumour in mind, participants were interested in shining a light on one of the two critical prerequisites to rumour: They expressed a strong desire to achieve a clearer understanding of what was going on in the case of ambiguous facts or evidence. However, because features requiring active contribution were low ranked, the participants may not have felt personally involved enough in the subject or situation in which there was a need for rebuttal and clarity.

The three most valued features can be assessed using the four-question truth assessment people undertake when evaluating a statement (Lewandowsky et al., 2012). All three features partially contribute to answering the second question of the assessment: "Is the story coherent?" Participants were interested in monitoring the spread of a story, expressing a desire to fill in gaps that a refutation may have left behind, as one piece of information cannot be assessed in isolation. Features 2 and 3a resort to pinpointing the communicator ("who flagged it," "who has shared it"), attending to the third question: "Is the information from a credible source?" To process the information, participants turned to verifying the communicator's credibility. The fourth question, "Do others believe this information?", is inherent in Feature 3b, in which importance is attributed to how many fact-checkers have debunked a story. The number of fact-checkers may therefore create the perception of a strong consensus, which participants may use to counterbalance an erroneous perceived social consensus. No wireframe responds to the first question, "Is the information compatible with what I believe?", as capturing belief structures was outside the scope of our study.

In the third Co-Inform co-creation workshop, we intended to determine the most preferred of the aforementioned 13 features among the stakeholders (citizens, journalists, and policymakers) in the Co-Inform project pilot countries (Greece, Sweden, and Austria). We thoroughly explain stakeholders' preferences for these features in Section 'Results'. However, we did not specifically gather participants' preferences for features of existing tools. During this workshop and discussions with participants, we observed that they desired that the Co-Inform project include the key features of existing misinformation tools. Existing tools, along with feature(s) preferred by the stakeholders, include the following:

Rbutr: includes sector-wise repositories of news and community rebuttal and feedback from the community about a news item or a tweet.

Foller.me and Botometer: provide generic information about social network users to help users know who shared information.

Fakespot: includes assessment/analysis of reviews about a news item or article.

NewsGuard: enables users to provide opinions along with supporting material about the authenticity of a news item or a tweet, informs users how reliable a news item or tweet is, and provides credibility statistics about a news item or a tweet.

Greek Hoaxes Detector: analyses news items or tweets and assigns them a label so that users can see whether they are misinformative.

The participants strongly prioritised these features over those requiring an active contribution to rebutting a story or claim. These observations thus indicate that automated tool support for reliable information detection, tools that support active reasoning, and training in becoming attuned to misinformation strategies (Roozenbeek and van der Linden, 2019) are of high importance. Future studies could further explore the reasons why people engage in a cognitive process of increasing links between nodes for coherence while taking a backseat when tasked with correcting misinformation. This would yield insights into whether and when a push or stimulus may be required when designing novel crowdsourced verification tools.

A main observation is that detection tools by themselves cannot combat misinformation; they must be complemented by other means. Automated solutions must work in a context of general societal awareness in combination with the detection mechanisms we investigated in this study. First, as our results indicate, no tool support is likely able to address all user preferences, which is why tools must be complemented by general awareness. There is thus a need to integrate automated systems with broad public information campaigns, including quick tips for citizens on how to conduct research online by themselves on news articles whose content seems to be dubious. Second, our results indicate that journalists take a rather passive approach to detecting misinformation; there is thus a need to increase awareness of the need for more active detection of misinformation. This can be done by, for instance, organising media and news literacy workshops that bring together, inter alia, fact-checkers and interested citizens with journalists. Third, because even the most sophisticated automatic tools cannot address the entire range of trust issues and preference structures for evaluating them, there is also a need for reinforced legislation to increase transparency among technology companies concerning the use of data and the origin of information. Fourth, there is a need to create diverse and cross-sectorial teams whose tasks are to spot misinformation and to warn the public by providing clear explanations. Finally, media and news literacy classes should be introduced into the school curriculum in parallel with, for instance, topics on information technology (e.g., Koulolias et al., 2018).

Although the first workshop had a global scope, the second and third ones were conducted in Europe. Within Europe, three different geographic locations were chosen (north, middle, south) with cultures that differ in both governance and social media. Although it is impossible to include every country, the sample set constituted good coverage of European countries. Thus, it seems feasible to generalise the findings to Europe as a whole. Because the first workshop focussed more on assessing the effects of misinformation in society, it provided an overview of the scope of the problem in different countries. It did not, however, yield data for classifying the extent of misinformation according to social media maturity, state constitution, or political traditions. Thus, beyond being generalised to Europe, our findings also form a relevant and important basis for conducting similar studies in other parts of the world.

Data availability

For individual privacy reasons data which were collected from stakeholders elicitations and preferences cannot be made available in the public repository.

Received: 21 July 2020; Accepted: 8 December 2020;
[redacted]

Notes

[1] See, for example, www.coinform.eu.

[2] We acknowledge that the preference for the more passive design options is sensitive to the availability of scientific evidence. For example, a passive attitude can be connected to the fact that tools for mitigating misinformation are new to users. Therefore, users might have vague and unclear associations about the features of such tools (Uekermann et al., 2010; Svahn and Lange, 2009).

[3] A description of the technical details of the evaluation procedures is beyond the scope of this paper, but a more in-depth explanation of them is provided in Danielson and Ekenberg (2019).

References

Allport GW, Postman L (1946) An analysis of rumour. Public Opinion Quart 10 (4):501–517

Aurum A, Wohlin C (2007) A value-based approach in requirements engineering: explaining some of the fundamental concepts. In: International Working Conference on Requirements Engineering: Foundation for Software Quality 2007. Springer, 109–115

Azar J, Smith RK, Cordes D (2007) Value-oriented requirements prioritization in a small development organization. IEEE Software 24(1):32–37

Biffl A, Aurum A, Boehm B, Erdogmus H, Grünbacher P (2006) Value-based software engineering. Springer Science & Business Media

Boehm B (2003) Value-based software engineering: reinventing "Earned Value" monitoring and control. ACM SIGSOFT Software Engineering Notes 28(2):3

Botometer Tool (2019) https://botometer.iuni.iu.edu/#!/. Accessed 12 Nov 2019

Brandtzaeg PB, Følstad A, Domínguez MAC (2018) How Journalists and Social Media Users Perceive Online Fact-Checking and Verification Services. J Pract 12(9):1109–1129

Burgoon JK, Blair JP, Qin T, Nunamaker JF (2003) Detecting deception through linguistic analysis. In: International Conference on Intelligence and Security Informatics 2003. Springer, 91–101

Chan MPS, Jones CR, Hall Jamieson K, Albarracín D (2017) Debunking: a meta-analysis of the psychological efficacy of messages countering misinformation. Psychol Sci 28(11):1531–1546

Danielson M, Ekenberg L (2019) An improvement to swing techniques for elicitation in MCDM methods. Knowledge-Based Syst 168:70–79

Danielson M, Ekenberg L, Larsson A,(2020) A second-order-based decision tool for evaluating decisions under conditions of severe uncertainty Knowledge-Based Syst 191. https://doi.org/10.1016/j.knosys.2019.105219

Del Vicario M, Bessi A, Zollo F, Petroni F, Scala A, Caldarelli G, Stanley HE, Quattrociocchi W (2016) The spreading of misinformation online. Proc Natl Acad Sci 113(3):554–559

Dyer J, Sarin R (1979) Measurable multiattribute value functions. Operat Res 27 (4):629–854

Ecker UK (2017) Why rebuttals may not work: the psychology of misinformation. Media Asia 44(2):79–87

Ecker UK, Lewandowsky S, Tang DT (2010) Explicit warnings reduce but do not eliminate the continued influence of misinformation. Memory Cogn 38 (8):1087–1100

Ekström M, Lewis SC, Westlund O (2019) Epistemologies of digital journalism and misinformation. *News Media and Society*. Guest Editorial for Special Issue

Ellinika Hoaxes Tool (2019) https://www.ellinikahoaxes.gr/. Accessed 15 Nov 2019

Fakespot Analyzer Tool (2019) https://www.fakespot.com/.Accessed 14 Nov 2019

Farrel T, Mensio M, Burrel G, Picollo L, Alani H (2018) D3.2 Survey of misinformation detection methods. Co-Inform Project

Foller.Me tool (2019) https://foller.me/. Accessed 12 Nov 2019

Freeze M, Baumgartner M, Bruno P, Gunderson JR, Olin J, Quinn Ross M, aSzafran J (2020) Fake Claims of Fake News: Political Misinformation Warnings, and the Tainted Truth Effect. Springer

Horne DB, Nørregaard J, Adalı S (2019) Different spirals of sameness: a study of content sharing in mainstream and alternative media. Proceedings of the Thirteenth International AAAI Conference on Web and Social Media (ICWSM 2019), 257–266

Giglietto F, Iannelli L, Rossi L Valeriani A (2016) Fakes, news and the election: a new taxonomy for the study of misleading information within the hybrid media system. Convegno AssoComPol

Gummesson E, Mele C, Polese F, Galvagno M, Dalli D (2014) Theory of value co-creation: a systematic literature review. Managing Service Quality

Khari M, Kumar N (2013) Comparison of six prioritization techniques for software requirements. J Global Res Comput Sci 4(1):38–43

Komendantova N, Mrzyglocki R, Mignan A, Khazai B, Wenzel F, Patt A, Fleming K (2014) Multi-hazard and multi-risk decision support tools as a part of participatory risk governance: feedback from civil protection stakeholders. Int J Disaster Risk Reduct 8:50–67

Koulolias V, Jonathan GM, Fernandez M, Sotirchos D (2018) Combating misinformation: an ecosystem in co-creation. OECD Publishing

Kujala S, Väänänem-Vainio-Mattila K (2009) Value of information systems and products: Understanding the users' perspective and values. J Informat Technol Theory Appl 9(4):4

Larsson A, Fasth T, Ekenberg L, Danielson M (2018) Policy analysis on the fly with an online multicriteria cardinal ranking tool. J Multi-Criteria Decision Anal 25(3-4):55–66

Lerman K (2016) Information is not a virus, and other consequences of human cognitive limits. Future Internet 2016(8):21

Lewandowsky S, Ecker UKH, Seifert CM, Schwarz N, Cook J (2012) Misinformation and its correction. Psychol Sci Public Interest 13(3):106–131. https://doi.org/10.1177/1529100612451018

Matatov H, Bechhofer A, Aroyo L, Ofr, A, Naaman M (2018) DejaVu: a system for journalists to collaboratively address visual misinformation. In: Computation + Journalism Symposium. Miami

Mensio M, Alani H (2019) MisinfoMe: Who's Interacting with Misinformation?. Proceeding of the 18th International Semantic Web Conference, New Zealand, http://oro.open.ac.uk/66341/1/paper526.pdf, Accessed 16 March 2020

Middleton SE (2017) Reveal project deliverable D5.2.2-modality models for trust and credibility, https://revealproject.eu/wp-content/uploads/D5.2.2-Modality-models-for-trust-and-credibility_PU.pdf. Accessed 12 Nov 2019

NewsGuard Tool (2019) https://www.newsguardtech.com/. Accessed 14 Nov 2019

Nyhan B, Reifler J (2010) When corrections fail: the persistence of political misperceptions. Springer

Pennycook G, McPhetres J, Zhang Y, Rand DG (2020) Fighting COVID-19 misinformation on social media: experimental evidence for a scalable accuracy nudge intervention. PsyArXiv https://doi.org/10.31234/osf.io/uhbk9

Pennycook G, Cannon TD, Rand DG (2018) Prior exposure increases perceived accuracy of fake news. J Exp Psychol 147(12):1865. https://psycnet.apa.org/record/2018-46919-001

Peters MA, Heraud R (2015) Toward a political theory of social innovation: collective intelligence and the co-creation of social goods. J Self-Govern Manag Econom 3(3):1–14

Piccolo, LS, Joshi, S, Karakanos, E, Farrell T (2019) Challenging misinformation: exploring limits and approaches, IFIP Conference on Human-Computer Interaction 2019. Springer, pp. 713–718

Rossi A, Lenzini G (2020) Transparency by design in data-informed research: a collection of information design patterns. Comput Law Security Rev 37:105402

Roozenbeek J, van der Linden S (2019) Fake news game confers psychological resistance against online misinformation. Pal Commun 5(1):1–10. https://doi.org/10.1057/s41599-019-0279-9

Schifferes S, Newman N, Thurman N, Corney D, Göker A, Martin C (2014) Identifying and verifying news through social media. Digi Journalism 2(3):406–418

Schwarz N, Newman E, Leach W (2016) Making the truth stick & the myths fade: Lessons from cognitive psychology. Behav Sci Policy 2(1):85–95

Sloan L, Quan-Haase A (2016) The SAGE Handbook of Social Media Research Methods. SAGE Publications Ltd, London

Smith EE, Medin DL (1981) Categories and concepts. 1st edn. Harvard University Press

Svahn M, Lange F (2009) Marketing the Category of Pervasive Games. In: Montola M, Stenros J, Waern A (eds) Pervasive games, theory and design, 1st edn. Morgan Kaufman

Tineye Tool (2019) https://tineye.com/. Accessed 12 Nov 2019

Tromble R, McGregor SC (2019) You break it, you buy it: the naiveté of social engineering in tech-and how to fix it. Politi Commun 36(2):324–332. https://doi.org/10.1080/10584609.2019.1609860

Uekermann F, Herrmann A, Wentzel D, Landwehr JR (2010) The influence of stimulus ambiguity on category and attitude formation. Rev Manag Sci 4(1):33–52

Vanenzuala S, Halpern D, Katz JE, Miranda JP (2019) The paradox of participation versus misinformation: social media, political engagement, and the spread of misinformation. Digi Journalism 7(6):802–823

Vetschera R (2006) Preference-based decision support in software engineering. In: Biffl S, Aurum A, Boehm B, Erdogmus H, Grünbacher P (eds) Value-based software engineering. Springer

Wardle C (2016) Six types of misinformation circulated this election season. Columbia Journalism Rev 18. Available at https://www.cjr.org/tow_center/6_types_election_fake_news.php

Wardle C, Derakhshan H (2017) Information disorder: toward an interdisciplinary framework for research and policymaking. Council of Europe Report 27

Wu K, Zhao Y, Zhu Q, Tan X, Zheng H (2011) A meta-analysis of the impact of trust on technology acceptance model: Investigation of moderating influence of subject and context type. Int J Informat Manag 31(6):572–581. https://doi.org/10.1016/j.ijinfomgt.2011.03.004

Yang K, Varo O, Davis CA, Ferrara E, Flammini A, Menczer F (2019) Arming the public with artificial intelligence to counter social bots. arXiv.org e-Print archive. https://arxiv.org/. Retrieved 26 March 2020

Acknowledgements

This work has been partially funded by the Co-Inform project (770302), under the Horizon 2020 call "H2020-SC6-CO-CREATION-2016–2017 (CO-CREATION FOR GROWTH AND INCLUSION)" of the European Commission. The authors acknowledge the support of the Co-Inform team and partners for the organisation of the co-creation workshops and the data collection efforts in Austria, Greece and Sweden.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to N.K.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.