EAI Endorsed Transactions
on Security and Safety
Research Article EAI.EU

Resonant Risk: A Sociotechnical Model for Understanding Cyber Threat Perception and Amplification in the Age of Artificial Intelligence (AI)

Muhammad Sajid Khan

Pakistan Space & Upper Atmosphere Research Commission, Islamabad - Pakistan

Abstract

INTRODUCTION: Cyber incidents are increasingly shaped not only by technical severity but also by how risk signals are amplified through AI-mediated information ecosystems. Deepfakes, synthetic media, algorithmic amplification, and AI-generated misinformation can intensify public fear, distort trust, and trigger disproportionate societal responses.
OBJECTIVES: This paper develops the Resonant Risk Model as a sociotechnical extension of the Social Amplification of Risk Framework for AI-era cybersecurity. It also proposes the Resonant Risk Management Framework to support perception-aware cyber risk governance.
METHODS: The study uses theory-building, interdisciplinary literature synthesis, structured case selection, and comparative case analysis. Six cases are assessed using dimensions including technical severity, amplification channels, resonance factors, public perception, societal ripple effects, and feedback loops.
RESULTS: The analysis shows that high technical severity does not always produce high public resonance. SolarWinds showed very high technical severity but moderate public resonance, while AI-driven misinformation and deepfake cases produced high trust erosion despite lower direct technical impact.
CONCLUSION: The paper argues that cyber resilience must include perception monitoring, rapid communication, misinformation correction, and trust recovery alongside technical controls.

Keywords: Cybersecurity, Risk Perception, Social Amplification of Risk, Artificial Intelligence, Deepfakes, Misinformation, Risk Communication, Resonant Risk, Risk Management

Received on 09 August 2025, accepted on 11 May 2026, published on 13 May 2026

doi: 10.4108/eetss.9915

1. Introduction

Cybersecurity threats have expanded in scope and impact with the advent of advanced AI capabilities. From AI-generated deepfake videos that can spark misinformation cascades, to intelligent bots that amplify malicious narratives, the “Age of AI” has transformed not only the threat landscape but also public perception of those threats. Today, a single cyber incident can resonate globally within minutes, fueled by social media algorithms and instant communication [1]. This resonance can inflate certain risks in the public mind while obscuring others, a phenomenon we term “Resonant Risk.” This paper addresses the critical question: How are cyber threat perceptions shaped and amplified in the age of AI, and how can we better understand and manage this social amplification of cyber risk?

Cyber risk is therefore no longer only a matter of technical probability and operational impact. It is also shaped by how information about an incident is framed, circulated, trusted, disputed, and emotionally interpreted. In AI-mediated information environments, synthetic media, automated accounts, recommendation algorithms, and generative misinformation can intensify or distort public

understanding of cyber events. This creates a perception layer of cyber risk, where social reaction may become disproportionate to technical reality or, conversely, where severe technical risks may remain socially invisible.

As cyber threats entwine with AI technologies, society faces a dual challenge. On one hand, AI tools can be weaponized by malicious actors to create convincing falsehoods (e.g. deepfake scams, automated disinformation), thereby amplifying the scale and reach of cyberattacks [2]. On the other hand, even benign AI systems can produce errors or "hallucinations" that, if misinterpreted, incite undue alarm or false confidence [3][4]. Misperceptions of cyber risk, whether overestimation via hype and panic, or underestimation via complacency, can lead to inappropriate responses. Over-amplified threats may cause public panic, economic disruption, or policy overreactions, while underplayed dangers may result in insufficient preparation for real attacks [1][5]. There is a pressing need for models that explain these dynamics of perception and guide stakeholders in responding appropriately.

Traditional cybersecurity risk analysis focuses on technical factors (vulnerabilities, attack likelihood, impact quantification), often neglecting the social and psychological dimensions of risk perception [1]. Conversely, risk perception research (e.g. in public health or environmental hazards) has rich theory on social amplification but has not fully incorporated the unique features of cyber threats and AI-driven information ecosystems. We lack an integrative theory that connects technical cyber risk assessments with the "human factor" of how those risks are perceived, magnified, or downplayed through modern media channels. This gap is evident in the contemporary context: for instance, widespread fears about AI-enabled cyberattacks coexist with apathy toward everyday security hygiene, suggesting inconsistent risk perceptions. To address this gap, we apply and extend the Social Amplification of Risk Framework (SARF) to the domain of cybersecurity in the AI era, formulating a Social Amplification Theory of Cyber Threat Perception.

Despite this need, current cybersecurity risk models provide limited explanation of how cyber threat perception is amplified or attenuated through AI-driven information ecosystems. Existing technical frameworks are effective for identifying vulnerabilities, controls, likelihood, and impact, but they rarely explain why some cyber incidents trigger public panic, institutional distrust, market reaction, or regulatory pressure while other technically severe incidents receive limited public attention. Similarly, classical risk perception theories explain social amplification but require adaptation to account for algorithmic curation, synthetic media, deepfakes, AI hallucinations, and platform-based misinformation. This gap creates the need for a sociotechnical model that connects technical cyber risk with perception dynamics and public response.

This paper makes four main contributions. First, it extends the Social Amplification of Risk Framework to the AI-era cybersecurity context by incorporating digital amplification channels, synthetic media, and algorithmic information flows. Second, it proposes the Resonant Risk Model as a conceptual model for explaining how cyber risk signals move from technical events to public perception and societal ripple effects. Third, it applies a structured comparative case analysis to examine how traditional and AI-driven cyber incidents produce different amplification patterns. Fourth, it introduces the Resonant Risk Management Framework as a practical approach for preparing, sensing, communicating, and recovering from perception-driven cyber risk events.

This study is guided by several key questions aimed at filling the identified gap:

How do AI technologies and digital media contribute to the social amplification (or attenuation) of cyber risk signals? For example, in what ways do deepfakes, algorithm-driven news feeds, or viral tweets influence public perceptions of a given cyber incident?

What psychological and contextual factors mediate cyber threat perception in the general public and organizations? We explore factors such as novelty/unknownness of the threat, dread (feelings of fear and lack of control), trust in institutions, and prior beliefs – and how these interact with AI-driven information flows.

What are the secondary impacts ("ripples") of amplified cyber risk perception on society and security? We examine potential outcomes like economic costs (e.g. market reactions to cyber news), policy changes, stigmatization of technologies, erosion of trust, or even social unrest [1][5].

How can a theoretical and practical framework be constructed to explain and manage resonant risk? We seek to design a model that integrates technical risk assessment with social amplification mechanisms, and to propose strategies (technical, communicative, policy-based) that mitigate harmful amplification while ensuring real threats are appropriately heeded.

By answering these questions, the paper aims to both advance theoretical understanding and offer actionable guidance for practitioners and policymakers.

Significance and Scope: The stakes are high and global. The World Economic Forum's Global Risks Report 2025 highlights that misinformation, disinformation, and adverse outcomes of AI technologies rank among the top global risks for the coming decade [5], alongside cyber warfare and espionage [5]. Cyber threats augmented by AI, whether in the form of sophisticated phishing, fake news influencing elections, or autonomous hacking tools, have the potential to undermine societal trust and stability worldwide. Recent incidents, such as a fake AI-generated image of an explosion at the Pentagon briefly causing stock market dips [6], demonstrate how perceptions driven by false cyber-related information can produce real economic shocks. In this global context, understanding resonant risk is crucial for international security and cooperation. This paper takes a global perspective, using examples and case studies from different regions and considering cross-cultural factors in risk perception. The goal is to provide a theory-informed foundation for understanding and

managing cyber risk amplification in AI-mediated information environments.

The scope of this paper is conceptual and analytical. It does not claim to provide a fully validated quantitative model of cyber risk perception. Instead, it develops a theory-informed model supported by structured case comparison and qualitative resonance assessment. This approach is appropriate for an emerging problem area where AI-driven amplification mechanisms are evolving rapidly and where empirical measurement methods are still developing.

2. Methodology

2.1 Research Design

This study adopts a theory-building and conceptual research design grounded in the extension of the Social Amplification of Risk Framework (SARF) into the cybersecurity domain under the influence of artificial intelligence. The research follows a qualitative, interdisciplinary approach, integrating insights from cybersecurity, risk communication, behavioral science, and information systems.

Rather than conducting purely empirical testing, the study aims to develop an explanatory and operational model that captures how cyber risks are perceived, amplified, or attenuated within modern digital ecosystems. To support this conceptual development, the research combines structured literature synthesis with comparative case analysis. This approach enables the identification of recurring amplification patterns across different cyber incidents and supports the formulation of the proposed Resonant Risk Model and the Resonant Risk Management Framework (RRMF).

The design is aligned with established practices in sociotechnical and risk research, where theoretical advancement is supported through analytical generalization rather than statistical inference.

2.2 Literature Selection Strategy

The literature used in this study was selected through a targeted and interdisciplinary approach. Sources were drawn from peer-reviewed journals, conference proceedings, institutional reports, and authoritative publications in the domains of risk analysis, cybersecurity, artificial intelligence, and communication studies.

The selection process focused on three key thematic areas:

Risk perception and amplification theory, particularly foundational and contemporary work related to SARF and psychometric risk analysis

Cybersecurity and digital risk communication, including studies on data breaches, cyber incidents, and crisis response

Artificial intelligence and misinformation, covering deepfakes, algorithmic amplification, and AI-generated content

Priority was given to recent and peer-reviewed sources to ensure relevance to the evolving AI-driven threat landscape. Non-academic sources, such as media reports, were used selectively to document real-world cyber incidents and their societal impacts, rather than as theoretical foundations.

2.3 Case Selection Criteria

To support the analytical validation of the proposed model, a set of representative cyber incidents was selected using predefined criteria. The objective was not to provide exhaustive coverage, but to include cases that illustrate diverse amplification dynamics across both traditional and AI-driven threat contexts.

The case selection was guided by the following criteria as in Table 1:

Table 1. Case/ Research Selection Criteria

Criterion	Description	Justification for Inclusion
Relevance to Cyber or AI Risk	The case must involve a cybersecurity incident or AI-mediated misinformation, manipulation, or system failure	Ensures alignment with the study's focus on cyber threat perception and AI-driven amplification
Public Visibility	The case must have received measurable attention through mainstream media, social media, or institutional reporting	Enables observation of amplification dynamics across communication channels
Amplification channels	The case must demonstrate one or more amplification pathways, such as news media, social media, expert commentary, or official communication	Supports analysis of how risk signals propagate and are transformed
Societal Impact	The case must show observable consequences, including behavioral change, economic disruption, regulatory response, or trust erosion	Ensures that resonance is evaluated in terms of real-world outcomes
Source Reliability	The case must be supported by credible sources, including peer-reviewed literature, official reports, or well-documented incident coverage	Strengthens analytical validity and reduces reliance on anecdotal evidence
Diversity of Context	Cases must represent both traditional cyber incidents and AI-driven or misinformation-related scenarios	Enables comparative analysis across different amplification mechanisms

The application of these criteria ensures that the selected cases provide a balanced representation of cyber risk amplification across different technological and social

contexts. By combining traditional cybersecurity incidents with AI-driven scenarios, the analysis captures both infrastructure-based and perception-driven amplification dynamics, supporting a more comprehensive evaluation of the proposed model.

2.4 Analytical Framework and Procedure

Each selected case was analyzed using a structured analytical framework derived from SARF and extended through the proposed Resonant Risk Model. The analysis focused on identifying how risk signals propagate through amplification channels and how they influence public perception and societal outcomes.

The analytical procedure involved the following dimensions (Table 2):

Table 2. Analytical Framework for Case Analysis

Analytical Dimension	Guiding Question	Evidence Considered	Link to Resonant Risk Model
Signal Origin	What type of cyber or AI-related event initiated the risk signal?	Incident type, affected systems, threat actor claims, technical description, and initial disclosure	Defines the starting point of the risk amplification process
Technical Severity	What was the actual or reported operational impact of the event?	Data exposure, service disruption, financial loss, critical infrastructure impact, or system compromise	Provides the baseline against which perceived risk can be compared
Amplification Channels	Through which channels did the risk signal spread?	News media, social media, expert commentary, official alerts, organizational statements, and interpersonal sharing	Identifies the stations through which the risk signal was amplified or attenuated
Resonance Factors	Why did the event become emotionally or socially salient?	Novelty, dread, uncertainty, personal relevance, institutional distrust, synthetic media realism, and perceived lack of control	Explains why some cyber events resonate more strongly than others
Public Perception	How was the risk interpreted by affected publics or institutions?	Public concern, confusion, panic, indifference, polarization, or sector-specific anxiety	Captures the point at which technical risk becomes perceived social risk
Societal Ripple Effects	What secondary consequences followed from the perception of risk?	Market reaction, behavioral change, policy debate, regulatory scrutiny, reputational damage, or trust erosion	Shows how amplified perception creates consequences beyond the initial incident
Feedback and Learning	Did the response influence future preparedness or governance?	New policies, updated authentication practices, communication changes, training, or incident response improvements	Captures the feedback loop from perception response back into future resilience

The framework was applied consistently across all selected cases to support analytical comparability. Each case was examined not only in terms of its technical characteristics, but also in terms of how the event was communicated, interpreted, amplified, and translated into societal or institutional response. This approach supports the central argument of the paper: cyber risk impact is increasingly shaped by the interaction between technical severity and perception dynamics, especially in AI-mediated information environments.

2.5 Validity and Limitations of the Method

The methodological approach provides analytical rigor through structured case comparison and interdisciplinary synthesis; however, several limitations must be acknowledged.

First, the study is primarily conceptual and does not rely on large-scale empirical data or statistical modeling. As such, the findings are intended for analytical generalization rather than predictive accuracy.

Second, the case analysis relies on publicly available information and secondary data sources, which may not capture the full internal dynamics of each incident.

Third, the assessment of public perception and amplification is interpretive in nature, as direct measurement (e.g., through surveys or large-scale sentiment analysis) is beyond the scope of this study.

Despite these limitations, the structured analytical framework, transparent case selection criteria, and cross-case comparison strengthen the validity of the findings. The methodology is sufficient for theory development and provides a foundation for future empirical validation, including quantitative measurement of risk resonance and large-scale perception analysis.

3. Theoretical Foundations: Risk Perception and Social Amplification in Cybersecurity

Understanding how cyber threats are perceived in the public arena requires grounding in risk perception theory and the social processes that amplify or attenuate those perceptions. In this section, we summarize key theoretical frameworks and examine their relevance to cybersecurity in the AI era.

3.1. Social Amplification of Risk Framework (SARF)

One foundational approach is the Social Amplification of Risk Framework (SARF), introduced by Kasperson and colleagues [1][7]. SARF is not a single theory but an integrative framework that links the technical assessment of risk with social, psychological, and cultural processes [1][7]. The core premise of SARF is that events pertaining to risk, such as accidents, attacks, or discoveries of hazards, are filtered and interpreted through various "amplification stations," including the media, individual risk assessors, cultural groups, and interpersonal networks [1][7]. These stations transmit information about the risk, often exaggerating or downplaying certain aspects, which in turn shapes public perception and behavior. As a result, some events that experts deem low risk can become the focus of intense public concern (risk amplification), whereas other objectively higher-risk issues receive comparatively little attention (risk attenuation) [1][8].

Amplification Stages: According to SARF, amplification typically occurs at two key stages [1]:

(i) Information Transmission: The initial communication of the risk through media and other channels. At this stage, signals about the risk can be strengthened or weakened. For instance, a news report might use dramatic language and imagery, thereby amplifying the signal, or conversely might bury the story on a back page, attenuating its impact. In modern terms, the virality of a tweet or the prominence of a story in search results can dramatically influence how amplified the risk signal becomes.

(ii) Social Response: The reaction by society, which can create "ripples" or secondary effects [1]. Public responses, such as public outcry, changes in consumer behavior, or demands for regulation, can magnify the consequences of the event beyond its direct harm. For example, fear-induced avoidance behavior or stigmatization can cause economic losses and social disruption even if the actual physical damage from the event was minimal [1].

The ripple effects are an important element of SARF. Amplified risk perceptions can lead to outcomes like stigma (where an entity associated with the risk is broadly shunned), policy changes, market impacts, or loss of trust [1]. In extreme cases, amplified perceptions might even provoke social disorder such as protests or violence. Attenuation, on the other hand, might result in complacency and inadequate preparation for a real threat.

Amplification Stations and Media: SARF identifies various actors as "amplification stations." Initially conceived in the late 1980s, these included scientists, news media, cultural groups, and interpersonal networks [1]. Today, we must add social media platforms and AI-driven content algorithms to this list [9]. These new stations differ in their dynamics: they are decentralized, fast, and often lack traditional gatekeeping. Indeed, social media-driven amplification tends to be "emotionally intense, time-compressed, and with less authority control over risk information" [10][11]. A rumor on Twitter or a viral video can circle the globe in minutes, reaching millions before official sources have a chance to respond [10].

SARF in Technological Risks: SARF was originally developed with hazards like nuclear accidents and chemical spills in mind. However, it has since been applied to a wide range of risks, from vaccines to climate change, and scholars note it remains highly relevant to understanding modern risk communication [10][11]. Notably, even though SARF predates the social media era, its principles anticipated the complex interactions now seen on global digital platforms [10][12]. For instance, an extension by Kasperson in 2003 [3] visualized "ripples in a pond" to represent how local risk events could spread via media into broader impacts [1][3], an analogy even more apt now that ubiquitous social networks act as a pond with almost no boundaries.

3.2. Psychometric and Cognitive Aspects of Cyber Risk Perception

While SARF emphasizes social processes and communication channels, understanding why certain risks resonate with the public requires looking at cognitive factors. Pioneering work by Slovic and colleagues developed the psychometric paradigm, identifying that people's perceptions of different risks depend on characteristics such as "dread" and "unknown risk" [7]. These factors are highly relevant to cyber threats in the AI age [1][3][7]:

- **Dread Risk**: This factor captures how fearful a risk is, often associated with lack of control, catastrophic potential, and the perceived worst-case outcomes. A risk that is highly dreaded evokes strong emotional responses and urgency. Cyber scenarios like an AI-powered cyberattack on critical infrastructure (e.g. causing widespread power grid failure) can evoke dread due to their potentially catastrophic scale and the public's limited ability to personally control or mitigate the threat. Research suggests that events with pronounced "dread" qualities tend to receive

sustained media coverage and public attention. In a study of data breach news, for example, breaches seen as not easily mitigated and potentially affecting many people (a source of dread) were associated with more prolonged coverage and secondary impacts.

- **Unknown Risk**: This factor involves how unfamiliar or new the risk is, and whether its effects are delayed or not observable. Novel AI-related threats (like entirely new forms of deepfake propaganda or unpredictable autonomous agents) score high on the unknown factor – the technology is new, its failure modes are not fully understood, and consequences might emerge only later. The novelty of AI threats can itself be an amplifier. Studies indicate that when a risk is novel and not well understood, media and public interest can be heightened. For instance, early reports on "deepfake" videos in politics garnered massive attention precisely because the public had never seen such a capability before, tapping into a fear of the unknown, "if we can't trust what we see or hear, what next?".
- **Personal Control and Voluntariness**: People perceive risks as higher when they feel a lack of personal control and when exposure is involuntary. Cyber risks often feel outside individual control, one cannot personally stop hackers on the internet, and exposure (e.g. one's data being hacked) is usually not a choice. This too can amplify fear. Conversely, if people feel empowered (through good cybersecurity practices or trusted tools), the perceived risk might attenuate.
- **Trust and Distrust**: Trust in institutions and technology developers is crucial. SARF-related studies have found that "social distrust of responsible institutions" can amplify risk perceptions. In cybersecurity, if the public distrusts that companies or governments can keep data safe, they may react more strongly to each breach or threat report. Unfortunately, sensational cyber incidents can erode trust further, creating a vicious cycle. This interplay of trust and perception is a key part of resonant risk theory.

3.3. Media, Misinformation, and the Age of AI

The digital communication revolution, especially social media and AI content generation acts as a force multiplier for risk amplification [14][15]. Three interrelated aspects are noteworthy:

- **Speed and Reach**: Information (and misinformation) travels instantaneously on platforms like Twitter, Facebook, or TikTok. This compresses the timeframe for reaction. There is little time for fact-checking or authoritative framing before public impressions form. As noted, today's "ripples" from a risk event can "fan out in minutes to global networks" [1], whereas in earlier decades the spread was slower and more geographically contained. A striking example occurred in May 2023, when a falsified AI-generated image of an explosion at the Pentagon went viral and briefly sent U.S. stocks tumbling before being debunked [6]. In this case, an alarming (but fake) risk signal resonated through social media and triggered real-world economic ripples within hours, a textbook case of rapid amplification via digital channels.
- **Algorithmic Amplification and Echo Chambers**: Social media algorithms, powered by AI, prioritize content that maximizes engagement, often sensational or emotionally charged posts. This can skew amplification toward extreme narratives. If a cyber threat story sparks anger or fear, algorithms may push it to more users, further amplifying those emotions. Moreover, online communities can form echo chambers where particular risks are either hyped up or dismissed, insulated from corrective viewpoints. For instance, a rumor about a certain software update being a spying scheme can bounce around in a conspiratorial forum and grow, even if unfounded. AI bots can also artificially inflate the apparent popularity of a narrative (e.g. bots retweeting a fabricated cyber threat), making it "trend" and thus appear credible or urgent [14][15].
- **Synthetic Media and False Signals**: AI enables the creation of synthetic media, "deepfakes" in video, audio, or text, that can simulate evidence of cyber threats or attacks. This is a new wildcard in risk perception. In the past, a claim of a security incident would be judged by available evidence; now, fake evidence can be manufactured to support false claims. For example, a deepfake video could be made of a government official "announcing" a cyberattack or of a CEO "admitting" a massive data breach, potentially fooling even experts temporarily. These false signals can greatly amplify perceived risk if believed. As Villasenor (2019) notes, deepfakes "can scramble our understanding of truth… exploiting our inclination to trust what we see with our own eyes", turning fiction into apparent fact [16]. This threatens to erode the baseline of trust in digital information, making risk communication extremely fraught.

Misinformation and disinformation are not just by-products; they are recognized as top-tier risks themselves. The World Economic Forum warned that "misinformation and disinformation remain top short-term global risks, undermining trust in governance" [2], and in the long term, technological risks like AI misuse in misinformation cloud the horizon. The interplay of AI and social media lies at the heart of these warnings. When considering cyber threat perception, we must acknowledge that the information ecosystem is polluted with intentional distortions (disinformation by malicious actors) and unintentional inaccuracies (misinformation, including AI hallucinations). These distortions can either amplify fear of certain threats disproportionately or attenuate legitimate

warnings. For instance, during a fast-evolving cyber crisis, conflicting reports (some false) might circulate, causing confusion and either undue panic or dangerous complacency if the wrong message gains traction.

4. Application to Cybersecurity Context

Literature specifically bridging SARF and cybersecurity has begun to emerge, confirming the framework's relevance. Research on major data breaches, for example, has applied social amplification concepts to understand public reactions. Dias & Mendonça [17] examined the Equifax and Capital One breaches, finding that media coverage volume correlated with amplification of perceived risk, but amplification also depended on qualitative factors of the incident and information flow. High exposure events with dramatic attributes (e.g. involving sensitive personal data, affecting millions, or indicating gross negligence) garnered sustained coverage and public anger, essentially creating a "reputational contagion" that extended beyond the immediate customers affected. The authors identified social amplification factors in cyber incidents, including: the dread and unknown qualities of the breach, the level of media dramatization and controversy, and social distrust in the institution responsible. These factors align closely with those in SARF and psychometric paradigms, reinforcing that cyber risks are subject to the same amplification dynamics observed in health or environmental risks, perhaps even more so, given their intangible nature.

Another study by Sutton et al. (2015) on risk sharing in Twitter during crises observed that the diffusion and retransmission of risk information on social media essentially is a form of amplification [11][10], albeit one that is hard to measure directly. It highlighted that in social media, traditional authority (experts, officials) has less control over the narrative, meaning false or emotive content can outpace measured, factual communications [1]. This underscores a challenge for cybersecurity: official advisories or clarifications might be drowned out by more sensational user-driven content. For example, when a new malware strain or data breach hits the news, cybersecurity agencies might issue a calm bulletin, but at the same time, individual commentary on Reddit or Twitter could exaggerate ("This is digital apocalypse!") or dismiss ("Fake news, don't worry!") the threat, leading to public confusion.

Global and Cultural Considerations: Risk perception and amplification can vary culturally. What resonates as a major fear in one society might not in another, due to differences in values, recent experiences, and media environment. For instance, countries that have recently experienced disruptive cyberattacks (like ransomware on hospitals or election interference) may have a heightened public sensitivity ("dread") toward cyber risks, whereas others might be more concerned with immediate economic issues. Cultural worldview (as per Douglas and Wildavsky's cultural theory of risk [18]) can influence which threats are amplified, e.g., individualistic societies might amplify cyber threats to personal liberty (like mass surveillance), while more collectivist ones might focus on threats to social stability [18][19]. While a full cultural analysis is beyond our scope, our model is intended to be adaptable globally, acknowledging that "resonant risk" may have different triggers in different contexts, even if the amplifying mechanisms are similar.

All things considered, the theoretical foundation is built on the insight that cyber threats become "resonant", echoing and amplifying through society, not merely by technical virtue of the threat, but through a complex interplay of communication dynamics and human psychology. The Social Amplification of Risk Framework provides a lens to examine these dynamics, and when combined with cognitive risk factors and the realities of AI-driven media, it sets the stage for our proposed model. We next introduce the Resonant Risk Model of cyber threat perception, which adapts these theoretical concepts to practical use in the cybersecurity domain.

5. The "Resonant Risk": A Model for AI-Era Cyber Threat Perception

Building on the above foundations, the study propose the Resonant Risk Model, a conceptual model that describes how cyber risks in the AI era are socially amplified or attenuated, and how they propagate through society. The model adapts classic SARF elements to the modern digital environment, incorporating AI-specific factors (like deepfakes and algorithms) and highlighting points where intervention can occur. Figure 1 below is a conceptual diagram of the model's components and their interactions (from the initial risk event through to societal outcomes).

(i) **Risk Event "Signal Origin"**: This refers to the initiating cyber threat or incident, either real (e.g., a data breach, malware attack, AI system malfunction) or anticipated (e.g., a rumored vulnerability or AI-predicted attack). Each event carries characteristics such as severity, uncertainty, and scope that influence the signal's strength. A technically complex AI exploit may carry an "unknown" quality, while a ransomware attack on a hospital evokes high dread. In this model, the strength of the initial signal is shaped not only by technical risk (probability and impact) but also by how alarming or novel it appears. A small but unprecedented incident may produce a stronger signal than a larger, routine one.

(ii) **Amplification/Attenuation Channels**: From the event, information flows out via multiple channels, each acting as an amplifier or attenuator:

- **News Media**: Traditional outlets report the event through press releases, investigations, and expert commentary. The prominence (headline vs. buried

story) and framing (tone, imagery) can amplify or mute public concern. Terms like "digital catastrophe" enhance perceived risk, while emphasizing containment reduces it. Media also rely on expert quotes, which may either lend authority or dramatize the event. Dias & Mendonça [17] observed that the frequency of news stories correlates strongly with public risk perception.

- **Social Media and Online Platforms**: Information spreads rapidly online, often shaped by unique amplification mechanisms:
  - **Viral Spread**: Engaging or shocking content spreads peer-to-peer rapidly. User shares, retweets, and likes to serve as multipliers.
  - **Algorithmic Curation**: Trending or emotionally engaging posts may be algorithmically promoted, exaggerating reach.
  - **User-Generated Content and Speculation**: Commentary, memes, and conspiracy theories can generate new narratives.
  - **Bots and Coordinated Influence**: Malicious actors can leverage bots or troll farms to amplify certain messages intentionally (for instance, a state-sponsored campaign might exaggerate an enemy nation's cyberattack to cause panic) [6].
  - **Memes and Emotional Hooks**: Social media often conveys complex issues through memes or emotive anecdotes, which can oversimplify but leave a strong impression, often amplifying aspects that resonate emotionally while ignoring mitigating details.
- **Official and Expert Communication**: Government agencies, cybersecurity firms, and technical experts can act as key attenuation agents through alerts, press briefings, and technical bulletins. Timely, clear communication can counteract misinformation and reduce overreaction. For instance, when the AI-generated image of a Pentagon explosion went viral, local police quickly tweeted to dismiss it [6]. However, poor messaging or low public trust can erode this effect. In some cases, experts unintentionally amplify fear by emphasizing severity ("most serious threat in decades"), especially if such claims gain wide media attention.
- **Interpersonal Networks**: People exchange risk information within communities, workplaces, or families. This form of transmission is shaped by social trust, such as advice from a knowledgeable colleague may prompt more action than a news report. Interpersonal exchanges often contextualize risk: "Should we be worried?" or "Is this just media hype?" Group sentiment can either amplify through shared concern or attenuate through reassurance. In tight-knit groups, a single voice (e.g., a CEO or IT lead) can shape internal perception significantly.

These channels rarely convey risk neutrally. They emphasize, distort, or omit aspects of the event, modulating the original signal like a sound wave. Interactions between channels, such as social media fueling traditional media attention, create feedback loops that can greatly enhance overall amplification [1]

(iii) **Public Risk Perception ("Resonance" Point)**: After traveling through various amplification stages, the signal converges into a public perception of the cyber threat. This "resonance point" reflects the collective sense of urgency or severity surrounding an event. Perception is not uniform, different groups (experts vs. laypeople, different regions, or age groups) respond differently, but there is typically an average perceived risk level discernible through sentiment analysis, surveys, or public discourse. Key outcomes at this stage:

- **Heightened Perception**: If amplification dominated, the public perception may be that the cyber threat is extremely dire or imminent, possibly exceeding the technical reality. Signs include widespread fear, extensive media attention, and people talking about the threat as a major concern.
- **Attenuated Perception**: If attenuation dominated (or if a high-risk event was ignored), the public might underestimate the threat. There may be a lack of concern or awareness, even among those potentially affected.
- **Polarized Perception**: Not uncommon in today's information environment, some groups are extremely worried (amplified perception) while others are dismissive (attenuated or in denial), often correlating with ideological lines or trust in different information sources.

At this stage, cognitive biases come into play such as "availability heuristics" in which People judge likelihood based on how easily examples come to mind. If a threat dominates the media, it appears more likely. Also, "confirmation bias", in which new information is interpreted to fit pre-existing beliefs. A fearful user may view all updates as reinforcing their concern.

Sometimes, the perception detaches from the original incident entirely. For example, a single deepfake scam could spawn broad societal fear that "nothing online can be trusted," extending perceived risk from one case to an entire technology category. Conversely, an advanced supply-chain attack may receive little concern if the public struggles to grasp its complexity.

(iv) **Societal Impacts and Responses (Risk Ripples)**: The final component of the model covers the real-world outcomes that result from the public perception. These can be direct responses to the perceived risk or secondary effects that unfold over time:

- **Behavioral Changes**: Individuals may avoid digital tools or platforms seen as risky. For example, panic over an AI voice cloning scam might cause users to disable smart assistants. Positive outcomes are also

possible: a ransomware surge in the news may push companies to invest in backup and patching practices.

- **Market and Economic Effects**: Markets are sensitive to perceived cyber threats. The false AI-generated Pentagon explosion led to a measurable drop in stock prices before clarification restored confidence [6]. Companies suffering widely publicized breaches may see disproportionate financial losses, while fintech use might drop if digital fraud stories trend. Cyber insurance premiums may spike in response to amplified perceptions of certain attack types.
- **Political and Regulatory Actions**: Public pressure can drive swift government responses. An amplified threat may prompt policy actions, investigations, or new regulations. For instance, increased visibility of deepfake abuse has fueled legislative momentum to regulate synthetic media [2][20]. Conversely, under-amplified threats may receive no political attention until crisis escalation forces action.
- **Social Movements and Trust Dynamics**: High resonance can catalyze public distrust in technology or institutions. Users may reject platforms, protest government failures, or rally around perceived technological threats. Technologies associated with high-profile failures may become stigmatized [1]. For example, AI voice tools may be seen as inherently dangerous after misuse cases, regardless of broader context.
- **Feedback to Risk Itself**: Perception can reshape the risk environment. Overreaction (e.g., taking critical systems offline) can introduce vulnerabilities or economic damage. On the other hand, amplified concern may drive better defenses, decreasing future risk. Underreaction may allow threats to grow unchecked. As SARF scholars observed, amplified behavior changes can "heighten or lower the risk itself" [1].

In the Resonant Risk Model, these stages form a chain, but with iterative loops. A particularly important feedback loop is crisis communication: once authorities see the public perception and response, they may adjust their messaging or actions to try to correct courses (e.g. issue stronger reassurances or, conversely, amplify warnings if they feel people are too complacent). This can either stabilize the situation or, if done poorly, further feed amplification (for instance, overly forceful assurances might backfire and increase public mistrust).

To bring all together, Resonant Risk is the end-to-end process by which a cyber risk event generates public resonance (amplified or dampened) and leads to tangible societal outcomes beyond the immediate technical damage. It extends SARF by explicitly incorporating AI-driven factors at the amplification stage and emphasizing the rapid, global scope of modern risk resonance. It also underscores the dual possibility of over-amplification (hype, panic) and under-amplification (silence, ignorance), both of which are problematic. The ideal scenario is accurate risk resonance, where public perception aligns with actual risk, prompting proportional response. Achieving that ideal is challenging, which motivates the strategies discussed later in this paper.

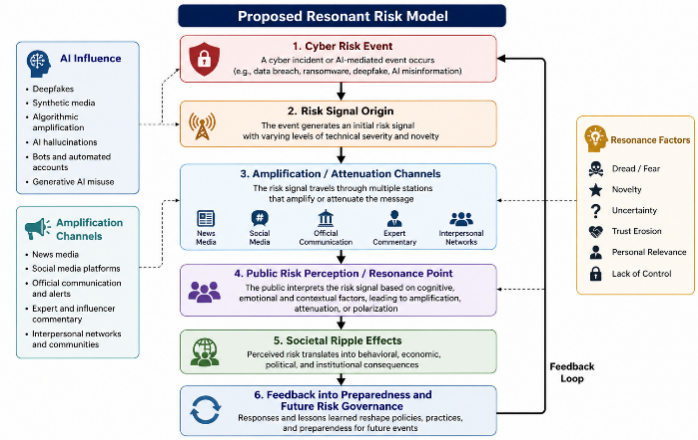

Figure 1. Illustration of The Proposed Resonant Risk Model [Source: Self]

6. Comparative Case Analysis of Cyber Risk Amplification

To assess the explanatory capability of the proposed Resonant Risk Model, this section applies a structured comparative case analysis across a set of representative cyber incidents. The selected cases span both traditional cybersecurity events and emerging AI-driven threat scenarios, enabling examination of how different types of risk signals are amplified or attenuated within modern information ecosystems.

The objective of this analysis is not to provide exhaustive empirical coverage, but to demonstrate how variations in amplification channels, cognitive resonance factors, and communication dynamics lead to disproportionate societal responses relative to technical severity. By applying a consistent analytical framework across cases, this section highlights recurring patterns in how cyber risks are socially constructed, perceived, and acted upon.

6.1. Traditional Cyber Incidents: Deep Resonance via Tangibility and Infrastructure Disruption

Several historical cyber incidents illustrate the social amplification mechanisms predicted by Kasperson et al.'s [1][7] Social Amplification of Risk Framework (SARF) and extended here through the Resonant Risk Model.

- **Equifax Data Breach (2017)**: The Equifax data breach (2017), which exposed personal data of

approximately 147 million individuals, represents a high-resonance cyber event driven by scale, personal relevance, and institutional trust failure [21]. While data breaches are not uncommon, the magnitude of exposure combined with delayed disclosure and perceived organizational negligence triggered sustained media attention, regulatory scrutiny, and public outrage. The incident illustrates how personalization of risk ("your identity is exposed") significantly amplifies perception, resulting in long-term reputational damage and erosion of trust in financial institutions [22].

- **SolarWinds Sunburst Attack (2020)**: The SolarWinds supply chain attack (2020) represents a technically severe but socially attenuated cyber event [23]. The attack compromised multiple U.S. government agencies and private sector organizations through a sophisticated software update mechanism. Despite its strategic significance, the incident generated limited public resonance due to its technical complexity, lack of immediate consumer-facing impact, and delayed visibility [24]. This case demonstrates that high technical severity does not necessarily translate into high public perception, particularly when cognitive accessibility and media framing are limited.
- **Colonial Pipeline Ransomware (2021)**: The Colonial Pipeline ransomware attack (2021) exemplifies a strongly amplified cyber event driven by visible societal disruption [25]. The attack led to fuel shortages across the U.S. East Coast, triggering panic buying, long queues, and widespread media coverage. Unlike more abstract cyber incidents, the tangible impact on daily life activated strong cognitive and emotional responses [26]. This case highlights how physical-world consequences, combined with real-time media and social amplification, can produce rapid and intense resonance, resulting in both behavioral and policy-level responses [26].

6.2. AI-Driven Cyber Incidents: New Amplification Vectors and Trust Disruptions

The integration of AI-generated content into cyber events introduces novel forms of amplification. These incidents, while sometimes lower in technical severity, disproportionately influence public and institutional reactions due to their memetic potential, novelty, and erosion of epistemic trust.

- **Deepfake CEO Voice Scam**: The deepfake CEO voice fraud case, first publicly reported in 2019, involved cybercriminals using AI-based voice synthesis to impersonate a senior executive and authorize a fraudulent financial transfer [27]. Although technically a form of social engineering, the use of AI-generated voice introduced a novel and psychologically disruptive threat vector. Media coverage and expert commentary emphasized the realism of synthetic impersonation, amplifying organizational concern and triggering widespread reassessment of authentication mechanisms [28]. This case demonstrates how AI-driven novelty and trust exploitation significantly increase perceived risk beyond traditional fraud scenarios [28].
- **ChatGPT Hallucination in Legal Filings (2023)**: The use of AI-generated content in legal proceedings gained attention in 2023 when fabricated case citations generated by a language model were submitted in court filings [29]. Although no direct cyberattack occurred, the incident exposed risks associated with unverified AI outputs in high-stakes environments [30][31]. The case generated institutional concern, legal scrutiny, and media attention, raising broader questions about trust in AI-assisted decision-making. This scenario illustrates how perceived risk can emerge not from malicious exploitation, but from the erosion of epistemic reliability in AI systems.
- **OpenAI Screenshot Misinformation (2024)**: The circulation of fabricated screenshots purporting to show outputs from AI systems represents a growing form of misinformation-driven amplification [32]. In such cases, visually convincing but unverifiable content is shared across social media platforms, often triggering public concern or outrage before verification occurs. The visual credibility of these artifacts, combined with rapid online diffusion, contributes to disproportionate trust erosion and reputational impact on AI platforms [33]. This case highlights how synthetic yet believable digital artifacts can act as powerful amplification triggers within modern information ecosystems.

6.3. Applying the Resonant Risk Model: Deepfake CEO Scam

This section applies the Resonant Risk Model to the deepfake CEO voice fraud case to illustrate how amplification dynamics operate across the full perception lifecycle. The scenario exemplifies how amplification dynamics, rooted in AI novelty, dread, and information asymmetry, trigger disproportionate societal reactions and policy ripples, despite the varying technical foundations of the risk events.

In this high-impact event, cybercriminals used AI-based voice synthesis to impersonate a corporate executive and defraud a multinational firm. Under the Resonant Risk Model:

- **Risk Event (Signal Origin)**: The incident embodied high novelty and dread, as it violated trust in direct voice communications. The impersonation bypassed traditional social engineering defenses, exploiting an unfamiliar threat vector.

- **Amplification Channels**: News media highlighted the "sci-fi" nature of the fraud, with headlines emphasizing AI cloning and substantial monetary loss. Social media added virality, with users personalizing and meme-ifying the threat ("What if this happened to me?"), increasing reach and psychological salience. Influencers and tech explainers further dramatized the risks through YouTube and podcasts.
- **Public Perception**: The resonance outcome was strong. The incident crystallized fear of AI-enabled fraud and generalized concern beyond corporations, inducing shifts in how individuals and organizations interpreted voice authenticity. Resonance extended well beyond the initial victim, reflecting widespread concern over the erosion of sensory trust.
- **Societal Impacts**: Corporate security protocols evolved, with increased uptake of multi-channel verification systems. Regulatory discussions intensified, and insurance providers updated cyber policy exclusions to address synthetic impersonation. Some opportunistic phishing campaigns emerged, capitalizing on the amplified fear. Thus, the resonance induced both productive and parasitic ripples.

This case provides preliminary analytical support for the model's core: AI-enhanced threat → rapid amplification → behavioral, regulatory, and trust effects, including feedback into future preparedness and public discourse.

All the above cases are summarized in Table 3 below as per the applied Resonant Risk Model:

Table 3. Case Comparison of Traditional and AI-Driven Cyber Incidents

Case	Type	Technical Impact	Amplification Channel	Public Perception
Equifax Data Breach (2017)	Traditional	Massive PII breach affecting 147M people, erosion of public trust in credit bureaus.	News media, Congressional hearings, regulatory scrutiny.	High distrust, long-term reputational damage.
SolarWinds Supply Chain Attack (2020)	Traditional	Sophisticated attack via software update compromised U.S. federal systems.	Technical reports, expert discussion, delayed mainstream coverage.	Moderate concern among public; high alert within cybersecurity community.
Colonial Pipeline Ransomware (2021)	Traditional	Operational shutdown of fuel pipelines, nationwide panic and fuel shortages.	Mainstream media, viral images, government alerts.	Severe panic, emergency behavior among consumers.
Deepfake CEO Voice Scam (2023)	AI-Driven	AI-generated audio used to impersonate a CEO, leading to fraudulent transfers.	corporate security discussions.	and corporate vulnerabilities.
ChatGPT Hallucination in Legal Filing (2023)	AI-Driven	AI hallucinations cited in legal filings, revealing issues of reliability in high-stakes domains.	Legal media, institutional alerts, academic scrutiny.	Mixed responses; concern within legal field, broader public confusion.
OpenAI Screenshot Misinformation (2024)	AI-Driven	Fake screenshots of AI responses used to spread misinformation and discredit AI models.	Meme culture, online forums, media speculation on AI risks.	Erosion of trust in AI platforms due to realistic-looking falsehoods.

6.4. Comparative Synthesis and Resonance Assessment

The cases summarized above demonstrate that cyber risk amplification is not solely determined by technical severity, but by a combination of amplification channels, cognitive resonance factors, and communication dynamics. Traditional incidents tend to achieve high resonance when they produce tangible societal disruption, whereas AI-driven incidents often derive amplification from novelty, symbolic impact, and erosion of trust in digital systems.

Table 4. Resonance Assessment Matrix Across Selected Cases

Case	Technical Severity	Media Visibility	Social Spread	Trust Erosion	Behavioral / Institutional Response	Overall Resonance
Equifax Data Breach	High	High	Moderate	High	High	High
SolarWinds Supply Chain Attack	Very High	Moderate	Low	Moderate	High within government and cybersecurity sectors	Moderate
Colonial Pipeline Ransomware	High	High	High	High	Very High due to panic buying and emergency response	Very High
Deepfake CEO Voice Fraud	Moderate	High	High	High	High due to authentication review and organizational concern	High
ChatGPT Legal Hallucination Case	Low to Moderate	High	Moderate	High in legal and professional settings	Moderate to High	High
Fabricated AI Screenshot Misinformation	Low technical severity	Moderate to High	High	High	Moderate due to reputational and public trust impact	Moderate to High

Table 4 shows that resonance does not always correspond directly with technical severity. SolarWinds, for example, represented a very high technical and strategic

risk, yet its broader public resonance remained comparatively moderate because of limited public visibility and technical complexity. In contrast, AI-driven misinformation cases may produce high trust erosion and social spread despite lower direct technical severity. This supports the central claim of the Resonant Risk Model: cyber risk impact in the AI era is shaped by the interaction between technical harm, communication channels, cognitive salience, and trust dynamics.

Note: The assessment is interpretive and based on structured comparison of publicly available case evidence. The categories Low, Moderate, High, and Very High are used as qualitative indicators rather than statistical measurements. They are intended to support analytical comparison and future operationalization of a quantitative resonance index.

6.5. Societal and Operational Implications

The presence of AI in the risk amplification loop alters not only threat perception but also institutional behaviors. For instance [34][35][36]:

- **Policy Overreach**: Some jurisdictions have considered blanket AI restrictions based on deepfake misuse rather than grounded risk analysis.
- **Alert Fatigue**: Continuous amplification via unverified content may desensitize users to genuine threats.
- **Cybersecurity Strategy Misalignment**: Security teams may allocate disproportionate resources to high-amplification/low-impact threats, neglecting silent but severe vulnerabilities.

Furthermore, these cases suggest that accurate risk resonance, where public perception aligns proportionally with technical severity, is rare. Instead, organizations and regulators must design communication and risk response models that consider the dual challenges of under- and over-amplification.

Across both categories of incidents, feedback loops are often mismanaged. In the AI related cases, official silence allowed misinformation to metastasize. In contrast, the Colonial Pipeline crisis saw effective coordination among government and media to stabilize panic. This reinforces the critical importance of crisis communication timing and tone as outlined in the Resonant Risk Model. Authorities must balance reassurance and transparency, avoiding both overconfidence and silence, especially when AI-related amplification mechanisms are involved [37].

This comparative analysis confirms that cyber threat perception is not determined solely by technical severity. Instead, resonance emerges from a complex interplay of amplification stations, cognitive and cultural factors, and media dynamics, now increasingly shaped by AI technologies. As the boundary between real and synthetic blurs, the future of cyber risk management must evolve beyond patching code toward managing perception itself.

7. Solutions and Strategies for Managing Resonant Risk

Managing cyber threat perception in the age of artificial intelligence requires a multidimensional strategy addressing technological, governmental, and societal domains. Building on the Resonant Risk Model, the following interventions (Figure 2) aim to balance amplification with accuracy, foster public trust, and mitigate both panic and complacency.

7.1. Technological and Industry Interventions

The following industry interventions are suggested:

- **AI-Augmented Detection Systems**: Industry stakeholders must invest in AI tools that counteract misuse of synthetic content. This includes deepfake detection algorithms integrated into social platforms and cybersecurity applications [38]. Voice biometrics vendors are upgrading systems to detect synthetic frequencies in spoofed calls [20]. Collective intelligence-sharing across firms, via consortia or ISACs, should improve detection accuracy through pooled datasets and model refinement.
- **AI Transparency Standards**: Embedding cryptographic watermarks or metadata into AI-generated outputs can enhance traceability and reduce confusion. As mandated under the EU AI Act, synthetic media must be clearly labeled [20]. Similar transparency efforts should become standard practice across global AI platforms, irrespective of legal jurisdiction.
- **Human-in-the-Loop Protocols**: In safety-critical applications, AI systems must be governed by human oversight. This includes requiring human verification in diagnosis, fraud detection, or threat alerting. Interface design should make uncertainty and anomaly scores visible to users, minimizing blind trust in AI-generated conclusions and mitigating alert fatigue.
- **Rapid Response Mechanisms**: Incident response teams must establish clear protocols for verifying and publicly addressing sensational cyber claims. Equivalents to “fact-checking desks” should be institutionalized within cybersecurity operations, offering vetted counter-narratives that reduce amplification time windows. Trusted CERTs or industry groups can coordinate synchronized messaging to dispel false alarms.
- **Perception Simulations and Preparedness Drills**: Organizations should conduct scenario-based simulations of information-based incidents, e.g., viral falsehoods or disinformation campaigns. As practiced in some cybersecurity awareness campaigns, these drills can test readiness for rapid containment and communications.

7.2. Policy and Regulatory Interventions

The following policy and regulatory interventions are recommended:

- **AI Risk Governance and Content Authentication Laws**: Governments should strengthen legislative frameworks that mandate labeling of AI-generated content, criminalize malicious deepfake use, and enforce responsible disclosure of cyber incidents. Adoption of authentication standards like C2PA can embed verifiable provenance into digital media, supporting forensic verification and legal recourse.
- **Public Education on Digital Risk Literacy**: Sustained efforts are needed to build "risk literacy" among the public, helping citizens evaluate digital content and detect manipulation. This includes curriculum development, community outreach, and media campaigns comparing real versus AI-generated materials. Training should be extended to professionals in journalism, law, and IT.
- **Cyber Crisis Communication Frameworks**: Governments should maintain designated spokespersons and verified digital channels for cyber emergencies. These entities must issue rapid, multilingual, and platform-amplified updates during high-profile incidents. Coordination across national CERTs can also reduce vulnerability to cross-border disinformation tactics.
- **Simulation-Based Policymaking and Research**: Funding should be directed toward agent-based models and risk simulation tools that predict societal perception trajectories. These tools can help determine thresholds for public communication, preemptive action, and countermeasures under various threat scenarios [1].
- **International Cooperation on Information Integrity**: Global norms against weaponizing misinformation (especially AI-generated) are essential. Multilateral frameworks should formalize consequences for states deploying synthetic media to incite cyber panic or fabricate incidents, akin to existing norms prohibiting bioweapon disinformation.

7.3. Societal and Organizational Practices

- **Trust Infrastructure Within Organizations**: To mitigate internal amplification, institutions should designate centralized communication hubs during cyber events. These should be used for official updates, countering rumors, and directing staff behavior. Identifying and training internal "information leaders" can foster rational response culture [1].
- **Ethical Self-Governance in Media and Social Platforms**: Media outlets and influencers must adopt norms discouraging speculative or exaggerated reporting of cyber threats. Establishing verification guidelines and encouraging clear attribution practices can reduce unnecessary escalation. Fact-checking teams should work in tandem with technical experts to ensure accuracy.
- **Community Resilience Campaigns**: Cyber preparedness should mirror natural disaster protocols. Public drills and simulated disinformation events (e.g., false reports of local cyber incidents) can educate communities on appropriate response mechanisms. Sweden's national defense materials on misinformation serve as a model for broader adoption [39].
- **Perception Monitoring and Early Warning Systems**: Institutions should monitor social media and news feeds for disproportionate amplification signals. Tools such as a "risk resonance index" could be developed by correlating perceived risk levels (social volume) with technical severity, prompting calibrated responses.
- **Post-Incident Learning Loops**: After major incidents (real or fabricated) organizations must assess not only technical outcomes but perception management. Evaluations should cover rumor spread patterns, messaging effectiveness, and audience behavior, feeding lessons into revised protocols and training.

By applying these measures cohesively, stakeholders can build resilience not only against technical threats, but also against the distortions of perception that can amplify (or dangerously mute) those threats. Addressing both the technical and social vectors of cyber risk is essential for building trust, ensuring proportional response, and preserving institutional legitimacy in an age of accelerating information volatility.

Figure 2. Solution domains for managing resonant risk across technological, regulatory, societal, and organizational layers [Source: Self]

7.4. Proposed RRMF Framework for Implementation

To implement the above recommendations, a Resonant Risk Management Framework (RRMF) has been proposed for organizations and governments. It consists of four cyclical phases explained below:

(i) **Prepare**: Before incidents, invest in technology (detection tools, backups), establish communication protocols, train spokespersons, and educate stakeholders. Simulation exercises and building trusted networks fall here.

(ii) **Sense & Analyze**: Continuously monitor for technical threats and perception signals (social/media monitoring). Use AI analytics to identify misinformation quickly. Analyze the potential impact on perception, essentially risk intelligence that covers social factors.

(iii) **Communicate & Act**: During an incident, execute the communication plan. Issue clear, honest, and frequent updates. Engage with media and influencers proactively to get correct narratives out. If misinformation is present, directly debunk it with evidence (transparency is key, share what you know, what you don't know yet, and what's being done). Also take technical containment actions in parallel but ensure communication doesn't lag behind the news cycle.

(iv) **Recover & Improve**: After the incident stabilizes, address any lingering public concerns (sometimes the perception harm lasts longer than the technical harm). Provide support where needed (e.g., hotlines to answer public questions after a breach). Then review and refine plans, technology, and training.

Figure 3 below summarizes the Resonant Risk Management Framework (RRMF)

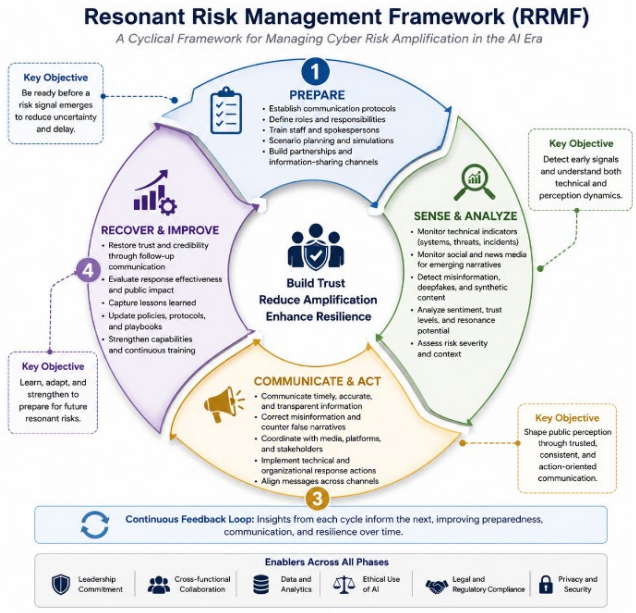

Figure 3. Proposed Resonant Risk Management Framework showing a cyclical approach to preparing for, detecting, communicating, and learning from cyber risk amplification events [Source: Self]

This framework aligns with general crisis management [40] but tailored to the nuances of cyber risk and AI-era amplification. It emphasizes that communication is not an afterthought to technical response, but an integral part of the incident response itself.

Importantly, the solutions must be adaptive. As AI technology evolves (e.g. deepfakes becoming indistinguishable, or new forms of synthetic media), so too must our defenses and communication techniques. We should expect adversaries to also adapt (e.g., planting fake "official" debunk messages to confuse, or timing attacks to media cycles). Therefore, an agile, learning-oriented approach is needed in implementing these strategies.

By combining these measures, technical safeguards, regulatory action, and savvy communication, stakeholders can strive to dampen the harmful resonances of cyber risks. The aim is a society where people are informed and vigilant but not panicked; where technology's benefits are utilized, but its outputs are verified; and where malicious attempts to distort perception are quickly exposed and neutralized. In the final section, we discuss the broader implications of this approach and future research directions to further strengthen our ability to navigate the social dimensions of cybersecurity.

7.5 Operationalizing the Resonant Risk Management Framework (RRMF)

To enhance practical applicability, the Resonant Risk Management Framework can be operationalized through clearly defined activities, sector-specific adaptations, measurable indicators, and supporting resources. Table 5 presents a structured implementation view across the four phases of the framework.

Table 5. Structured Implementation Across the Phases of The Framework

Phase	Key Activities	Sector Example	Measurable Indicators	Required Resources
Prepare	Establish communication protocols, define roles and responsibilities, conduct simulation	A financial institution prepares for AI-enabled fraud scenarios such as deepfake voice	Training completion rate, simulation performance scores, incident readiness level	Incident response team, communication specialists, training platforms,

	exercises, train spokespersons, and build trusted information channels	attacks by training staff and updating verification procedures		simulation tools
Sense & Analyze	Monitor technical systems and threat intelligence, track social media and news signals, detect misinformation, and assess sentiment and trust indicators	A government CERT monitors emerging narratives about a suspected cyber incident while correlating them with technical alerts	Detection time for misinformation, sentiment shift indicators, correlation between technical and perception signals	Security operations center (SOC), threat intelligence tools, social listening platforms, data analytics capability
Communicate & Act	Issue verified updates, counter misinformation, coordinate with media and stakeholders, and execute technical response actions	A healthcare organization communicates transparently during a ransomware incident while correcting false reports spreading online	Time to issue first public statement, misinformation correction time, message reach and engagement	Trained spokesperson, legal and compliance support, media coordination channels, crisis communication team
Recover & Improve	Restore trust, evaluate communication effectiveness, capture lessons learned, and update policies and response strategies	A public sector organization conducts post-incident review after a misinformation-driven cyber panic and updates its communication protocols	Trust recovery indicators, stakeholder feedback, completion of post-incident review actions	Post-incident review team, governance and audit functions, policy development resources

The operationalization of the RRMF highlights that effective cyber risk management in the AI era requires integration of technical detection, perception monitoring, and strategic communication. The inclusion of measurable indicators ensures that organizations can assess not only technical response performance but also the effectiveness of perception management and trust recovery. This reinforces the core premise of the framework: resilience depends equally on managing cyber threats and managing how those threats are understood and communicated.

The framework is intentionally adaptable across sectors. While critical infrastructure and government entities may require advanced monitoring and coordination capabilities, small and medium-sized enterprises can implement simplified versions focusing on communication readiness and basic perception monitoring. This scalability supports broader adoption without imposing uniform resource requirements.

7.6 Implementation Barriers and Alignment with Existing Frameworks

Implementation Barriers

While the Resonant Risk Management Framework provides a structured approach to managing cyber risk amplification, several practical challenges may affect its adoption across organizations.

First, organizational culture and communication maturity play a critical role. Many organizations prioritize technical incident response but lack established processes for perception monitoring or coordinated public communication. This gap can delay response to misinformation or amplify inconsistent messaging.

Second, resource constraints may limit implementation, particularly for small and medium-sized enterprises. Capabilities such as social media monitoring, sentiment analysis, and rapid communication coordination require specialized tools and trained personnel, which may not be readily available.

Third, trust and credibility challenges can reduce the effectiveness of communication strategies. In environments where public or stakeholder trust is already weakened, even accurate messaging may fail to attenuate amplified risk perception.

Fourth, integration complexity arises when aligning perception-focused activities with existing cybersecurity operations. Security teams may operate independently from communication or public relations functions, creating coordination gaps during high-impact incidents.

Finally, information ecosystem volatility, driven by algorithmic amplification and rapid misinformation spread, makes it difficult to maintain control over narrative dynamics. Even well-executed communication strategies may be outpaced by viral content or coordinated disinformation campaigns.

Alignment with NIST CSF and ISO/IEC 27001

The proposed RRMF is not intended to replace existing cybersecurity frameworks but to extend them by incorporating perception and amplification dynamics into risk management processes.

The framework aligns naturally with the National Institute of Standards and Technology Cybersecurity Framework (NIST CSF 2.0), particularly across its core functions:

- Govern and Identify: The Prepare phase supports governance structures, risk awareness, and organizational readiness
- Detect: The Sense & Analyze phase extends detection to include perception signals such as misinformation and public sentiment
- Respond: The Communicate & Act phase complements incident response by integrating communication and narrative management
- Recover: The Recover & Improve phase aligns with recovery planning and continuous improvement practices

Similarly, the RRMF complements the International Organization for Standardization ISO/IEC 27001:2022 standard, particularly in areas such as:

- Incident Management: Enhancing response processes with coordinated communication strategies
- Monitoring and Logging: Extending monitoring to include external perception and information signals
- Communication Controls: Supporting structured internal and external communication during incidents
- Continual Improvement: Embedding perception-based learning into post-incident reviews

By integrating with these established frameworks, the RRMF introduces a sociotechnical layer that addresses how cyber risks are interpreted, amplified, and acted upon by society. This integration ensures that organizations can enhance their existing cybersecurity posture without requiring complete restructuring of established processes.

The combined application of technical frameworks and perception-aware risk management enables a more holistic approach to cybersecurity resilience. While traditional frameworks focus on system protection and recovery, the RRMF extends this capability to include trust preservation, misinformation management, and proportional response to perceived threats. This integration is essential in the AI era, where the impact of cyber incidents increasingly depends on how they are communicated and understood rather than solely on their technical characteristics.

8. Discussion

The Resonant Risk Model offers a multidimensional reframing of cyber threat perception by synthesizing the Social Amplification of Risk Framework (SARF) with the evolving information ecosystems shaped by artificial intelligence, algorithmic curation, and social media virality. It moves beyond traditional technical paradigms by demonstrating that the social construction of cyber risk is increasingly shaped by how threats are mediated, interpreted, and circulated across platforms and communities [1][2].

The first research question "how AI technologies and media contribute to amplification or attenuation of cyber risk signals" is addressed through an analysis of AI-enhanced virality mechanisms. The model identifies how algorithm-driven feeds, real-time recommendation systems, and generative media such as deepfakes introduce volatility into the perception lifecycle of a cyber incident [6][16]. As observed in the deepfake voice scam targeting a CEO [38], it is not the technical exploit alone but its rapid reproduction and emotional framing that escalate societal concern. Furthermore, AI-generated hallucinations embedded within large language models [20] show how false narratives may be unintentionally amplified, creating pseudo-incidents that mobilize public or organizational responses.

The second research question explores psychological and contextual factors mediating cyber threat perception. The model confirms that emotional dimensions, such as dread, lack of control, and unfamiliarity, are central in how individuals and institutions assess cyber risk salience [8]. These findings are consistent with cognitive heuristics in risk communication literature, wherein emotionally charged or novel stimuli are more likely to resonate, particularly when institutional trust is low [1]. In this context, platform design becomes a key factor: echo chambers, virality mechanics, and lack of content provenance deepen perceptual distortions.

Addressing the third research question, the study shows that amplified or distorted risk perception leads to concrete societal and security impacts, ranging from behavioral overreactions (e.g., premature withdrawals of services), economic consequences (e.g., market volatility in response to cyber rumors), to policy misalignment and institutional erosion. Case studies revealed that the degree of public engagement was not proportionate to technical severity but to media framing and timeliness of official clarification. This perception-response disconnect can drive costly misallocations, stigmatization of technologies, or even geopolitical tensions [5][41].

The fourth research question focused on constructing theoretical and practical models to manage resonant risk. The model's value lies in its operability: it not only maps the socio-technical feedback loops that constitute risk amplification but also provides an implementation-ready framework encompassing four phases "Prepare, Sense & Analyze, Communicate & Act, and Recover & Improve". These phases institutionalize perceptual resilience as a continuous, adaptive process rather than a post-incident concern. The model's emphasis on embedding perception monitoring into cyber incident response infrastructures represents a paradigm shift. For instance, real-time misinformation detection using AI and rapid narrative correction are now not optional but integral components of modern cybersecurity resilience [2][20].

A critical insight from the simulations and framework implementation is the dual risk of over-amplification and under-amplification. While exaggerated responses can generate public panic, alert fatigue, or economic disruption, insufficient visibility into complex but high-risk vulnerabilities can render systems socially invisible until failure occurs. This duality aligns with empirical research showing that risk attention is often disproportionately distributed [6]. It calls for recalibrated threat communication systems capable of modulating amplification based on both technical reality and perceptual distortion indices.

Furthermore, the policy review suggests that existing regulatory frameworks are technologically reactive but perceptually underprepared. While the EU AI Act mandates labeling of synthetic content, its fragmented international uptake leaves critical vulnerabilities in cross-border media ecosystems [20]. Without harmonized risk communication standards or shared amplification indices, adversarial actors exploit timing, platform differences, and jurisdictional inertia to sow asymmetric disinformation.

To bring it all together, this study advances the discourse on cyber risk by foregrounding perception as a threat surface. It argues that AI's role in cybersecurity is both catalytic and corrosive: it enables faster detection and deeper distortion simultaneously. The findings support a shift toward perceptual intelligence as a new frontier in cyber governance, where predictive analytics, sentiment monitoring, and narrative countermeasures are operationalized alongside technical defenses. Future research should quantify risk resonance using amplification indices and scenario-based simulations, developing predictive tools to anticipate not just incidents, but their perceptual cascades.

9. Limitations

This study has several limitations that should be acknowledged. First, the research is primarily conceptual and theory-building in nature. The proposed Resonant Risk Model and Resonant Risk Management Framework are developed through interdisciplinary synthesis and comparative case analysis rather than large-scale empirical testing. Therefore, the findings should be interpreted as analytical and explanatory rather than statistically generalizable.

Second, the case analysis relies on publicly available sources and secondary evidence. Although this supports transparency and reproducibility, it may not capture the internal decision-making processes, communication strategies, or stakeholder perceptions that shaped each incident response.

Third, the resonance assessment matrix uses qualitative indicators such as media visibility, trust erosion, social spread, and institutional response. These indicators provide preliminary analytical support but do not constitute a fully validated quantitative measurement model.

Fourth, public perception may vary across cultural, political, and institutional contexts. Levels of trust in government, media, technology providers, and cybersecurity authorities can influence whether a cyber risk signal is amplified, attenuated, or ignored. As a result, the model may require contextual adaptation when applied across different regions or sectors.

Finally, the study does not conduct real-time social media analytics, stakeholder interviews, or expert validation. Future research should address these limitations through empirical testing, sentiment analysis, expert surveys, and cross-cultural comparative studies.

10. Conclusion and Future Work

The Resonant Risk Model proposed in this paper reframes cyber risk not solely as a function of system vulnerabilities but as a product of perception dynamics, media architectures, and AI-driven signal distortion. In doing so, it addresses a growing gap in cybersecurity discourse: the absence of integrated frameworks to understand and govern the societal consequences of cyber threats as mediated experiences.

Findings reveal that AI technologies, while improving detection and efficiency, also amplify misinformation, heighten public anxiety, and complicate attribution and response. The case studies underscore that technical severity alone is no longer the dominant factor in determining risk impact; rather, resonance (how risk is perceived, shared, and reacted to) often defines societal outcomes. This realization challenges conventional cybersecurity models that prioritize technical control without addressing narrative control.

The proposed Resonant Risk Management Framework offers a strategic response: integrating technical monitoring with perception analytics, embedding crisis communication protocols in incident response, and fostering trust networks to stabilize perception. By positioning communication as a parallel vector of defense, the framework promotes a dual resilience: technical and social.

Future research should operationalize the concept of "risk resonance" through empirical studies, including sentiment tracking, amplification index development, and cross-platform threat perception modeling. Simulations and real-world exercises must test not only system readiness but perception handling under uncertainty. Interdisciplinary collaboration, spanning cybersecurity, AI ethics, behavioral science, and public policy, will be essential to advance this agenda.

As generative AI continues to evolve, so too must society's tools to interpret and moderate its influence on risk perception. Only by addressing both the technical and perceptual vectors can cyber resilience be meaningfully achieved.

References

[1] R. E. Kasperson, T. Webler, B. Ram, and J. Sutton, "The social amplification of risk framework: New perspectives," Risk Analysis, vol. 42, no. 7, pp. 1367–1380, Jul. 2022, doi: https://doi.org/10.1111/risa.13926.

[2] T. Brooks and J. Heatley, "Increasing Threat of Deepfake Identities," Department of Homeland Security, 2023. Available: https://www.dhs.gov/sites/default/files/publications/increasing_threats_of_deepfake_identities_0.pdf

[3] M. Salvagno, F. S. Taccone, and A. G. Gerli, "Artificial intelligence hallucinations," Critical Care, vol. 27, no. 1, May 2023, doi: https://doi.org/10.1186/s13054-023-04473-y.

[4] M. S. Khan, "Slopsquatting (AI Hallucinations) and the Future of Secure Prompt Engineering," Medium, Jul. 22, 2025. https://medium.com/@sajidmkd/slopsquatting-ai-hallucinations-and-the-future-of-secure-prompt-engineering-24705c1225ed (accessed Aug. 01, 2025).

[5] WEF, "Global Risks Report 2025: Conflict, Environment and Disinformation Top Threats," Jan. 30, 2025. https://ghhin.org/news/global-risks-report-2025-conflict-environment-and-disinformation-top-threats/ (accessed Aug. 01, 2025).

[6] D. Alba, "How a fake AI photo of a Pentagon blast went viral and briefly spooked stocks," May 22, 2023. https://www.latimes.com/business/story/2023-05-22/how-fake-ai-photo-of-a-pentagon-blast-went-viral-and-briefly-spooked-stocks

[7] R. E. Kasperson et al., "The Social Amplification of Risk: A Conceptual Framework," Risk Analysis, vol. 8, no. 2, pp. 177–187, Jun. 1988, doi: https://doi.org/10.1111/j.1539-6924.1988.tb01168.x.

[8] W. J. Burns, P. Slovic, R. E. Kasperson, J. X. Kasperson, O. Renn, and S. Emani, "Incorporating Structural Models into Research on the Social Amplification of Risk: Implications for Theory Construction and Decision Making," Risk Analysis, vol. 13, no. 6, pp. 611–623, Dec. 1993, doi: https://doi.org/10.1111/j.1539-6924.1993.tb01323.x.

[9] I. J. Chung, "Social Amplification of Risk in the Internet Environment," Risk Analysis, vol. 31, no. 12, pp. 1883–1896, May 2011, doi: https://doi.org/10.1111/j.1539-6924.2011.01623.x.

[10] S. C. Vos et al., "Retweeting Risk Communication: The Role of Threat and Efficacy," Risk Analysis, vol. 38, no. 12, pp. 2580–2598, Aug. 2018, doi: https://doi.org/10.1111/risa.13140.

[11] J. Sutton et al., "A cross-hazard analysis of terse message retransmission on Twitter," Proceedings of the National Academy of Sciences, vol. 112, no. 48, Dec. 2015, doi: https://doi.org/10.1073/pnas.1508916112.

[12] X. A. Zhang and R. Cozma, "Risk sharing on Twitter: Social amplification and attenuation of risk in the early stages of the COVID-19 pandemic," Computers in Human Behavior, vol. 126, Jan. 2022, doi: https://doi.org/10.1016/j.chb.2021.106983.

[13] J. X. Kasperson, R. E. Kasperson, N. Pidgeon, and P. Slovic, "The social amplification of risk: assessing fifteen years of research and theory," The Social Amplification of Risk, pp. 13–46, Jul. 2003, doi: https://doi.org/10.1017/cbo9780511550461.002.

[14] N. Kshetri, "Disinformation and Misinformation in the Age of Artificial Intelligence and the Metaverse," Computer, vol. 57, no. 12, pp. 110–116, Dec. 2024, doi: https://doi.org/10.1109/mc.2024.3461325.

[15] R. Ma, X. Wang, and G.-R. Yang, "Fighting fake news in the age of generative AI: Strategic insights from multi-stakeholder interactions," Technological Forecasting and Social Change, vol. 216, Apr. 2025, doi: https://doi.org/10.1016/j.techfore.2025.124125.

[16] J. Villasenor, "Artificial intelligence, deepfakes, and the uncertain future of truth," Feb. 14, 2019. https://www.brookings.edu/articles/artificial-intelligence-deepfakes-and-the-uncertain-future-of-truth/

[17] D. Dias and P. Mendonça, "Lessons from the Equifax and Capital One Data Breaches on Social Amplification of Risk," Review of Integrative Business and Economics Research, vol. 11, 2022, Available: http://buscompress.com/uploads/3/4/9/8/34980536/riber_11-4_06_t22-044_71-90.pdf

[18] M. Douglas and A. Wildavsky, Risk and culture : an essay on the selection of technological and environmental dangers. Berkeley, Calif.: Univ. Of California Press, 1982.

[19] S. Rippl, "Cultural theory and risk perception: a proposal for a better measurement," Journal of Risk Research, vol. 5, no. 2, pp. 147–165, Apr. 2002, doi: https://doi.org/10.1080/13669870110042598.

[20] B. Colman, "EU AI Act Cheat Sheet: Understanding Regulations in 2 Minutes," Reality Defender — Enterprise-Grade Deepfake Detection, Jun. 04, 2025. https://www.realitydefender.com/insights/understanding-eu-ai-act-in-2-minutes

[21] I. K. Miyashiro, "Case study: Equifax Data Breach," Seven Pillars Institute, Apr. 30, 2021. https://sevenpillarsinstitute.org/case-study-equifax-data-breach/

[22] K. D. Martin, A. Borah, and R. W. Palmatier, "Data Privacy: Effects on Customer and Firm Performance," Journal of Marketing, vol. 81, no. 1, pp. 36–58, Jan. 2017, doi: https://doi.org/10.1509/jm.15.0497.

[23] S. Oladimeji and S. M. Kerner, "SolarWinds hack explained: Everything you need to know," TechTarget, Nov. 03, 2023. https://www.techtarget.com/whatis/feature/SolarWinds-hack-explained-Everything-you-need-to-know

[24] Fortinet, "SolarWinds Supply Chain Attack," Fortinet, 2025. https://www.fortinet.com/resources/cyberglossary/solarwinds-cyber-attack

[25] S. Srinivasan, "Ransomware Attack at Colonial Pipeline Company," Mar. 2023. https://www.hbs.edu/faculty/Pages/item.aspx?num=63756

[26] J. Beerman, D. Berent, Z. Falter, and S. Bhunia, "A Review of Colonial Pipeline Ransomware Attack," IEEE Xplore, May 01, 2023. https://ieeexplore.ieee.org/abstract/document/10181159

[27] D. Citron and R. Chesney, "Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security," California Law Review, vol. 107, no. 6, Dec. 2019, Available: https://scholarship.law.bu.edu/faculty_scholarship/640/

[28] Dark Reading, "Cybercriminals Impersonate Chief Exec's Voice with AI Software," 2019. https://www.darkreading.com/cyber-risk/cybercriminals-impersonate-chief-exec-s-voice-with-ai-software

[29] H. Hart and A. Aronsson-Storrier, "False citations: AI and 'hallucination,'" Society for Computers & Law, Feb. 26, 2024. https://www.scl.org/uk-litigant-found-to-have-cited-false-judgments-hallucinated-by-ai/

[30] B. A. Herrera-Tapias and D. H. Guzmán, "Legal Hallucinations and the Adoption of Artificial Intelligence in the Judiciary," Procedia Computer Science, vol. 257, pp. 1184–1189, Jan. 2025, doi: https://doi.org/10.1016/j.procs.2025.03.158.
[31] T. Khan, "Law, Lies, and Language Models: Responding to AI Hallucinations in UK Jurisprudence," Jun. 12, 2025. https://thebarristergroup.co.uk/blog/responding-to-ai-hallucinations-in-uk-jurisprudence
[32] L. Arvanitis, M. Sadeghi, and J. Brewster, "NewsGuard's Misinformation Monitor: GPT-4 produces more misinformation than predecessor," NewsGuard, Feb. 2024. https://www.newsguardtech.com/misinformation-monitor/march-2023/
[33] C. Metz, "OpenAI Says It Disrupted an Iranian Misinformation Campaign," The New York Times, Aug. 16, 2024. Available: https://www.nytimes.com/2024/08/16/technology/openai-chatgpt-iran-misinformation.html
[34] R. Koppel, "Role of Computerized Physician Order Entry Systems in Facilitating Medication Errors," JAMA, vol. 293, no. 10, Mar. 2005, doi: https://doi.org/10.1001/jama.293.10.1197.
[35] E. M. Redmiles, S. Kross, and M. L. Mazurek, "How I Learned to be Secure," Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security, Oct. 2016, doi: https://doi.org/10.1145/2976749.2978307.
[36] A. Mulahuwaish et al., "A survey of social cybersecurity: Techniques for attack detection, evaluations, challenges, and future prospects," Computers in Human Behavior Reports, p. 100668, Apr. 2025, doi: https://doi.org/10.1016/j.chbr.2025.100668.
[37] D. C. Glik, "Risk Communication for Public Health Emergencies," Annual Review of Public Health, vol. 28, no. 1, pp. 33–54, Apr. 2007, doi: https://doi.org/10.1146/annurev.publhealth.28.021406.144123.
[38] L. Fitzgerald, "How Deepfakes Are Impacting Public Trust in Media," Jan. 17, 2025. https://www.pindrop.com/article/deepfakes-impacting-trust-media/
[39] J. Trevithick, "Sweden's New Civil Defense Guide Tells Citizens To Resist Fake News As They Would An Invasion," The War Zone, Jun. 06, 2018. https://www.twz.com/21343/swedens-new-civil-defense-guide-tells-citizens-to-resist-fake-news-as-they-would-an-invasion (accessed Aug. 03, 2025).
[40] P. Chadha, "The Four Phases of Crisis Management," AGB, Oct. 06, 2020. https://agb.org/blog-post/the-four-phases-of-crisis-management/
[41] M. D. Cavelty, C. Eriksen, and B. Scharte, "Making cyber security more resilient: adding social considerations to technological fixes," Journal of Risk Research, vol. 26, no. 7, pp. 801–814, May 2023, doi: https://doi.org/10.1080/13669877.2023.2208146.