Forthcoming in Risk Analysis

How Can Multicriteria Portfolio Decision Analysis Help Mitigate Mis/Disinformation?

Ahti Salo, Leevi Olander and Tuomas Raivio
Systems Analysis Laboratory
Department of Mathematics and Systems Analysis
Aalto University School of Science
P.O. Box 11100, 00076 Aalto,
FINLAND
emails: ahti.salo@aalto.fi (corresponding author),
leevi.olander@aaltofi, tuomas.raivio@aalto.fi

Abstract: In this paper, we examine and illustrate the relevance of multicriteria portfolio decision analysis as an analytical approach to addressing mis/disinformation risks. After summarizing representative applications of these methods, we draw on the framework for societal resilience developed in the Finnish COVID-19 Science to Panel to elaborate on the range of capabilities that can be leveraged to mitigate mis/disinformation, for example, by maintaining alertness, implementing risk management actions, and interacting with stakeholders to promote awareness. We then present a numerical case study that illustrates how multicriteria portfolio decision analysis can be combined with probabilistic approaches for describing voting behavior and the spread of mis/disinformation to identify cost-efficient strategies consisting of combinations of risk management actions. We conclude by discussing some of the challenges posed by mis/disinformation.

Keywords: Decision analysis, disinformation, misinformation, multicriteria decision making, resilience, governance.

1. Introduction

The creation, dissemination, adoption, and use of false information are a growing concern in modern societies. Rapid dissemination of information across media channels, particularly social media platforms, has become easier, while verifying claims has become more challenging due to limited transparency and access to trustworthy sources (Agarwal and Alsaeedi, 2020). Whether the creation and dissemination of false information result from inadvertent errors (misinformation) or are deliberate, with the intent to cause harm

(disinformation), their influence on subsequent decision-making processes can pose substantial risks to individuals and societal structures (Aven and Thekdi, 2022). Therefore, it is crucial to investigate potential risk management actions that can be implemented at various stages of the process to mitigate harmful consequences.

Finland has been reported to be more resilient to false and misleading content than most other countries, as indicated by its highest average truth quest score in a comprehensive survey of 21 countries (OECD, 2024b; see also Saari et al., 2025). From this perspective, it is instructive to examine how the challenges posed by mis/disinformation are portrayed in the Report of the Future, published by the Finnish Government in September 2025 (Government of Finland, 2025). This Report is a forward-looking strategic document that aims to prepare policymakers for potential societal, technological, environmental, and geopolitical changes, drawing on comprehensive coverage of domestic and international developments.

The most recent Report (Government of Finland, 2025) devotes considerable attention to mis/disinformation, which is seen as a primary driver of the decline in democracy and institutional trust. In particular, the Report warns that AI systems can threaten institutional independence and undermine the rule of law. Rapid technological advancements can create an unpredictable landscape in which traditional democratic safeguards may fail to withstand sophisticated manipulation. The rise of disjointed “information bubbles” and “parallel realities” can make it increasingly difficult for opposing groups to understand one another, undermining social cohesion, fueling social unrest, and weakening the factual basis of democratic discourse. The information environment is considered a central arena for geopolitical rivalry, where both state and non-state actors manipulate information to advance their strategic goals.

Overall, the Report portrays mis/disinformation as a systemic challenge that calls for coordinated policy responses across regulatory, educational, technological, and social sectors, with the aim of strengthening capabilities at multiple levels, spanning international, national, and regional organizations, as well as local communities and individual citizens. The shaping of an adequate risk management strategy against mis/disinformation therefore involves the development, selection, and implementation of actions that are jointly effective in countering harmful impacts. In practice, the preparation, selection, and implementation of such actions may have to satisfy various feasibility constraints related to the availability of human and financial resources, as well as technological capabilities.

An important distinction can be made between methods intended to support *preventive risk management* and those intended to select *reactive responses* to mis/disinformation in an

ongoing crisis (cf. Boin et al., 2010). In the former, preventive setting, conditions for conducting extensive studies are better, as much more time and effort can be devoted to the analysis. Such studies can, for example, be conducted by synthesizing a broad range of documents to extract a systematically structured set of value-focused objectives that can then be leveraged to support subsequent decision-making (see, e.g., Simon et al., 2014). Conversely, during an ongoing crisis, pressure to react quickly is likely to limit the breadth (scope of topics) and depth (level of detail in exploring these topics) of the analysis. This notwithstanding, it may be possible to lay a stronger foundation for reactive risk management through structured decision processes that are rehearsed in advance with decision-makers and, where appropriate, external stakeholders. In defining such processes, one can draw on experiences from other contexts, such as selecting remediation options in nuclear risk management (see, e.g., Geldermann et al., 2009).

In the context of mis/disinformation, multicriteria methods, most notably multicriteria decision analysis (MCDA) (see, e.g., French, 1993; Greco et al., 2016), are helpful because they enable the systematic evaluation of risk management actions, taking into account the various objectives involved in developing a sound risk management plan and decision-makers' preferences concerning the attainment of these objectives. These can encompass a broad range of impacts, such as limiting the size of population segments reached by mis/disinformation or mitigating its effects on their behavior. The range of alternative actions can be equally wide. For example, these actions can seek to enhance technical capabilities for detecting mis/disinformation, impose tighter regulatory restrictions, or raise awareness through targeted educational campaigns.

In general, the formulation of alternative actions from which risk management strategies are built can be guided by frameworks for societal resilience. The dimensions of such frameworks can represent, for example, specific sectors in which actions are applied (e.g., healthcare, finance, transportation) or different phases of the risk management process (e.g., identification, implementation, monitoring). Conceptually and operationally, these frameworks can serve as checklists to ensure that the selected risk management actions adequately cover these dimensions (see, e.g., Rød et al. 2025 and references therein), recognizing that best strategies typically involve combinations (portfolios) of actions that must be implemented together to ensure effectiveness. For example, the ability to identify mis/disinformation does not suffice unless accompanied by actions that limit its impacts.

As a result, methods of multicriteria portfolio decision analysis (see, e.g., Salo et al., 2010; Liesiö et al., 2021) can help assess the performance of different strategies with respect to relevant objectives, such as minimizing the extent to which mis/disinformation propagates or affects those exposed to it. Specifically, these methods make it possible to determine which strategies, defined as portfolios of actions, offer the most value in supporting the

attainment of these objectives, subject to relevant resource and other constraints that may apply. This can be achieved by constructing a multicriteria model that, on the one hand, captures decision-makers' preferences for criteria representing the attainment of multiple objectives and, on the other hand, aggregates the evaluations of strategies with regard to these criteria into an overall measure of value, thereby identifying the strategy for which this value is highest.

Against the above background, the methodological contribution of this paper lies in proposing a systemic approach that combines multicriteria portfolio decision analysis with models describing the uncertain propagation of mis/disinformation in networks of information users who may be affected by or spread such information. Our case study highlights the need for such an approach, as the impacts of risk management actions often depend on what other actions are implemented. Specifically, the impact interactions among the actions is captured by the propagation model, while the attainment of overall risk-mitigation objectives (such as minimizing exposure to mis/disinformation) is assessed with the multicriteria model. Furthermore, we relate different types of risk management actions to a capability framework for societal resilience, noting that using this framework as a checklist can stimulate the development of additional risk management actions.

Because the proposed approach helps identify cost-efficient risk management strategies to counter mis/disinformation, it can guide investment planning. It also allows these strategies to be optimized under different scenarios that reflect alternative assumptions about the future. These two uses constitute contributions to the future research directions outlined by Hunt (2020).

The rest of this paper is organized as follows. Section 2 briefly summarizes the key characteristics of multicriteria methods, and Section 3 reviews selected case studies in which these methods have been used to counter mis/disinformation. Drawing on the work of the Finnish COVID-19 Science Panel, Section 4 outlines a capability framework for promoting societal resilience and interprets it in the context of mis/disinformation. Section 5 presents a case study showing how multicriteria portfolio decision analysis can be combined with probabilistic models of disinformation propagation to formulate cost-efficient risk-mitigation strategies. Section 6 discusses possible extensions of the proposed approach, and Section 7 concludes.

2. Multicriteria methods

In the literature and professional practice, there is a wide variety of multicriteria methods (see, e.g., Greco et al., 2016), which differ in their theoretical foundations and in the procedural steps for eliciting and synthesizing preference information. As a rule, one key

rationale for using multicriteria methods is to improve the quality of decision-making, for example, by promoting adherence to reasonable axioms of rational decision-making, such as the independence of irrelevant alternatives (Keeney and Raiffa, 1976; French, 1993). Against this background, we advocate sound methods that adhere to widely embraced axioms of rationality.

This is the case with multi-attribute value theory (MAVT; see, e.g., Belton, 1998; French, 1993), which is based on the explicit articulation of the values (e.g., ensuring safety) that decision-makers seek to realize and can therefore be viewed as objectives. Specifically, the attainment of each objective is assessed by specifying a corresponding criterion, which is then used as an attribute to measure the performance of all alternatives. Mathematically, the structure of MAVT models is, in its simplest form, given by the additive sum

$$V(x) = \sum_{i=1}^{n} w_i \, v_i(x_i) \quad , \tag{1}$$

where $V(x)$ is the overall value of alternative $x = (x_1, \ldots, x_n)$, characterized by its performance levels x_i on each of the n criteria; w_i is the criterion weight associated with the i-th criterion; and $v_i(x_i)$ is the score that is attached to the performance level x_i by employing a single-attribute value function $v_i(.)$ from the domain of performance levels to non-negative real numbers. The scores are usually normalized to the interval [0,1], following the convention that the score for the worst performance level is 0, and that for the best performance level is 1.

By convention, the weights are w_i positive and are usually normalized to sum to one, i.e., $\sum_{i=1}^{n} w_i = 1$. Thus, given the additive structure of equation (1), the weight w_i represents the additional contribution to the overall value that can be gained by improving the performance on criterion i from its worst level to the best one. In particular, this weight is not an absolute measure of importance but reflects how much additional value is attained by improving performance on this criterion, relative to the improvements that can be made on other criteria (Keeney and Raiffa, 1976). Moreover, the additive value function in equation (1) is compensatory in the sense that weaker performance on some criterion can be compensated through improved performance on other criteria.

Once the additive value function in (1) has been built and its numerical parameters, most notably alternatives' scores $v_i(x_i)$ and criteria weights w_i, have been elicited, the corresponding overall values $V(x)$ can be used to inform decisions by recommending the alternative with the highest overall value. The process of eliciting these parameters, for instance through structured interviews or facilitated workshops, encourages respondents to articulate how the alternatives contribute to the objectives represented by the criteria. This process can offer insights even if equation (1) is not formally employed to prioritize alternatives. As a result, we do not view multicriteria methods merely as an approach to

produce explicit recommendations about choices among alternatives. Rather, we consider multicriteria methods as enablers of systematically structured processes in which alternatives are explicitly defined and evaluated against criteria that represent stated objectives, without requiring that these evaluations are necessarily combined to yield a strict ranking of alternatives.

In general, offering a structured framework for systematically evaluating alternative actions provides process benefits by creating a traceable, transparent record that can be used to justify the chosen actions and revisited later in light of new evidence (Geldermann et al., 2009). This can be a notable benefit that promotes trust and can therefore enhance the positive impact of the formal decision process on the decision outcome. Further benefits (for an overview, see Salo et al., 2021) include the possibility to reach decisions more quickly, for example, by developing and applying suitable decision templates. Consistency can be enhanced through the systematic consideration of all alternatives, whilst collective learning is supported by a constructive dialogue among the participants.

3. Applications of multicriteria methods in countering mis/disinformation

We next summarize selected applications of multicriteria methods in the context of mis/ disinformation. As our selection is necessarily limited in scope, we refer to comprehensive reviews (e.g., de Almeida, 2017 et al.; Maček et al., 2020).

- Komendantova et al. (2021) propose a decision-analytic framework to evaluate tools and systems for combating misinformation on social media. They present a multi-criteria approach, in which preferences are elicited from stakeholders, including policymakers, journalists, and citizens. Based on the experiences using this approach in Austria, Greece, and Sweden, they note that some of the most valued features of misinformation tools include the ability to study the accountability of distributors of false content, the lifecycle and pathways of misinformation spread, and the perceived intention to deceive. The experiences also suggest that cultural and stakeholder differences shape preferences for tools and should thus be accounted for in countering misinformation.

- To support the detection of fake news, Abrar et al. (2023) conducted a comprehensive Multivocal Literature Review based on 161 published and 61 unpublished studies. They identified 14 credibility factors, such as “reporter reputations” and "impartiality," which were then analyzed using statistical tests and the Analytic Hierarchy Process (AHP) to determine the most important factors across different domains.

- Bajpai and Chaturvedi (2024) propose an integrated approach that combines MCDA with epidemic diffusion modeling to prevent the early spread of rumors in social networks. They motivate their approach by noting the need to (1) debunk rumors quickly, (2) reduce delays by starting the diffusion of counter-rumors promptly, and (3) select trusted nodes for spreading the counter-rumors. The identification of influential nodes is supported using a variant of the AHP, while the diffusion of rumors is modeled by extending the Susceptible–Infected–Recovered (SIR) epidemic model. Illustrative examples are provided for six real datasets of social networks.

- Luo et al. (2023) examine the reputational and economic losses that companies may face due to the spread of negative online content in complex public-opinion environments. They present an extensive framework that combines (i) a multicriteria approach to determine indicator weights for detecting such content; (ii) probabilistic hesitant fuzzy sets to account for uncertainties in experts' judgments about these weights (see also Viviani and Pasi, 2017); and (iii) a compromise solution method to evaluate possible pathways of information spread. They demonstrate their framework by prioritizing information pathways in a case study from the catering industry, but do not explicitly address how to handle these pathways.

- Da Costa et al. (2025) design a structured decision-making framework for evaluating and choosing countermeasures against fake news and disinformation. Specifically, they view fake news as a form of digital risk to be addressed through interventions whose attractiveness and effectiveness enforcement and policy actors can assess. The framework systematically considers multiple criteria, such as costs and impact on risk reduction, when developing strategies in response to high-impact disinformation events. In particular, effective mitigation requires adequate risk prioritization and targeted resource allocation.

- Ganin et al. (2020) note that tools for assessing cybersecurity risks often fail to fully incorporate the three core components (threats, vulnerabilities, and consequences) that need to be considered when managing risks across the physical, informational, and social domains of cyber systems. Motivated by this insight, they propose an MCDA model to quantify these three components and compare cybersecurity management strategies. They also describe a hypothetical case study on evaluating five different countermeasures. The model offers a transparent, value-based process that aligns chosen strategies with stakeholders' preferences.

- Nielsen et al. (2019) examine the implications of potential 'fake' information in decision analysis, emphasizing the need to treat all information, whether fake or real, using the same Bayesian decision analysis approach. Specifically, they propose a decision-analytic approach that explicates multiple system models. Their framework helps assess the value of obtaining better information, for example, by removing false data. This can provide useful insights, even if the pressing realities of on-going crises may constrain the ability to explicate and validate such system models.

To assess the potential for managing mis/disinformation risks, it is instructive to examine realized risks with similar characteristics. In this regard, the COVID-19 pandemic is relevant, as false claims about vaccine safety were a major concern for health care leaders, and individual choices to refuse vaccination posed significant risks to the population (see, e.g., Vujovich-Dunn et al., 2021; Lazarus et al., 2022). Furthermore, in both cases—pathogens and mis/disinformation—the processes leading to harmful consequences have been analyzed using similar concepts, such as contagiousness, susceptibility, infection, and recovery (cf. Jamalzadeh et al., 2022). Even the notion of an incubation period, which has a clear meaning in epidemiology, can be interpreted as the delay between the spread of mis/disinformation and its impact on behavior (see, e.g., Lv et al., 2022).

Against this background, Dillon et al. (2023) offer an enlightening review of how decision analysis can guide decision-makers during an ongoing pandemic by characterizing key uncertainties, eliciting decision-makers' preferences, and systematically evaluating alternative courses of action. However, they focus primarily on individual decisions, providing case-by-case guidance for each decision. This is useful for preparing decision templates that can be documented in advance and used during training to improve the consistency and comprehensiveness of decisions regarding alternative risk management actions in an eventual crisis.

Nonetheless, focusing on individual actions has limitations for addressing the complex relationships among them, because it does not recognize that the strategy to be implemented will most likely comprise many actions. In particular, in an ongoing crisis, the de facto strategy can be viewed as the realization of interconnected decisions taken across various levels of administration, under significant time pressure and based on incomplete, ambiguous, and highly contested information. Furthermore, in a crisis, the pressure to act can be overwhelming, leaving less time and resources for detailed analysis. From this perspective, it is instructive to consider how risk management actions can be analyzed jointly with portfolio decision analysis against the backdrop of general frameworks of societal resilience that draw attention to interdependencies among risk

management actions and the ensuing need for adequate coordination among them (see, e.g., Agarwal and Alsaeedi, 2020; Raetze et al., 2021; Westover, 2024).

4. A capability framework for societal resilience

The concept of resilience, and more specifically societal and social resilience, has been examined from complementary perspectives that span a range of capacities needed in the face of disruptions. Keck and Sakdapolrak (2013), for example, distinguish between coping, adaptive, and transformative capacities, which, respectively, refer to the abilities to overcome adversity, to learn from past experiences, and to adjust to future challenges through institutions that foster sustainable robustness. Lorenz (2013) pinpoints the need for participative capacity, which refers to the system's ability to change its own structure in response to interventions by other systems. Lucini (2014) highlights the capacities, abilities, and competencies needed to cope with stress, a crisis, or a disaster and to bounce back to preexisting conditions of life. Moya and Goenechea (2022) maintain that social resilience must be conceived as a dynamic, multi-level, and evolutionary process to help societies adapt and transform themselves. In general, these perspectives share many similarities: they recognize the need to absorb and learn from shocks and to pursue transformative change to be better prepared for future crises in specific contexts. For example, Hochrainer-Stigler et al. (2020) present a standardized community resilience framework tailored to measuring disruptions due to flooding.

It is instructive to examine mis/disinformation risks against the backdrop of resilience frameworks that help societies prepare for, endure, and recover from them. For example, Nasery et al. (2023) adapt the Security Action Cycle model (Straub and Welke, 1998) to the context of fake news on social media by distinguishing the sequential phases of *deterrence*, *prevention*, *detection*, and *mitigation/remedy*. Building on an extensive literature review, they discuss approaches to address the risks of fake news across these four phases, elaborating on strategies, relevant actors, their roles, and research gaps for each phase. They argue that, because of the complexities of how fake news spreads, effective countermeasures need to span multiple disciplines and stakeholder groups. They also stress the need for integrated tools and longitudinal studies to facilitate a shift from isolated tactics to a more systemic, process-oriented approach to conducting analytical studies, which is one of the motivations of the present paper (see also Bondielli and Marcelloni, 2019; Zhang et al., 2019; Westover, 2024).

Rød et al. (2025) provide a comprehensive literature review of research published from 2018 to 2022 on disinformation in highly digitized societies, in which state infrastructure and civil society use integrated digital technologies in daily governance and civic engagement. Many (but not all) of these societies can be viewed as digitally mature

democracies thanks to governing institutions that promote regulatory frameworks to protect democratic values from digital threats (see, e.g., OECD, 2024a). Rød et al. also propose a structured framework that categorizes counter-disinformation measures across multiple dimensions, including *legal/regulatory, educational, political/governance, psychological/social-psychological*, and *technological* factors. This framework can serve as a practical tool to evaluate the maturity of counter-disinformation efforts and, more specifically, to help stakeholders benchmark and develop strategies for improved resilience.

These two frameworks are complementary. The framework in Nasery et al. (2023) has a strong temporal 'process' focus on the operational management of risks related to false news in the four different phases. In contrast, the framework of Rød et al. (2025) adopts a more static, 'vertical' approach in which the preconditions of and contributors to societal resilience are itemized separately within each sectoral dimension.

As a complement to these two frameworks, we next outline the capability framework for societal resilience developed by Finland's COVID-19 Science Panel. This framework differs from those above in its emphasis on *capabilities* (rather than phases or sectors in society) that may be needed to navigate dynamic phenomena characterized through interactions between communities, social contagion, significant uncertainties, and interdependent decision options. Although originally motivated by the acute needs of managing the COVID-19 pandemic, this framework is generic in providing a transferable tool for examining and structuring capabilities that strengthen resilience by helping societies anticipate, absorb, and recover from disruptions. Here, for the purposes of this paper, capabilities are understood as the *planning and execution of specific risk management actions that contribute toward these aims* (for a review of related definitions of capabilities, see Lindbom et al., 2015).

Administratively, the COVID-19 Science panel was appointed by the Prime Minister's Office in April 2020 to support the Finnish Government in managing the pandemic. It consisted of 13 leading researchers from different fields[1] who were asked to synthesize information to support recovery planning by offering recommendations in its 115-page main report, aimed primarily at the Government but also at many other institutional actors and stakeholders, including individual citizens (Government of Finland, 2020).

[1] These included fields such as epidemiology, psychology, economic policy, and environmental science. Resilience as a field was represented by the first author.

Figure 1. The four pillars in the capability framework for societal resilience.

Specifically, this framework (see Figure 1) features four dimensions, referred to as capability pillars (see also Kontogiannis et al., 2017), which represent the different kinds of capabilities that are needed to prepare for and navigate societal crises even in the presence of high stakes, limited information, pressing schedules, and major uncertainties about the effectiveness of countermeasures and their broader consequences:

1. **Maintain forward-looking alertness** by establishing a timely, comprehensive foundation consisting of validated information for decision-making.
2. **Enact decision-making processes** that build on this foundation to help decision-makers prepare and choose among alternative risk management actions based on a coherent, systematic evaluation, which may involve stakeholder consultations.
3. **Implement risk management actions** effectively to support the attainment of objectives for chosen actions.
4. **Interact with stakeholder**s through purposeful, continual, participatory, and multidirectional communication efforts that build trust, shape risk perceptions, and aid the successful implementation of risk management actions, such as encouraging voluntary collaboration among stakeholders.

We next elaborate on these dimensions in the context of mis/disinformation.

I Maintain forward-looking alertness

This first capability pillar refers, in a broad sense, to the ability to build and maintain required information resources for building up-to-date, balanced, comprehensive, and forward-looking situational awareness and alertness as a basis for decision-making. It includes data resources and analytical tools to help detect mis/disinformation, but it also covers other sources of information, such as solicitation of expert judgments and even

laypeople's opinions on the susceptibility of specific population segments to targeted disinformation campaigns. Because timeliness is a key concern, there may be a need to conduct continuous horizon scanning to monitor the environment and identify relevant trends, observations, or hypotheses about mis/disinformation that, after validation, can be added to the approved pool of information resources.

The identification of weak signals (see Hines et al., 2019 for a literature review) can be supported by a systematic multicriteria assessment based on criteria such as relevance, novelty, and probability of occurrence (see, e.g., Könnölä et al., 2012). Concretely, building on the recognition that weak signals are likely to be those that are given a high score on such criteria by some but not all respondents, the computation of dispersion metrics such as the variance of the respondents' scores on the criteria can be used as an indicator of whether a proposed signal could be viewed as a weak one. For a detailed case study on such an approach, applied to the generation of innovation ideas, see Könnölä et al. (2007).

II Enact decision-making processes

The second capability pillar consists of the capabilities required to integrate relevant information resources into structured decision-making processes. Apart from covering the characteristics of the specific occurrence of mis/disinformation and its potential impacts, these resources need to include information about the risk management actions currently available, or those that can be developed. The latter perspective is vital, as focusing only on mis/disinformation, without links to actions that may be taken to curtail it, will not limit harmful consequences.

By design, multicriteria methods strengthen this capability pillar by helping make choices among available courses of action (cf. Komendantova et al., 2021). When selecting actions, constraints often arise, for example, from the availability of scarce resources or specialized expertise. Such constraints affect the feasibility of selected actions, which may compete for resources from the same pool. Here, methods of portfolio decision analysis (see, e.g., Salo et al., 2010; Liesiö et al., 2021) can be applied to select which actions to implement jointly to build a portfolio that performs as well as possible with respect to relevant criteria, subject to resource availability and other constraints. Such analyses can also be used to identify what resources would be needed to attain performance targets, or what kinds of new actions, enabled by technological advances, for instance, would complement those actions that have been available earlier.

III Implement risk management actions

This capability pillar refers to the operational implementation of the selected actions as planned. However, as the effectiveness of the actions may involve uncertainties, it is important to monitor the targeted effectiveness as well as possible impacts more generally, and if necessary, to make adjustments 'along the way' (for example, improving measures that aim to inoculate the population against mis/disinformation; see Bessarabova et al., 2024), rather than choosing irreversible strategies that could have unintended effects (such as denying all access to social media platforms, which could lead to widespread resentment and mistrust of authorities).

IV Interact with stakeholders

This fourth capability pillar links to those above. It is a vital part of (i) raising broader societal awareness of mis/disinformation, which is ultimately influenced by interactions among different stakeholder groups in society; (ii) identifying effective risk management actions by involving stakeholders who often have a better 'grass-root' level understanding of information sharing patterns than administrators and policymakers; (iii) enabling commitment and successful implementation of these actions through collaboration among key societal stakeholders (e.g., willingness to alert peers about false statements within one's social networks).

Interactions related to this capability pillar should not be viewed as one-way, based on the idea that 'correct' information comes from and is shared by policymakers and administrators. Instead, when considered alongside the other three capability pillars, these interactions can more fruitfully be understood as ongoing, multi-directional, participatory activities that span local communities and social networks across different time horizons to enhance resilience to mis/disinformation, for example by fostering cohesion and strengthening adaptive and transformative capacities ex-ante as well as coping capabilities ex-post (see, e.g., Aldrich and Meyer, 2014; Keck and Sakdapolrak, 2013; Lorenz, 2013; Townshend et al., 2015). Furthermore, because mis/disinformation can cause greater harm when believed and acted upon, activities that focus on informing, educating, and empowering stakeholder groups about the risks of mis/disinformation are essential (Carmi et al., 2020). In low-trust societies, the successful implementation of such activities may pose unique challenges, as administrative initiatives are often viewed with greater suspicion (see, e.g., Löfstedt, 2005; Lenton et al., 2022).

The systemic view

Because the four capability pillars represent distinct capabilities that can be manifested through execution of different kinds of actions, part of the value of this framework lies in

examining them together holistically. This can be especially important in organizations where the processes of information gathering, decision-making, implementation, and communication span across administrative silos, as is often the case in public administrations.

In this regard, the capability framework can be harnessed as a checklist to ensure that all capability pillars receive due attention. It can also serve as a basis for a structured assessment process, in which possible actions associated with the different capability pillars are first elaborated and then evaluated against the aims of respective pillars, to derive indicators that reflect the maturity of risk management capabilities. Such an evaluation can be carried out using measurement scales as in multicriteria methods (for an example, see Youshuira et al., 2023).

This notwithstanding, the above capability framework should not be viewed as an MCDA model that represents objectives in the same way as the additive model in equation (1). By construction, this additive model assumes that low performance on a criterion reflecting one objective can be compensated for by higher performance on another criterion. In contrast, this assumption of additivity does not hold when evaluating capabilities. To see this concretely, if the capabilities to prevent the spread of mis/disinformation are entirely lacking, this deficiency cannot be compensated for by improved capabilities for maintaining alertness. Conversely, alertness alone does not suffice, unless accompanied by effective means to counter mis/disinformation through appropriate actions.

Accordingly, the capability framework in Figure 1 should more aptly be seen as a tool for multicriteria mapping, which is a reflective and participatory approach for systematically addressing complex problems from multiple perspectives with the aim of developing an improved shared understanding (see, e.g., Kuipers-Dirven et al., 2023; Stirling and Coburn, 2006). Compared with MAVT models, multicriteria mapping is less normative, as its benefits often arise from a structured dialogue about the range of policy options rather than from clear-cut decision recommendations (White, 2017). This was also the case in the COVID-19 Science Panel, as this framework catalyzed constructive discussions during joint meetings between the Panel and the Chiefs of Staff from different ministries.

Many of the responses to mis/disinformation in the Government's Report on the Future (Government of Finland, 2025) can be examined in light of the capability framework in Figure 1. For example, responses that enhance citizens' skills in identifying mis/disinformation contribute to heightened awareness (capability pillar I); the development of proactive regulation is an enabler of timely decision-making processes (capability pillar II); enhanced data protection measures and improved technical capabilities support the implementation of effective risk management actions (capability

pillar III); and stakeholder interactions that promote media literacy and uphold the value of scientific thinking and research-based information lessen the impacts of mis/disinformation among recipients (capability pillar IV).

Against this backdrop, Figure 2 shows how the proposed approach of multicriteria portfolio decision analysis builds on this framework. Here, the labels attached to the arrows characterize relationships between adjacent elements: for instance, capabilities enable the execution of risk management actions. This figure clearly distinguishes individual risk management actions from strategies, because strategies typically consist of several actions. It also highlights that multicriteria evaluation should focus on the consequences associated with strategies (instead of individual actions, because these may exhibit strong interactions). In predicting these consequences, computational models representing the emergence, propagation, and impact of mis/disinformation can be helpful. For example, they make it possible to assess the benefits of introducing the possibility to implement new risk management actions in addition to those that are already available, as illustrated by the following case study.

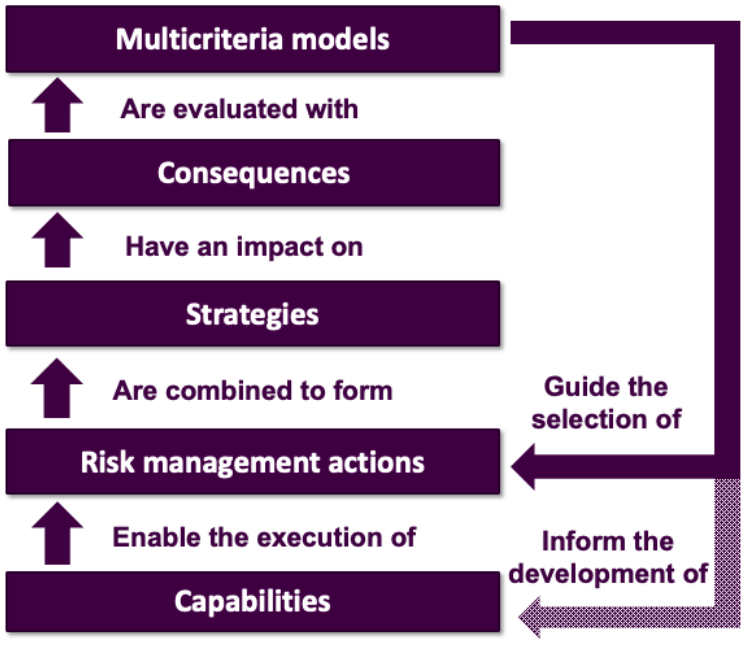

Figure 2. The main elements of the proposed multicriteria approach to evaluating strategies consisting of risk management actions.

5. A case study on mitigating the impact of disinformation on voting

The following case study illustrates how multicriteria portfolio decision analysis can be integrated with other analytical techniques, such as Bayesian modeling of voting behavior

and cascade models for describing its propagation. This study, downsized and streamlined for transparency, builds partly on a real-world scenario analysis engagement[2] conducted in collaboration with authorities in preparation for national parliamentary elections.

Consider a scenario in which a municipal council consisting of 17 voters decides by majority vote whether to grant a permit for the construction of a contested production facility. At the outset, slightly more support exists for granting this permit. However, disinformation posted on one of three social media platforms can undermine this support by propagating to voters through the connections in Figure 3 (cf. Schneider et al., 2013). Before the vote, one platform, each with a one-third probability, will be targeted with disinformation containing false claims that speak against granting the permit.

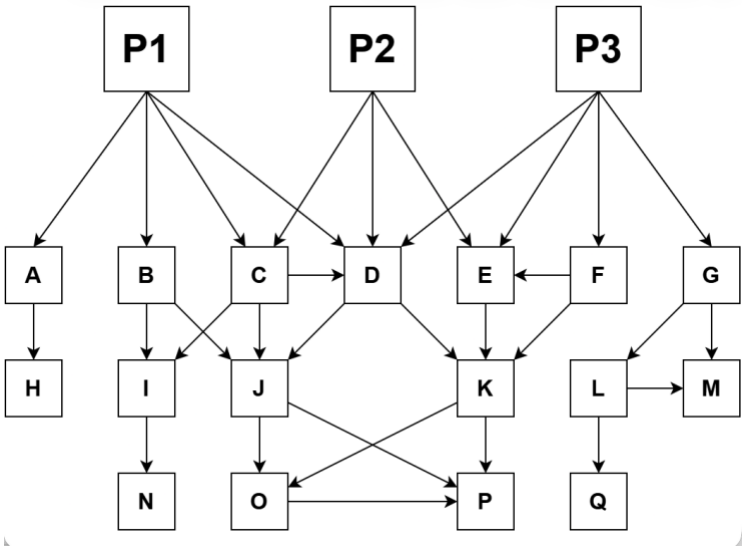

Figure 3. Network of platforms (large nodes at the top), voters (small nodes), and communication connections among platforms and voters (directed arrows).

The propagation of disinformation in Figure 3 is modeled by the independent cascade model (see, e.g., Wang et al., 2012; see Appendix A for details). If enough voters receive it and are influenced by it, the council's decision may be tipped toward permit denial. To counter the possibility of this inadvertent shift, there are three types of risk management actions:

1. *Monitoring*: Each platform can be monitored to detect disinformation.
2. *Isolation*: The attacked platform can be isolated to prevent the propagation of disinformation from it.

[2] The scenario analysis was carried out in the research project "Adversarial Risk Analysis for Impact Assessment", funded by the Scientific Advisory Board for Defence in Finland (see Roponen, 2023).

3. *Education*: Voters can be informed through an education program that reduces the impact of disinformation on voting behavior.

We assume that actions to monitor platforms and educate voters can be implemented at two levels of effectiveness. For monitoring, the less effective level costs 1 unit, and the more effective level costs 2 units. For education, the costs for the two levels are 2 and 4, respectively. Even if a platform is monitored and attacked, disinformation may go undetected, but there are no false indications of disinformation if the platform has not been attacked. In response to a positive monitoring indication, the attacked platform can be isolated for a cost of 8 units, payable up front. Thus, the costs arising from the choice of actions are not uncertain.

Note that the three types of risk management actions above can be related to the pillars of the capability framework in Figure 1. Specifically, monitoring attacks on the platforms helps *maintain enhanced awareness*; isolating platforms is a means to *implement risk management actions*; both variants of educating voters call for the capability to *interact with stakeholders*; and combining these elements to prepare a systemic analysis with ensuing recommendations helps *enact decision-making processes*.

Specifically, voting behavior follows the beta distribution (see, e.g., Rigdon et al., 2009) in Figure 4 so that the impact of education is modeled by adjusting likelihood ratios (see Appendix A). The prior distribution of voters' initial probabilities, shown in blue, implies that the permit is granted with probability 0.572. The three other distributions, indicated by dashed lines, show the distributions for disinformed voters who have received either no, limited, or substantial education. Note that the distribution for more educated voters deviates less from the initial blue distribution.

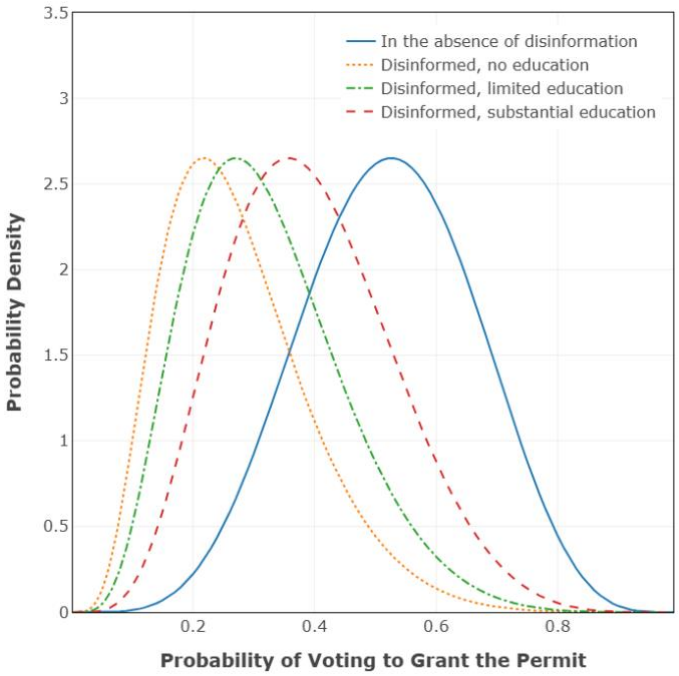

Figure 4. Beta-distributed voting behavior based on disinformation and education status.

The three main objectives in this scenario are to:

(i) minimize the costs incurred due to the implementation of actions,

(ii) minimize the probability of disinformed voters (meaning that at least one voter is exposed to disinformation), and

(iii) minimize the impact of disinformation by lowering the probability that the permit is denied as a result of it.

In this example, there are $27 \times 2 \times 3 = 162$ different strategies, each a portfolio of several risk management actions. Specifically, there are 3 platforms with 3 monitoring possibilities for each (i.e., no monitoring = 0, light monitoring = 1, close monitoring = 2), which gives 3 x 3 x 3 = 27 options; the action to isolate the attacked platform is binary, which gives 2 options; and, finally, there are 3 options in providing education (i.e., no = 0, limited = 1, or substantial education = 2).

Of these 162 strategies, it is of interest to identify those that are Pareto-optimal, in the sense that there is no other strategy that would perform as well on all objectives and, in addition, be strictly better on at least one of them.

To this end, we ran Monte Carlo simulations for each strategy to estimate its expected consequences with respect to the three objectives; see Figures 5 and 6. Out of the 162 strategies, only 11 are Pareto-optimal in keeping with the above definition.

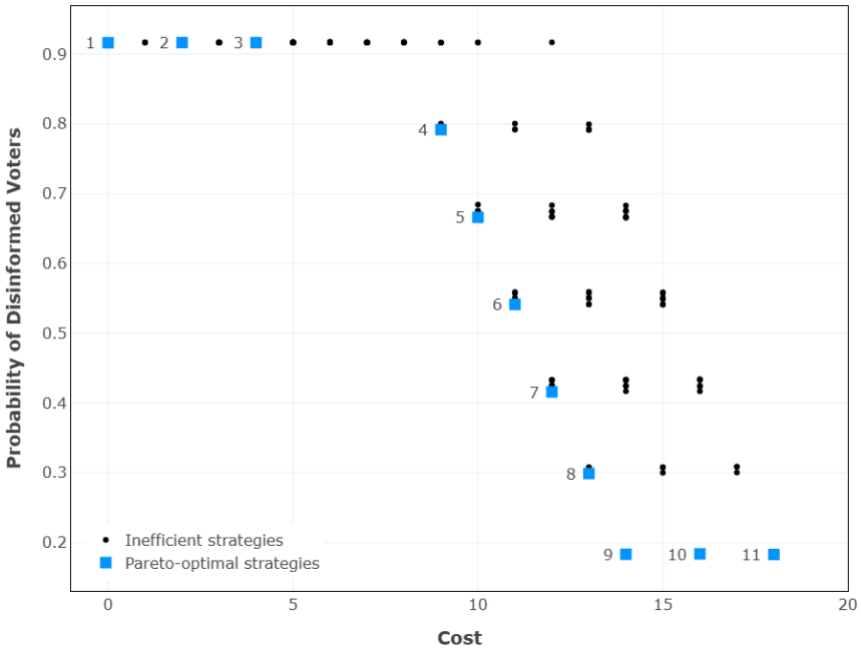

Figure 5. Costs and probabilities of disinformed voters for all strategies. The labels correspond to the rows in Table 1.

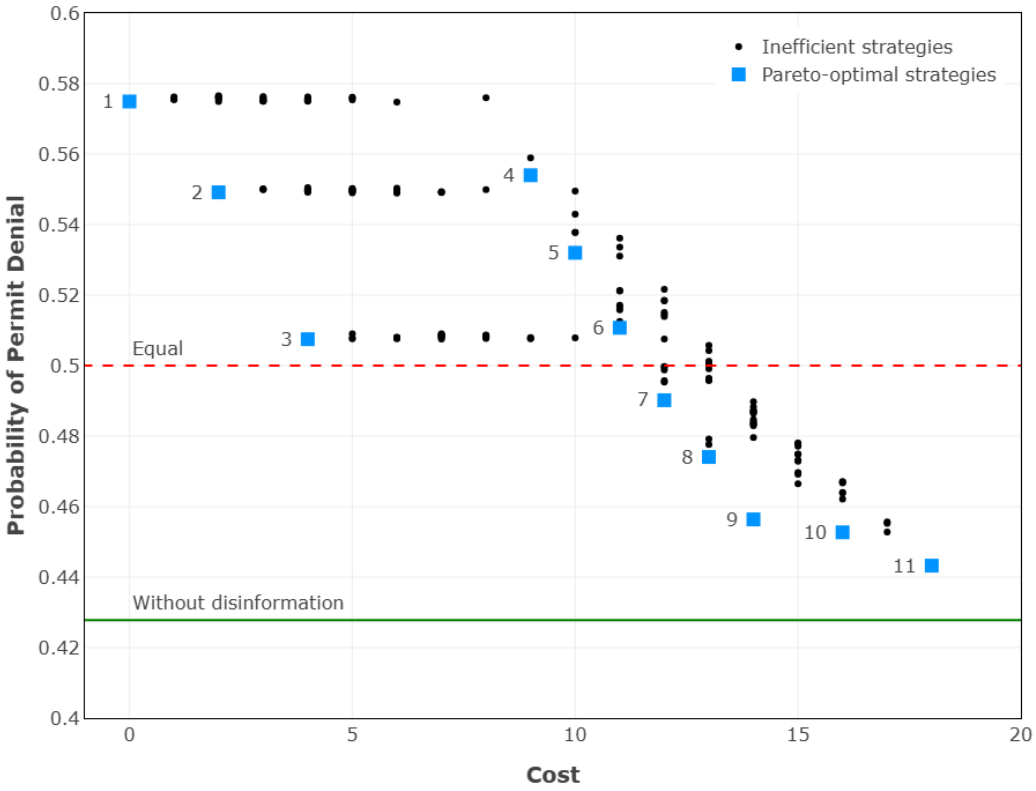

Figure 6. Costs and probabilities of permit denial for all strategies. The labels correspond to the rows in Table 1.

Figures 5 and 6 show the trade-offs between the costs of risk management actions, the probability of disinformed voters, and permit denial. For the strategies above the red dashed line at 0.5 in Figure 6, permit denial is likely. The four strategies #2, #3, #10, #11 and the three strategies #4, #5, #6 are Pareto-optimal, but they do not appear on the Pareto fronts in Figures 5 and 6, respectively, because these figures show only two of the three objectives (recall that Pareto-optimality is defined with respect to *all* three objectives). Furthermore, some Pareto-optimal strategies are unsupported, as they are dominated by linear combinations of other strategies. This is the case, for example, for strategy #7, which lies in Figure 5 and Figure 6 above the line segments connecting strategies #3 and #9.

#	Monitoring			Iso-lation	Edu-cation	Cost of strategy	Probability of disinformed voters	Probability of permit denial	Overall value (equal weights)
	P1	P2	P3						
1	0	0	0	0	0	**0** (1.000)	**0.916** (0.000)	**0.575** (0.000)	0.333
2	0	0	0	0	1	**2** (0.943)	**0.916** (0.000)	**0.549** (0.023)	0.322
3	0	0	0	0	2	**4** (0.882)	**0.916** (0.000)	**0.507** (0.211)	0.365
4	1	0	0	1	0	**9** (0.707)	**0.792** (0.170)	**0.554** (0.014)	0.297
5	2	0	0	1	0	**10** (0.667)	**0.666** (0.341)	**0.532** (0.074)	0.361
6	2	0	1	1	0	**11** (0.624)	**0.541** (0.512)	**0.511** (0.189)	0.441
7	2	0	2	1	0	**12** (0.577)	**0.416** (0.682)	**0.490** (0.359)	0.540
8	2	1	2	1	0	**13** (0.527)	**0.299** (0.842)	**0.474** (0.538)	0.636
9	2	2	2	1	0	**14** (0.471)	**0.183** (1.000)	**0.456** (0.784)	0.752
10	2	2	2	1	1	**16** (0.333)	**0.183** (1.000)	**0.453** (0.841)	0.724
11	2	2	2	1	2	**18** (0.000)	**0.183** (1.000)	**0.443** (1.000)	0.667

Table 1. Pareto-optimal strategies, actions contained in them, costs, probabilities of disinformed voters and permit denial, scores (in parentheses), and overall values.

The composition of Pareto-optimal strategies and their consequences (in boldface) are shown in Table 1. Analyzing the composition of Pareto-optimal strategies can yield further insights. For example, strategies that cost at most 4 units consist solely of educational actions and will not reduce the probability of permit denial below 0.5. In Figure 5, the many dots to the right of strategies #1, #2, and #3 include actions in which platforms are monitored to varying degrees of effectiveness without isolating them, but this enhanced monitoring as such does not lower the probability of permit denial. This illustrates that consequences must be analyzed systemically, because the impacts of additional investments in some risk management actions depend on what other actions are implemented.

To reduce the probability of a permit denial below 0.5, a combination of monitoring and isolation actions must be deployed, and the attacked platform must be isolated. Across Pareto-optimal strategies #7 through #11, platforms 1 and 3 should always be monitored closely (at level 2), but platform 2 may be monitored at any one of the three possible levels. Also, because disinformation from platforms 1 and 3 can reach all voters, monitoring them together is more effective than monitoring either of the other two platform pairs (platforms 1 and 2, platforms 2 and 3). Even if all risk management actions are implemented, at a total cost of 18 units, the probability of permit denial will not be reduced to the baseline level of 0.428 = 1 – 0.572 (where 0.572 is the initial probability of granting the permit, implied by the distribution in Figure 4). This is because monitoring does not give a fully reliable indication of the disinformation attack.

Isolation actions are essential for reducing the probability that voters become disinformed. This is because education only affects the impact of disinformation, not its propagation. For example, consider strategies #3 and #4. Strategy #3, which consists of education actions, outperforms strategy #4 as it has a lower probability of permit denial. Strategy #4, which consists of monitoring and isolation actions, has a lower probability of disinformed voters because it helps prevent the propagation of disinformation from the attacked platform. However, because monitoring does not provide a fully reliable indication of the disinformation attack on one platform, the probability of disinformed voters cannot be reduced below 0.183, compared with the initial probability of 0.916 when no isolation actions are implemented.

The information in Table 1 can be further synthesized by with the help of value functions, defined on the range of impacts, and the criteria weights in equation (1). For illustration, assume that the functions are power functions $v_i(x_i) = {x_i}^{r_i}$, where $r_i > 0$ is the exponent and $x_i \in [0,1]$ are the normalized performance scores, obtained through a linear mapping of the worst and best levels of the objectives to 0 and 1, respectively. Here, the worst and best levels can be readily defined for each objective from the Pareto-optimal strategies. Thus, for strategy #11 whose cost is 18, the normalized performance score is 0, and 1 for the "do nothing" strategy 1 which does not cost anything. Furthermore, assume that the single-attribute value score associated with the cost level 10 is 0.667 (representing a preference for low costs, as two thirds of the maximum value on the cost criterion is gained by limiting costs to 10), and that the single-attribute value score associated with the probability of disinformed voters is linear. Finally, the single-attribute score associated with probability 0.5 of permit denial is 0.2 (representing a preference for small changes in election results, given that only a fifth of the maximum value on this criterion is gained by lowering it from 0.567 to 0.5). Thus, the respective exponents are $r_1 = 0.5,\ r_2 = 1$ and $r_3 \approx 2.32$. These assumptions imply the scores in parentheses in columns 7, 8, and 9 of Table 1.

Furthermore, if the criteria have equal weights in the additive value function representation in equation (1), meaning that lowering costs from 18 to 0 contributes as much to the overall value as reducing the probability of disinformed voters from 0.916 to 0.183, and reducing the probability of permit denial from 0.567 to 0.442, the scores can be combined to obtain the overall values in the last column in Table 1. In this case, strategy #9, which consists of close monitoring of all platforms and platform isolation without providing education, would have the highest overall value and be the recommended strategy.

The model can also be used to conduct sensitivity analyses on how changes in risk management actions would affect strategy recommendations. For example, assume that the effectiveness of the more substantial education program can be improved so that, for voters exposed to disinformation, the ratio between the probabilities of voting in support of granting the permit versus voting against the permit is reduced only by 20% instead of 50% (i.e., the parameter $\Delta(2)$ in Appendix A changes from 1/2 to 4/5). Then, such education *alone* would lower the probability of permit denial to 0.454, which is better than 0.456 for strategy #9. Yet, the overall value function with equal criterion weights would still recommend strategy #9, because education does not in any way limit the spread of disinformation, and strategy #9 has the lowest probability of disinformed voters. However, if the weight of the probability of disinformed voters were less than 0.243 (with the remaining weight divided equally between the other two criteria), the improved substantial education would become the recommended strategy.

More generally, one could also explore the impacts of actions that would weaken the communication connections shown in Figure 3. Platform isolation is representative of such an action, as it severs the connection between the attacked platform and its voters. For example, one could assess the impacts of weakening but not fully cutting communication connections between voters. The attractiveness of doing so would need to be evaluated by accounting for the full set of possible actions (and their costs) from which strategies can be built.

Methodologically, the novelty of this case study lies in its span of different types of risk management actions, including those focused on specific stages of the process (such as monitoring), localized interventions (such as platform isolation), and broader cross-cutting impacts (such as education). This enables more systematic analyses and can also support the development of additional risk management actions that, when added to those previously introduced, would complement them effectively. From this perspective, the proposed approach of multicriteria portfolio decision analysis contributes to two of the four future research directions identified in Hunt (2020), as it helps inform investment planning in mitigating mis/disinformation and makes it possible to optimize strategies for scenarios that reflect such propagation.

In the case study above, the focus was on the impacts of disinformation on a collective vote to grant a permit for a physical production facility. However, because the proposed modeling approach can (i) track the disinformation status of all information users and (ii) characterize how this status affects their behavior, it can be extended to situations where users make other kinds of decisions, such as which transportation routes they take and how much critical resources they consume. This makes it possible to link the information model with models representing the physical infrastructures affected by users' decisions. Thus, the consequences layer in Figure 2 has effectively two levels: the information level and the physical level (see Barker et al., 2025). Hence, objectives concerning the performance of the physical level (e.g., maximizing reliability at different locations over time) can be examined to develop corresponding criteria for multicriteria models. In addition, criteria from the information level can also be introduced (e.g., minimizing user exposure to disinformation). The resulting multicriteria models can then be leveraged in much the same way as in the case study above to identify strategies, implemented through portfolios of risk management actions at the information level, that offer the most value in terms of contributing to achieving overall objectives.

6. Discussion

The case study above suggests that countering the risks of mis/disinformation can benefit from a systemic approach that supports coordinated strategy decisions. Specifically, rather than making choices among 'single' actions independently–such as deciding how to monitor a given platform–it is instructive to consider portfolios of complementary risk management actions and to use modeling to account for interdependencies in their impacts. Here, multicriteria portfolio decision analysis (see, e.g., Salo et al., 2010; Liesiö et al., 2021) helps identify Pareto-optimal strategies from the potentially very large number of alternative strategies that can be constructed by combining distinct risk management actions. For example, Żebrowski et al. (2022) present an instructive case study on selecting portfolios of risk management actions against cyberattacks. To accommodate uncertainties in assessing the consequences of the strategies, one can explore ranges of parameter values instead of numerical point estimates, as in our case study (see, e.g., Liesiö and Salo, 2012).

In the case study above, the scenario was that only one platform would be attacked. One could develop many other scenarios depicting alternative futures to identify which strategies perform particularly well in one or more of them. However, given the large number of possible scenarios, this creates a need to choose which scenarios to select (Haimes et al., 2002). Such choices can be assisted by multicriteria methods based on evaluation criteria such as diversity and consistency (Seeve and Vilkkumaa, 2022). One can even introduce scenario probabilities that depend on the actions taken (Vilkkumaa et al., 2025). In contrast, comprehensiveness, referring to the systematic exploration of *all*

possible combinations of realizations of relevant uncertainties (see, e.g., Tosoni et al., 2018), may be an unrealistic aim, as there are extraordinarily many ways in which mis/disinformation can be propagated and used in society.

While our numerical example has many parameters, these could nevertheless be estimated from data analysis (see Jiang et al., 2021; Hunt et al., 2020, 2022 for examples of extensive empirical studies) or through structured expert judgment interviews with specialists, decision makers, or stakeholders (Cooke, 1991). Yet the temptation to build increasingly complex models merely for the sake of analytical sophistication should be resisted. Feedback from the users, especially through longitudinal case studies (see, e.g., Mild et al., 2015), can help understand how models can be best embedded in decision processes.

In our literature review, we encountered many applications of multicriteria methods that have paid little attention to rigor and defensibility or, more formally, the requirements for prescriptive support to rational decision-making (see, e.g., Abrar et al., 2023; Bajpai and Chaturvedi, 2024). One of these is the alignment with decision-theoretic axioms of rationality, such as the transitivity of preferences[3], whilst the exclusion of rank reversals is another requirement for rationality[4]. Methods that violate such requirements and axioms, such as the AHP and DEMATEL, may offer benefits, as applying any structured approach may be better than using none. Still, one can make a strong case for methods with a solid axiomatic foundation, most notably multi-attribute value theory (MAVT) and multi-attribute utility theory (MAUT) (see, e.g., French, 1993). These methods are appealing in that the numerical parameters of decision models, such as alternative assessment scores and criterion weights, have well-defined meanings.

7. Conclusion

Mis/disinformation poses complex, systemic risks that weaken the foundations of evidence-based decision-making and democratic governance. While traditional risk management approaches, which proceed from the identification of failure scenarios to the estimation of scenario probabilities and the assessment of consequences (see, e.g., Kaplan and Garrick, 1981), can be applied to study these risks, mis/disinformation involves unique challenges that arise from the recursive and actor-dependent communication dynamics. For

[3] If alternative A is preferred to alternative B which is preferred to alternative C, then it logically follows that alternative A should be preferred to alternative C.

[4] A rank reversal occurs, for example, when the preference ranking of two alternatives changes as a result of adding the possibility to choose a third alternative which is irrelevant in the sense that it is less preferred than either one of the two alternatives being considered (see, e.g., Salo and Hämäläinen, 1997; Maleki and Zahir, 2013).

example, it may be difficult to obtain reliable numerical parameter estimates for the behavior of adversaries who act surreptitiously. Also, the societal consequences of the propagation of unprecedented false information are harder to predict than those of previously encountered physical hazards that evolve following well-known laws of physics.

Another differentiating feature is that, from the viewpoint of upholding democratic governance and social cohesion (cf. Bennett and Livingston, 2018), mis/disinformation can pose particularly pervasive and subtle threats because those in power may seek to produce and propagate it. In effect, problems arising from mis/disinformation cannot be fully isolated and tackled as an external threat, because they are part of its communication fabric. Here, the close attention that the Finnish Government's Report (Government of Finland, 2025) gives to media literacy, digital citizenship, and broader awareness through educational efforts is pertinent (cf. Carmi et al., 2020). Unless mis/disinformation is perceived to pose risks to society and its administrative apparatus, the associated risks are likely to persist and intensify. In the face of this reality, complacency may be the greatest risk of all.

From the perspective of policy processes, multicriteria methods can be advocated on the grounds that they improve the traceability of decision processes. This can be valuable as the distinction between the "external" risks (caused by natural phenomena such as extreme weather events) and the "internal" administrative apparatus (responsible for countering risks in accordance with approved standards and ethical norms) is not necessarily clear in cases of mis/disinformation, especially in the presence of political forces that may seek to disrupt legitimate governance processes. In effect, this apparatus itself may be permeated with mis/disinformation and take part in its propagation, even to the extent that the trustworthiness of administrative processes will be severely compromised.

Against this backdrop, there is a growing need for analytical rigor, methodological transparency, and stakeholder participation, all hallmarks of well-designed and well-executed decision processes. Beyond their instrumental role in providing structured multicriteria decision support, such processes promote accountability in risk governance by clarifying links among modeling assumptions, resulting decision recommendations, and actual decisions. This may reduce the risk that decisions will be reached on shallow or even arbitrary grounds. Here, methods of multicriteria portfolio decision analysis can play a role in making these processes more transparent, inclusive, and reflective. Thanks to their ability to incorporate a broad range of criteria—spanning technical and social considerations—these methods are well positioned to support the formulation of balanced policy portfolios that go beyond choices among individual risk management options.

Inspired by this recognition, we see the need for further methodological and applied research that seeks to integrate these methods with other analytical approaches.

References

Abrar, M.F., Khan, M.S., Khan, I., ElAffendi, M., and Ahmad, S. (2023). Towards fake news detection: a multivocal literature review of credibility factors in online news stories and analysis using analytical hierarchical process, *Electronics* 12(15), 3280. https://doi.org/10.3390/electronics12153280

Agarwal, N.K. and Alsaeedi, F. (2020). Understanding and fighting disinformation and fake news: towards an information behavior framework, *Proceedings of the Association for Information Science and Technology* 57, 327. https://doi.org/10.1002/pra2.327

Aldrich, D.P. and Meyer, M.A. (2014). Social capital and community resilience, *American Behavioral Scientist* 59(2), 254-269. https://doi.org/10.1177/0002764214550299

de Almeida, A.T., Alencar, M.H., Garcez, T.V., and Ferreira, R.J.P. (2017). A systematic literature review of multicriteria and multi-objective models applied in risk management, *IMA Journal of Management Mathematics*, 28(2), 153-184, https://doi.org/10.1093/imaman/dpw021

Aven, T. and Thekdi, S.A. (2022). On how to characterize and confront misinformation in a risk context, *Journal of Risk Research*, 25(11-12), 1272-1287. https://doi.org/10.1080/13669877.2022.2142950

Bajpai, S. and Chaturvedi, A. (2024). A multi-criteria decision making based integrated approach for rumor prevention in social networks, *Multimedia Tools and Applications* 83(8), 1–26. https://doi.org/10.1007/s11042-024-18419-1

Barker, K., Bessarabova, E., Radhakrishnan, S., González, A.D., Weber, M.S., Ramirez Marquez, J.E., Vorobeychik, Y., and Jiang, J.N. (2025). Risk analysis of disinformation weaponized against critical networks, *Risk Analysis*, published online 7 July. https://doi.org/10.1111/risa.70062

Belton, V. (1998). Multi-criteria problem structuring and analysis in a value theory framework. In: M. Ehrgott, M. and Figueira, J.R. (Eds.), *Multiple Criteria Decision Making: Current State-of-the-Art Surveys*. Springer. 335-366.

Bennett, W.L. and Livingston, S. (2018). The disinformation order: disruptive communication and the decline of democratic institutions, *European Journal of Communication* 33(2), 122-139. https://doi.org/10.1177/0267323118760317

Bessarabova, E., Banas, J. A., Reinikainen, H., Talbert, N., Luoma-aho, V., and Tsetsura, K. (2024). Assessing inoculation's effectiveness in motivating resistance to conspiracy propaganda in Finnish and United States samples, *Frontiers in Psychology* 15, 1416722. https://doi.org/10.3389/fpsyg.2024.1416722

Boin, A., Comfort, L.K., and Demchak, C.C. (2010). *Designing Resilience: Preparing for Extreme Events*, University of Pittsburgh Press. 351 p.

Bondielli, A. and Marcelloni, F. (2019). A survey on fake news and rumour detection techniques, *Information Sciences* 497(3), 38-55. https://doi.org/10.1016/j.ins.2019.05.035

Carmi, E., Yates, S.J., Lockley, E., and A. Pawluczuk, A. (2020). Data citizenship: Rethinking data literacy in the age of disinformation, misinformation, and malinformation, *Internet Policy Review* 9(2), 1-22. https://doi.org/10.14763/2020.2.1481

Cooke, R.M. (1991). *Experts in Uncertainty*, Oxford University Press. 336 p.

da Costa, V.J., Dongo, D.F., and Mira da Silva, M. (2025). Using MCDA to select countermeasures against fake news, *Journal of Information, Communication and Ethics in Society* 23(1), 54-103. https://doi.org/10.1108/JICES-07-2024-0089

Dillon, R.L., Bier, V.M., John, R.S., and Althenayyan, A. (2023). Closing the gap between decision analysis and policy analysts before the next pandemic, *Decision Analysis* 20(2), 109-132. https://orcid.org/0000-0003-4738-7807

French, S. (1993) *Decision Theory: An Introduction to the Mathematics of Rationality*, Ellis Horwood. 448 p.

Ganin, A.A., Quach, P., Panwar, M., Collier, Z.A., Keisler, J.M., Marchese, D., and Linkov, I. (2020). Multicriteria decision framework for cybersecurity risk assessment and management, *Risk Analysis* 40(1), 183-199. https://doi.org/10.1111/risa.12891

Greco, S., J Figueira, J., Ehrgott, M. (eds.) (2016). *Multiple Criteria Decision Analysis: State of the Art Surveys*. Springer. 1356 p.

Geldermann, J., Bertsch, V., Treitz, M., French, S., Papamichail, K.N., and Hämäläinen, R.P. (2009). Multi-criteria decision support and evaluation of strategies for nuclear remediation management, *Omega* 37(1), 238-251. https://doi.org/10.1016/j.omega.2006.11.006

Government of Finland (2020). *Preventing the Harmful Effects of the COVID-19 Crisis in Accordance with the Principles of Sustainable Development*. Report of the COVID-19 Science Panel. https://urn.fi/URN:NBN:fi-fe2020101584121

Government of Finland (2025). *Strategic Operating-Environment Analysis (Government Report on the Future, Part I)*. Publications of the Government 2025:82, Helsinki. http://urn.fi/URN:ISBN:978-952-383-695-2

Haimes, Y.Y., Kaplan, S., and Lambert, J.H. (2002). Risk filtering, ranking, and management framework using hierarchical holographic modeling, *Risk Analysis* 22(2), 383-397. https://doi.org/10.1111/0272-4332.00020

Hines, P., Hiu Yu, L., Guy, R.H., Brand, A., and Papaluca-Amati, M. (2019). Scanning the horizon: a systematic literature review of methodologies, *BMJ Open* 9, e026764. https://doi.org/10.1136/bmjopen-2018-026764

Hochrainer-Stigler, S., Laurien, F., Velev, S., Keating, A., and Mechler, R. (2020). Standardized disaster and climate resilience grading: A global scale empirical analysis of community flood resilience, *Journal of Environmental Management* 276, 111332. https://doi.org/10.1016/j.jenvman.2020.111332

Hunt, K., Agarwal, P., and Zhuang, J. (2022). Monitoring misinformation on Twitter during crisis events: A machine learning approach, *Risk Analysis* 42(8), 1782-1748. https://doi.org/10.1111/risa.13634

Hunt, K., Wang, B., and Zhuang, J. (2020). Misinformation debunking and cross-platform information sharing through Twitter during Hurricanes Harvey and Irma: A case study on shelters and ID checks, *Natural Hazards* 106, 861-883. https://doi.org/10.1007/s11069-020-04016-6

Jamalzadeh, S., Barker, K., González, A.D., and Radhakrishnan, S. (2022). Protecting infrastructure performance from disinformation attacks, *Scientific Reports* 12, 12707. https://doi.org/10.1038/s41598-022-16832-w

Jiang, M., Gao, Q., and Zhuang, J. (2021). Reciprocal spreading and debunking processes of online misinformation: A new rumor spreading–debunking model with a case study, *Physica A* 565, 125572. https://doi.org/10.1016/j.physa.2020.125572

Kaplan, S. and Garrick, B.J. (1981). On the quantitative definition of risk, *Risk Analysis* 1(1), 11-27. https://doi.org/10.1111/j.1539-6924.1981.tb01350.x

Keck, M. and Sakdapolrak, P. (2013). What is social resilience? Lessons learned and ways forward, *Erdkunde* 67(1), 5 -19. https://www.jstor.org/stable/23595352

Keeney, R.L. and Raiffa, H. (1976). *Decisions with Multiple Objectives: Preferences and Value Trade-Offs*. New York: Wiley.

Komendantova, N., Ekenberg, L., Svahn, M., Larsson, A., Shah, S.I.H., Glinos, M., Koulolias, V., and Danielson, M. (2021). A value-driven approach to addressing misinformation in social media, *Humanities and Social Sciences Communications* 8(33), pp.1–12. https://doi.org/10.1057/s41599-020-00702-9

Kontogiannis, T., Leva, M.C., and Balfe, N. (2017). Total Safety Management: Principles, processes and methods, *Safety Science* 100, Part B, 128-142. https://doi.org/10.1016/j.ssci.2016.09.015.

Kuipers-Dirven, R., Janssen, M., and Hoekman, J. (2023). Assessing university policies for enhancing societal impact of academic research: A multicriteria mapping approach, *Research Evaluation* 32(2), 371–383.

Könnölä, T., Brummer, V., and A.Salo, A. (2007) Diversity in foresight: Insights from the fostering of innovation ideas, *Technological Forecasting and Social Change* 74(5), 608-626. https://doi.org/10.1016/j.techfore.2006.11.003

Könnölä, T., Salo, A., Cagnin, C., Carabias. V., and Vilkkumaa, E. (2012). Facing the future: scanning, synthesizing and sense-making in horizon scanning, *Science and Public Policy* 39(2), 222-231. https://doi.org/10.1093/scipol/scs021

Lazarus, J.V., Wyka, K., White, T.M., Picchio, C.A., Rabin, K., Ratzan, S.C., Leigh, J.P., Hu, J., and El-Mohandes, A. (2022). Revisiting COVID-19 vaccine hesitancy around the world using data from 23 countries in 2021. *Nature Communications* 13, 3801. https://doi.org/10.1038/s41467-022-31441-x

Lenton, T.M., Boulton, C.A., Scheffer, M. (2022). Resilience of countries to COVID-19 correlated with trust, *Scientific Reports* 12, 75. https://doi.org/10.1038/s41598-021-03358-w

Liesiö, J. and Salo, A. (2012). Scenario-based portfolio selection of investment projects with incomplete probability and utility information, *European Journal of Operational Research* 217(1), 162-172. https://10.1016/j.ejor.2011.08.025

Liesiö, J., Salo, A., Keisler, J.M., and Morton, A. (2021). Portfolio Decision Analysis: Recent developments and future prospects, *European Journal of Operational Research* 293/3, 811-825. https://doi.org/10.1016/j.ejor.2020.12.015

Lindbom, H., Tehler, H., Eriksson, K., and Aven, T. (2015). The capability concept – On how to define and describe capability in relation to risk, vulnerability and resilience, *Reliability Engineering and System Safety* 135, 45-54. https://doi.org/10.1016/j.ress.2014.11.007

Lorenz, D. F. (2013). The diversity of resilience: Contributions from a social science perspective, *Natural Hazards* 67(1), 1-18. https://doi.org/10.1007/s11069-010-9654-y

Lucini, B. (2014). What is resilience: The state of the art. *Disaster Resilience from a Sociological Perspective, Humanitarian Solutions for the 21st Century*, Springer. https://doi.org/10.1007/978-3-319-04738-6_3

Luo, Z., Xue, Y., and Su, J. (2023). Evaluating information risk propagation in complex public opinion environments based on the Improved grey relational analysis—Decision making trial

and evaluation laboratory method, *Systems* 11(9), 11090472. https://doi.org/10.3390/systems11090472

Lv, W., Zhou, W., Gao, B., Han, Y., and Fang, H. (2022). New insights into the social rumor characteristics during the COVID-19 pandemic in China, *Frontiers in Public Health* 10, 864955. https://doi.org/10.3389/fpubh.2022.864955

Löfstedt, R. (2005). *Risk Management in Post-Trust Societies*, Palgrave Macmillan.

Maček, D., Magdalenić, I., and Ređep, N.B. (2020). A systematic literature review on the application of multicriteria decision making methods for information security risk assessment, *International Journal of Safety and Security Engineering* 10(2), 161-174. https://doi.org/10.18280/ijsse.100202

Maleki, H. and Zahir, S. (2013). A comprehensive literature review of the rank reversal phenomenon in the Analytic Hierarchy Process, *Journal of Multi-Criteria Decision Analysis* 20(3-4), 141-155. https://doi.org/10.1002/mcda.1479

Mild, P., Liesiö, J., and Salo, A. (2015). Selecting infrastructure maintenance projects with Robust Portfolio Modeling, *Decision Support Systems* 77, 21-30. https://doi.org/10.1016/j.dss.2015.05.001

Moya, J. and Goenechea, M. (2022). An approach to the unified conceptualization, definition, and characterization of social resilience, *International Journal of Environmental Research and Public Health* 19(9), 5746. https://doi.org/10.3390/ijerph19095746

Nasery, M., Turel, O., and Yuan, Y. (2023). Combating fake news on social media: A framework, review, and future opportunities, *Communications of the Association for Information Systems*, 53(35), 833–876. https://doi.org/10.17705/1CAIS.05335

Nielsen, L., Tølbøll Glavind, S., Qin, J., and Faber, M.H. (2019). Faith and fakes – dealing with critical information in decision analysis, *Civil Engineering and Environmental Systems* 36(1), 32-54. https://doi.org/10.1080/10286608.2019.1615476

OECD (2024a). *The OECD Reinforcing Democracy Initiative: Monitoring Report – Assessing Progress and Charting the Way Forward*, OECD Public Governance Reviews. OECD Publishing, Paris, https://doi.org/10.1787/9543bcfb-en.

OECD (2024b). *The OECD Truth Quest Survey: Methodology and Findings*. OECD Digital Economy Papers No. 369. OECD, Paris. https://www.oecd.org/content/dam/oecd/en/publications/reports/2024/06/the-oecd-truth-quest-survey_a1b1739c/92a94c0f-en.pdf

Oniśko, A., Druzdzel, M. J., and Wasyluk, H. (2001). Learning Bayesian network parameters from small data sets: Application of Noisy-OR gates, *International Journal of Approximate Reasoning* 27(2), 165-182. https://doi.org/10.1016/S0888-613X(01)00039-1

Raetze, S., Duchek, S., Maynard, M.T., and Kirkman, B.L. (2021). Resilience in organizations: an integrative multilevel review and editorial introduction. *Group & Organization Management* 46(4), 607-656. https://doi.org/10.1177/10596011211032129

Rigdon, S. E., Jacobson, S. H., Tam Cho, W. K., Sewell, E. C., and Rigdon, C. J. (2009). A Bayesian prediction model for the US presidential election, *American Politics Research* 37(4), 700-724. https://doi.org/10.1177/1532673X08330670

Rocco, C.M., Barker, K., Radhakrishnan, S., and Ramirez-Marquez, J.E. (2025). Multi-objective model to protect infrastructure networks from disinformation diffusion, *Social Network Analysis and Mining* 15(28), 19-28. https://doi.org/10.1007/s41125-025-00105-4

Roponen, J. (2023). *Computational Models for Adversarial Risk Analysis and Scenario Planning*, Aalto University publication series Doctoral Theses 67/2023. http://urn.fi./URN:ISBN:978-952-64-1261-0

Rød, B., Pursiainen, C., and Eklund, N. (2025). Combatting disinformation – how do we create resilient societies? Literature review and analytical framework, *European Journal for Security Research* 10, 73-114. https://doi.org/10.1007/s41125-025-00105-4

Saari, D., Moilanen, P. and Hautala, M. (2025). Disinformation landscape in Finland. Finnish factsheet, EU Disinfolab. 11 p. https://www.disinfo.eu/wp-content/uploads/2025/04/20250402_Disinfo-landscape-in-Finland-V2.pdf

Salo, A. and Hämäläinen, R.P. (1997). On the measurement of preferences in the Analytic Hierarchy Process, *Journal of Multi-Criteria Decision Analysis* 6, 309-319. https://doi.org/10.1002/(SICI)1099-1360(199711)6:6%3C309::AID-MCDA163%3E3.0.CO;2-2

Salo, A., Hämäläinen, R.P., and Lahtinen, T.J. (2021). Multicriteria methods for group decision processes: an overview. In: D.M. Kilgour and C. Eden (eds.), *Handbook of Group Decision and Negotiation*, 2nd edition, Springer, Cham, 2021. 29 p.

Salo, A., Keisler, J., and A. Morton, A. (eds) (2010). *Portfolio Decision Analysis: Methods for Improved Resource Allocation*, Springer International Series in Operations Research & Management Science, Vol. 162, Springer, New York, 2011. 409 p.

Schneider, K., Pohl, E., Rainwater, C., Ramirez-Marquez, J.E., and Hernandez, I. (2013). Social network analysis via multi-state reliability and conditional influence models, *Reliability Engineering and System Safety* 109(1), 99-109. https://doi.org/10.1016/j.ress.2012.07.007

Seeve, T. and Vilkkumaa, E. (2022). D Identifying and visualizing a diverse set of plausible scenarios for strategic planning, *European Journal of Operational Research* 298(2), 596-210. https://doi.org/10.1016/j.ejor.2021.07.004

Simon, J., Regnier, E., and Whitney, L. (2014). A value-focused approach to energy transformation in the United States Department of Defence, *Decision Analysis* 11(2), 117-132.

Stirling, A. and Coburn, J. (2006). *Multicriteria Mapping Manual*. Brighton: SPRU, University of Sussex.

Straub, D.W. and Welke, R.J. (1998). Coping with systems risk: security planning models for management decision making, *MIS Quarterly* 22(4), 441-469. https://www.jstor.org/stable/24955

Tosoni, E., Salo, A., and Zio, E. (2018). Scenario analysis for the safety assessment of nuclear waste repositories: a critical review, *Risk Analysis* 38(4), 755-776. https://doi.org/10.1111/risa.12889

Townshend, I., Awosoga, O., Kulig, J. C., and Fan, L. (2015). Social cohesion and resilience across communities that have experienced a disaster, *Natural Hazards* 76:913–938. https://doi.org/10.1007/s11069-014-1526-4

Vilkkumaa, E., Rikkinen, P., Liesiö, J. (2025). A Framework for selecting action portfolios with action-dependent scenario probabilities, *Management Science*, published online 5 November 2025. https://doi.org/10.1287/mnsc.2024.05383

Viviani, M. and Pasi, G. (2017) A multi-criteria decision making approach for the assessment of information credibility in social media. In: Petrosino, A., Loia, V., and Pedrycz, W. (eds.) *Fuzzy Logic and Soft Computing Applications*. WILF 2016. Lecture Notes in Computer Science, vol 10147, Springer, Cham, 107-207. https://doi.org/10.1007/978-3-319-52962-2_17

Vujovich-Dunn, C., Kaufman, J., King, C., Skinner, S.R., Wand, H., Guy, R., and Leask, J. (2021). A systematic review and meta-analysis of effectiveness of decision aids for vaccination decision-making, *Vaccine* 39(28), 3655-3665. https://doi.org/10.1016/j.vaccine.2021.05.021

Wang, C., Chen, W., and Wang, Y. (2012). Scalable influence maximization for independent cascade model in large-scale social networks, *Data Mining and Knowledge Discovery* 25(3), 545-576. https://doi.org/10.1007/s10618-012-0262-1

White., R. (2017). *Multicriteria Mapping*. In: Spash, C.L. (Ed.), Routledge Handbook of Ecological Economics: Nature and Society, 1st Edition. 321-330.

Westover, J. H. (2024). Designing for resilience: principles for building organizational adaptability. *Human Capital Leadership Review*, 14(1). https://doi.org/10.70175/hclreview.2020.14.1.8

Yoshiura, L.J.M., Martins, C.L., and Costa, A.P.C.S. (2023) A multicriteria decision model for risk management maturity evaluation, *Pesquisa Operacional,* 43. https://doi.org/10.1590/0101-7438.2023.043.00270186

Żebrowski, P., Couce-Vieira, A., and Mancuso, A. (2022). A Bayesian framework for the analysis and optimal mitigation of cyber threats to cyber-physical systems, *Risk Analysis* 42(10), 2275-2290. https://doi.org/10.1111/risa.13900

Zhang, C., Gupta, A., Kauten, C., Deokar, A.V. & Qin, X. (2019). Detecting fake news for reducing misinformation risks using analytics approaches, *European Journal of Operational Research* 279(3), 1036–1052. https://doi.org/10.1016/j.ejor.2019.06.022

Appendix A

The network in Figure 3 representing possible information flows is modeled as a directed graph $G = (V, E)$, where nodes $V = V^P \cup V^V$ correspond to platforms V^P and voters V^V, connected by edges $E \subseteq \{(v, v') \mid v, v' \in V\}$. In the absence of disinformation, the probability that voter $v \in V^V$ votes to grant the permit, denoted by $D(v) = 1$, is Beta-distributed according to $p_v \sim B(\alpha, \beta) \in [0, 1]$ (see, e.g., Rigdon et al., 2009), where $\alpha = 6$ and $\beta = 5.5$ are the distribution parameters. Voting for permit denial is denoted by $D(v) = 0$. For these parameters, the permit is initially granted with probability 0.572 and denied with probability 0.428.

The disinformation status is represented by a binary variable, such that $S(v) = 1, v \in V^P \cup V^V$, if the platform $v \in V^P$ is attacked with disinformation or the voter V^V becomes disinformed, respectively. In our scenario, all platforms are equally likely to be attacked $\mathbb{P}[S(v) = 1] = \frac{1}{3}, \forall v \in V^P$. We assume there is no recovery from disinformation (see, e.g., Rocco et al., 2025).

The policymaker can take action to monitor each platform at three different levels of intensity, represented by the decision variable λ_v^M whose value is 0, 1 or 2 in case of no monitoring, light monitoring, and close monitoring; these three values are also the costs of respective monitoring actions. For the monitoring action λ_v^M, the policymaker obtains a binary indication of the attack $S(v) = 1, v \in V^P$ with probability $\mathbb{P}[S^M(v) = 1 \mid S(v) = 1] = \lambda_v^M \delta^M$, where $\delta^M = 0.4$. Thus, for close monitoring $\lambda_v^M = 2$, there is still a 1 – 2x0.4 = 0.20 probability of not detecting the disinformation. There are no false positive indications: if the indication is received, the attack has surely occurred.

If the monitoring indication is received, preparations for isolating the attacked platform $v \in V^P$ can be made by the action represented by the binary decision variable $\lambda^I = 1$ at a cost of 8 units. If isolated, the disinformation cannot propagate from the platform to any voters. Because only one platform is attacked, there is no need to introduce separate decision variables for the different platforms.

If the platform $v' \in V^P$ is attacked but not isolated, a voter $v \in V^V$ who is directly connected only to this platform becomes disinformed with probability $q_{v',v} = \delta^P$, where the parameter $\delta^P = 0.5$ represents the strength of this connection. Disinformation can propagate from disinformed voters to their respective connections so that if the voter $v' \in V^V$ is disinformed $S(v') = 1$, then the influence of this on a voter $v \in V^V$, connected to v' by the edge $(v', v) \in E$, is $q_{v',v} = (1 - p_{v'})\delta^V$, where the parameter $\delta^V = 0.4$ represents the strength of the connection between voters. This equation reflects the reality that supporters of granting the permit, who have a high value of $p_{v'}$, are less likely to promote disinformation in support of permit denial. Finally, a voter $v \in V^V$ with incoming

connections from disinformed nodes $V^D(v) = \{v' \in V \mid (v', v) \in E, S(v') = 1\}$ becomes disinformed with probability

$$\mathbb{P}[S(v) = 1 \mid V^D(v), \lambda^I] = 1 - \prod_{v' \in V^D(v)} \left(1 - q_{v',v}\right),$$

which is the noisy OR-gate models (Oniśko et al., 2001). The disinformation status of voters is simulated sequentially in the order given in Figure 3.

Disinformation influences voters' behavior by affecting the likelihood ratio representing their probabilities of voting to grant or deny. Specifically, for a disinformed voter $v \in V^V, S(v) = 1$ this ratio changes from $p_v/[1 - p_v]$ to

$$\frac{q_v}{1 - q_v} = \Delta(\lambda^E) \frac{p_v}{1 - p_v},$$

where q_v the updated probability of granting the permit and the parameter $\Delta(\lambda^E)$ reflects the impact that the decision λ^E to educate voters has on the impact of disinformation. Education can be provided at three levels such that $\lambda^E = 0$ for no education, $\lambda^E = 1$ for limited education, and $\lambda^E = 2$ for substantial education. The associated costs are twice these values. Parameters $\Delta(\lambda^E)$ are assumed to be $\Delta(0) = 1/4, \Delta(1) = 1/3, \Delta(2) = 1/2$, which yields the corresponding distributions shown in Figure 4. For example, if the initial probability of voting to grant is $p_v = 3/4 = 0.75$ and no education is provided, then $\lambda^E = 0$ and $\Delta(\lambda^E) = 1/4$, in which case the probability of voting to grant is reduced to $q_v = 3/7 = 0.423$.

The possible strategies are portfolios $\lambda = \{\lambda_v^M \mid v \in V^P\} \cup \{\lambda^I, \lambda^E\} \in \Lambda$ defined by combinations of monitoring, isolation, and education actions. The first objective is to minimize the total cost of actions. Secondly, the probability of disinformed voters is to be minimized. Thirdly, the impact on the outcome of the vote is to be minimized, which is interpreted as minimizing the probability that the permit is denied, further to the disinformation attack. Thus, the multicriteria optimization problem is

$$\mathrm{v} - \min_{\lambda \in \Lambda} \begin{pmatrix} \sum_{v \in V^P} (c^M \lambda_v^M) + c^I \lambda^I + c^E \lambda^E, \\ \mathbb{P}\left[\sum_{v \in V^V} S(v) > 0 \mid \lambda\right], \\ \mathbb{P}\left[\sum_{v \in V^V} D(v) < \frac{|V^V|}{2} \mid \lambda\right] \end{pmatrix},$$

where c^M, c^I, c^E are the costs associated with monitoring, isolation, and education, respectively, and $|V^V|$ is the number of voters.