Engage Society for
Risk Awareness and Resilience

Deliverable 3.4 – Directions for innovative communication and social media solutions

Authors: Nathan Stolero (TAU), Tomer Simon (TAU), Bruria Adini (TAU), Rachele Gianfranchi (O2M), Menno Bot (O2M), Solange VanDerKolff (O2M).

Contributors: O2M-EverBridge, SINTEF, TECNUN, DBL, CA.

Abstract: D3.4 revisits the findings of D3.2, adapting to the rapid advancements in AI and chatbot technologies, notably the emergence of GPT-4 and ChatGPT. Aiming to bolster societal resilience, this deliverable refines the AI chatbot blueprint, integrating stakeholder feedback, technological progress, and insights from the eDelphi consensus. The revised blueprint, underpinned by expert consensus, emphasizes user-centricity, real-time information, and misinformation counteraction. The prototype showcases the chatbot's potential in emergency warning systems, while additional solutions explore the roles of SML, VR, and AR in emergency communication. Recommendations advocate for a user-focused, culturally sensitive approach, emphasizing continuous feedback and collaboration for effective AI integration in emergency management.

INFORMATION TABLE

Deliverable Number	3.4
Deliverable Title	Directions for innovative communication and social media solutions
Project Acronym	ENGAGE
Dissemination Level PU: Public; CO: Confidential; CI: Classified	PU
Grant	882850
Project Duration	July 2020 – December 2023
Call	SU-DRS01-2018-2019-2020
Topic	Human factors, and social, societal, and organisational aspects for disaster-resilient societies
Consortium Coordinator	SINTEF
Edition date	30 Aug. 2023
Version	01.01.04

AUTHORSHIP & APPROVAL INFORMATION

EDITOR **Bruria Adini (TAU)**	DATE 22.6.2023 – PCOS 8.8.2023 – INTERMEDIATE 22.8.2023 - EXTERNAL
CONTRIBUTORS **Koen Vogel (O2M), Matthieu Branlat, Martina Ragosta, Jacqueline Floch (SINTEF), Alberto Pasquini, Laura Moens (DBL), Maya Batisti (CA).**	DATE 22.6.2023 – PCOS 20.7.2023 – INTERMEDIATE 24.8.2023 - EXTERNAL

The research leading to these results has received funding from Horizon 2020, the European Union's a Programme for Research and Innovation (H2020/2014-2020) under grant agreement n° 882850.

REVIEWED BY **Leire Labaka (TECNUN)** **Jan Woerlein (EMS)** **Marita Hoel Fossen (TRC)** **Håkon Straume (EverBridge-Norway)**	DATE 7.7.2023 – PCOS 17.8.2023 – INTERMEDIATE 28.8.2023 - External
APPROVED BY **Matthieu Branlat (SINTEF)**	DATE 4.9.2023

ETHICS BOARD REVIEW REQUIRED?	SECURITY BOARD REVIEW REQUIRED?
No	NO

DOCUMENT HISTORY

Version	Date	Version description / Milestone description
00.00.01	22 Jun. 2023	PCOS Proposed
00.00.02	3 Jul. 2023	PCOS Revised
00.00.03	4 Jul. 2023	PCOS Approved
00.01.01	8 Aug. 2023	Intermediate Proposed
00.01.02	14 Aug. 2023	Intermediate Revised
00.01.03	15 Aug. 2023	Intermediate Approved
01.01.01	21 Aug. 2023	External Proposed
01.01.02	28 Aug. 2023	External Revised
01.01.03	30 Aug. 2023	External Approved
01.01.04	11 Sep. 2023	Released

*The project uses a multi-stage internal review process with defined milestones. Milestone names include terms (in bold) as follows:

» **PCOS**

- **proposed:** Describes planned content and structure of different sections. Document authors submit for internal review.

The research leading to these results has received funding from Horizon 2020, the European Union's a Programme for Research and Innovation (H2020/2014-2020) under grant agreement n° 882850.

- **revised:** Document authors produce new version in response to internal review comments.
- **approved:** Internal project reviewers accept the document.

» **Intermediate**

- **proposed:** Document is approximately 50% complete – review checkpoint. Document authors submit for internal review.
- **revised:** Document authors produce new version in response to internal reviewer comments.
- **approved:** Internal project reviewers accept the document.

» **External**

- **proposed:** Document is approximately 100% complete – review checkpoint. Document authors submit for internal review.
- **revised:** Document authors produce new version in response to internal reviewer comments.
- **approved:** Internal project reviewers accept the document.

» **Released:** Executive Board accepts the document. Coordinator releases the deliverable to the Commission Services.

The research leading to these results has received funding from Horizon 2020, the European Union's a Programme for Research and Innovation (H2020/2014-2020) under grant agreement n° 882850.

MEMBERS OF THE ENGAGE CONSORTIUM

SINTEF	SINTEF AS (SINTEF) NO-7465 Trondheim Norway www.sintef.com	**Project Coordinator:** Matthieu Branlat Matthieu.Branlat@sintef.no
deepblue CONSULTING & RESEARCH	Deep Blue Srl (DBL) IT-00198 Rome Italy www.dblue.it	**Contact:** Alberto Pasquini alberto.pasquini@dblue.it
tecnun Universidad de Navarra	University of Navarra (TECNUN) SP-31009 Pamplona Spain www.tecnun.unav.edu	**Contact:** Leire Labaka llabaka@tecnun.es
TEL AVIV UNIVERSITY אוניברסיטת תל אביב	Tel Aviv University (TAU) IL-6997801 Tel Aviv Israel www.english.tau.ac.il	**Contact:** Bruria Adini adini@netvision.net.il
Røde Kors Trondheim	Trondheim Red Cross (TRC) NO-7465 Trondheim Norway www.rodekors.no/en/	**Contact:** Marita Hoel Fossen marita.fossen@redcross.no
eena EUROPEAN EMERGENCY NUMBER ASSOCIATION	European Emergency Number Association (EENA) BE- 1060 Brussels Belgium www.eena.org	**Contact:** Alexis Gizikis ag@eena.org
	Ministry of Internal Affairs, Department for Emergency Situations (DSU) RO- 010086 Bucharest Romania www.dsu.mai.gov.ro	**Contact:** Raed Arafat arafatr@smurd.ro
everbridge	Everbridge Norway (EVBG) NO-0663 Oslo Norway www.everbridge.no	**Contact:** Håkon Straume haakon.straume@everbridge.com
ENS ÉCOLE NORMALE SUPÉRIEURE	Ecole Nationale Supérioure (ENS) FR-75005 Paris France www.ens.psl.eu	**Contact:** J. Peter Burgess james.peter.burgess@ens.psl.eu
ERTZAINTZA	ERTZAINTZA - Departamento de Seguridad – Gobierno Vasco (ERTZ) ES- 01010 San Sebastian Spain www.ertzaintza.eus/wps/portal/ertzaintza	**Contact:** Iñaki Gangoiti igangoiti@seg.euskadi.eus

The research leading to these results has received funding from Horizon 2020, the European Union's a Programme for Research and Innovation (H2020/2014-2020) under grant agreement n° 882850.

	Cittadinanzattiva (CA) IT- 00183 Rome Italy www.cittadinanzattiva.it	**Contact:** Annalisa Mandorino a.mandorino@cittadinanzattiva.it
	Azienda Sanitaria Locale Roma 1 – Dipartimento di Epidemiologia (ASL) IT- 00198 Rome Italy www.aslroma1.it	**Contact:** Francesca de'Donato f.dedonato@deplazio.it
	Katastrofmedicinskt Centrum (KMC) SE-58330 Linköping Sweden www.lio.se/kmc	**Contact:** Carl-Oscar Jonson carl-oscar.jonson@regionostergotland.se
	NTNU Social Research Ltd. (NTNUSR) NO- 7491 Trondheim Norway www.ntnu.edu	**Contact:** Stian Antonsen stian.antonsen@ntnu.no
	One2Many (O2M) NL-7411 CL Deventer The Netherlands www.one2many.eu	**Contact:** Rachele Gianfranchi rachele.gianfranchi@everbridge.com

The research leading to these results has received funding from Horizon 2020, the European Union's a Programme for Research and Innovation (H2020/2014-2020) under grant agreement n° 882850.

Contents

The research leading to these results has received funding from Horizon 2020, the European Union's a Programme for Research and Innovation (H2020/2014-2020) under grant agreement n° 882850.

The research leading to these results has received funding from Horizon 2020, the European Union's a Programme for Research and Innovation (H2020/2014-2020) under grant agreement n° 882850.

The research leading to these results has received funding from Horizon 2020, the European Union's a Programme for Research and Innovation (H2020/2014-2020) under grant agreement n° 882850.

List of Figures

The research leading to these results has received funding from Horizon 2020, the European Union's a Programme for Research and Innovation (H2020/2014-2020) under grant agreement n° 882850.

Executive summary

Background: D3.4 serves as a revision of the results from D3.2, reflecting on the advancements and shifts in the technological landscape over the past two years. Since the release of D3.2, the domain of AI and chatbots has witnessed significant evolution, most notably with the emergence of GPT-4 and ChatGPT. These advancements have redefined the capabilities and potential applications of chatbots, offering more sophisticated, context-aware, and user-friendly interactions. The rapid pace of innovation in this space underscores the need for continuous evaluation and adaptation, ensuring that the strategies and solutions remain relevant and effective in the face of ever-evolving technological frontiers.

Goal: The goal of Deliverable 3.4 is to refine and expand upon the outcomes of T3.2, integrating feedback from the consortium, recent technological advancements, and insights from the eDelphi consensus process. This revision aligns with ENGAGE's objectives to bolster societal resilience by showcasing the impact of project solutions across various disaster contexts. D3.4 emphasizes the enhancement of communication and social media strategies, particularly the AI-enabled chatbot blueprint, to ensure rapid public response and efficient information management during adversities. By refining the chatbot design and proposing a prototype tailored for emergencies, D3.4 aims to bridge informal societal resilience mechanisms with formal authority-driven disaster management efforts, underscoring ENGAGE's dedication to collaborative, technologically advanced, and trustworthy emergency solutions.

Progress and Development Post Deliverable 3.2: Following the completion of Deliverable 3.2, significant advancements were made in refining the AI chatbot's design and functionality. Feedback from various stakeholders, including emergency authorities, first responders, and the public, was instrumental in this iterative development process. Furthermore, additional progress has been made in the fields of social media listening, virtual reality and augmented reality.

Process: The empirical work in this deliverable is based on an eDelphi consensus technique, design of a prototype based on the blueprint developed in D3.2 and revised in D3.4, and a systematic review of the scientific and gray literature for new solutions in the fields of social media listening, virtual reality and augmented reality. The eDelphi method was employed to gather expert consensus, resulting in a revised blueprint that better aligns with user needs and expectations. This blueprint served as the foundation for the prototype chatbot, which underwent rigorous testing and validation. Additionally, innovative solutions incorporating Social Media Listening (SML), Virtual Reality (VR), and Augmented Reality (AR) were explored to further enhance the chatbot's capabilities and user experience.

The Results of the eDelphi Process: The eDelphi process, which involved two rounds of expert consultations, yielded a consensus on the design and functionality of the AI chatbot for emergencies and disasters. Experts emphasized the importance of the chatbot being user-centric, ensuring it provides accurate, timely, and contextually relevant information. They also highlighted the need for the chatbot to be adaptable to various emergency scenarios, integrate seamlessly with existing emergency communication systems, and possess the capability to counteract misinformation. The feedback from this process was instrumental in refining the chatbot's blueprint, ensuring it aligns with the needs of both emergency professionals and the general public.

The Revised Blueprint: Building upon the feedback from the eDelphi process, the updated blueprint for the AI chatbot was designed to be more robust, adaptable, and user-friendly. The revised design emphasizes a modular approach, allowing for easy integration with various emergency communication systems. It also prioritizes real-time information dissemination, ensuring that users receive accurate and contextually relevant data during emergencies. Enhanced features include the ability to counteract misinformation, a user-centric interface, and adaptability to a wide range of emergency scenarios. The blueprint also outlines the technological infrastructure, data sources, and algorithms that power the chatbot, ensuring its efficiency and reliability.

The research leading to these results has received funding from Horizon 2020, the European Union's a Programme for Research and Innovation (H2020/2014-2020) under grant agreement n° 882850.

Prototype: The prototype, developed as an extension of the revised blueprint, serves as a tangible representation of the envisioned AI chatbot for emergency warning systems. It's designed to provide real-time alerts and information during crises, leveraging advanced algorithms to ensure accuracy and timeliness. The prototype integrates seamlessly with existing emergency communication infrastructures and utilizes a user-friendly interface to enhance public engagement. Its capabilities include detecting and countering misinformation, offering multilingual support, and providing context-specific guidance to users. The prototype's development underscores the potential of AI in bolstering societal resilience during emergencies.

Additional SML, VR and AR Solutions: In the exploration of additional solutions, the document delves into the potential of Social Media Listening (SML), Virtual Reality (VR), and Augmented Reality (AR) tools in the context of emergency communication. SML tools can assist authorities in real-time by monitoring public sentiment, enabling them to swiftly address concerns and counter misinformation. VR and AR offer immersive experiences beneficial for public awareness campaigns, training, and even real-time guidance during emergencies. However, the effective deployment of these technologies necessitates a profound comprehension of public perceptions, cultural nuances, and technological challenges. For instance, certain cultures might be skeptical of AI-driven tools, while others may face technological accessibility issues. The ultimate goal is to seamlessly integrate these tools into emergency communication strategies, fostering safer and more resilient communities.

Recommendations: Authorities and emergency responders should adopt a cautious yet progressive approach towards integrating AI chatbots and associated technologies. The eDelphi results underscore the importance of a user-centric design, ensuring that the chatbot addresses genuine public concerns and needs. The revised blueprint offers a comprehensive guide, but its implementation should be tailored to specific regional and cultural contexts. The potential of SML, VR, and AR tools is vast, but their deployment should be grounded in a deep understanding of public perceptions and technological accessibility. Continuous monitoring, feedback collection, and iterative improvements are crucial to ensure the chatbot's efficacy and trustworthiness. Lastly, collaboration between tech developers, emergency responders, and the public is vital for the holistic development and success of these technological solutions.

The research leading to these results has received funding from Horizon 2020, the European Union's a Programme for Research and Innovation (H2020/2014-2020) under grant agreement n° 882850.

1 INTRODUCTION

1.1 SCOPE OF THE DELIVERABLE

The deliverable reports on the results of Task 3.4: "Innovative use of communication and social media technologies". It builds upon and expands on T3.2, "Exploration of innovative use of communication and social media technologies". Mainly, it expands the knowledge about the innovative use of communication and social media technologies, in several ways:

(A) Revision of Results from D3.2: This includes a thorough examination of the results previously obtained, with an emphasis on identifying both the strengths and weaknesses of the AI-enabled chatbot blueprint for emergencies and disasters we designed. The insights gained from this analysis inform the subsequent development and refinement of the technologies considered. This includes improved design directions for communication and social media technology, tailored to the unique challenges and opportunities presented by emergencies and disasters.

(B) Innovative Communication and Social Media Solutions: The deliverable outlines new directions for enhancing communication between authorities and first responders to the public, and within the community members, through innovative communication and social media solutions.

(C) Transfer of Research Results to the Public: A key aspect of Deliverable 3.4 is the reporting on approaches for making research findings accessible to the public. This aligns with the broader goal of contributing to societal resilience by ensuring that the knowledge generated is not confined to academic circles but reaches those who can benefit from it most.

As part of the deliverable we utilized a modified eDelphi process, in order to seek consensus concerning the designed blueprint of an AI chatbot for emergencies and disasters developed in D3.2. This process involved engaging experts from various fields, including AI, machine learning, and emergency management, to refine and validate the proposed blueprint.

The deliverable also includes a detailed description of a prototype AI chatbot, showcasing the practical application of the theoretical concepts and design principles outlined in the blueprint developed in D3.2. This prototype serves only as a tangible example of how the guidelines and recommendations can be translated into a functional technology, in a specific case study of warning systems.

The intended readers of Deliverable 3.4 include the ENGAGE Consortium, the European Commission and project reviewers, EU emergency authorities, first responders, NGOs, and the public. The document aims to serve as a comprehensive guide, offering valuable insights and actionable recommendations for leveraging AI and social media technologies to enhance societal resilience in the face of emergencies and disasters.

1.2 GOALS

The goal of Deliverable 3.4 is to revise the outcomes of T3.2. This will be achieved through the evaluation conducted in WP4, incorporating feedback from the consortium, incorporating cutting-edge technological advancements, and utilizing the eDelphi consensus procedure. Additionally, we aim to enrich the communication and social media solutions previously collected.

The research leading to these results has received funding from Horizon 2020, the European Union's a Programme for Research and Innovation (H2020/2014-2020) under grant agreement n° 882850.

1.3 Objectives

This deliverable continues, as in D3.2, to follow ENGAGE's objectives to produce validated actionable knowledge on societal resilience by demonstrating the benefits and impact of the project solutions in different types of disasters, in addition to following one of the desired Knowledge and Innovation Community of Practice (KI-CoP). By revising the AI-enabled chatbot's blueprint and enriching the communication and social media solutions, we offer to continue to follow these objectives.

1.4 Fit within ENGAGE

D3.4 contributes to revising the directions for innovative communication and social media solutions for communication with citizens before, during and following disasters. Therefore, it corresponds with the following deliverables:

- **D2.5 Revision and update of solutions to improve societal resilience:** Deliverable 3.4 builds on additional exploration of formal solutions as extended in D2.5 by further analyzing how these solutions can be integrated and adopted in the context of innovative communication and social media.
- **D3.2 Initial directions for innovative communication and social media solutions:** Deliverable 3.4 directly extends the work of D3.2 by revising the blueprint of AI-enabled chatbot technology for enhancing societal resilience. The design concept and blueprint of an AI chatbot for emergencies and disasters developed in D3.2 serve as the basis for further refinement and development in D3.4. The strengths and weaknesses identified in existing AI solutions are revisited, and new directions for machine learning approaches are explored, leading to a more comprehensive and practical prototype.
- **D3.3 Final catalogue of solutions:** D3.4 corresponds with D3.3 by focusing on the revision of initial results based on feedback from evaluation efforts and validation of needs and challenges. The updated set of proposed solutions and recommendations in D3.4 informs the final version of the catalogue of solutions.
- **D4.3 Final validation of solutions:** The validation exercises defined and executed in D4.2 provide essential input for D3.4. The scenarios and exercises involving various actors, including first responders and citizens, contribute to the evaluation of the blueprint and the additional communication and social media solutions. The feedback and outcomes of these exercises support D3.4 in extracting insights and refining the approach for engaging citizens, enhancing the alignment of the AI chatbot prototype with real-world needs and practices. In addition, the results of the experiment in the chatbot's prototype for warning system, based on the blueprint, will be reported in D4.3.
- **D5.2 Communication and dissemination activities:** D3.4 aligns with D5.2 by contributing to the overall communication and dissemination strategy of the project. The innovative solutions and recommendations developed in D3.4 are integral to the communication and dissemination activities, ensuring that the project's advancements are communicated to relevant stakeholders with the most innovative cutting-edge social media technologies.
- **D5.5 Knowledge and Innovation Community of Practice (KI-CoP):** The collaboration strategy, workshops, and webinars organized in D5.5 provide valuable platforms for feedback and engagement that inform D3.4. The active participation of citizens, authorities, first responders, and other stakeholders in these activities ensures that the deliverable reflects diverse perspectives and needs. The focus on societal resilience, validation, experimentation, and exploitation in the workshops aligns with the goals and content of D3.4, fostering a collaborative and informed approach to developing and implementing innovative solutions.

The research leading to these results has received funding from Horizon 2020, the European Union's a Programme for Research and Innovation (H2020/2014-2020) under grant agreement n° 882850.

The research leading to these results has received funding from Horizon 2020, the European Union's a Programme for Research and Innovation (H2020/2014-2020) under grant agreement n° 882850.

2 SIGNIFICANCE

2.1 CONTRIBUTION TO THE FIELD

Based on the initial directions suggested in D3.2, deliverable 3.4 suggests a revised version of the AI-enabled blueprint, which fits the consensus of experts in the fields of AI, Machine Learning and Chatbots. Therefore, reflecting a version which is more fit to the technological experts. In addition, this deliverable contributes to the expansion of the knowledge of innovative communication and social media solutions, aiming to improve the pace of response to the public's queries and decrease the load on information centres during adversities, thus increasing the capacity to manage the surge of seeking data before and during disasters effectively.

2.2 SPECIFIC CONTRIBUTION TO THE ENGAGE PROJECT

Deliverable 3.4 contributes to project ENGAGE by furthering the exploration and refinement of innovative communication and social media solutions, particularly focusing on the strengths, weaknesses, and recommendations for the development and implementation of AI-enabled chatbot technologies. Building on the foundational work of previous deliverables, D3.4 aligns with ENGAGE's aim to link informal societal resilience with the formal work of authorities in disaster prevention, preparation, response, and recovery. By revising the results from D3.2, improving the design of the AI-enabled chatbot's blueprint and suggesting a prototype of an AI chatbot for emergencies and disasters, D3.4 supports the project's objectives of validating solutions and demonstrating their benefits and impact in different disaster scenarios. The deliverable's focus on innovative design directions, improved communication channels, and the transfer of research results to the public resonates with ENGAGE's commitment to fostering collaboration, enhancing societal resilience, and leveraging technology to create effective and trustworthy solutions for emergency management.

The research leading to these results has received funding from Horizon 2020, the European Union's a Programme for Research and Innovation (H2020/2014-2020) under grant agreement n° 882850.

3 Previous Work: Summary of Deliverable 3.2

3.1 Theoretical Base

From a theoretical perspective, D3.4 is based similarly on the scientific review which was conducted as part of D3.2.

3.1.1 Artificial Intelligence and Machine Learning

Artificial Intelligence (AI), a term coined in computer sciences, refers to the simulation of human-like intelligence and capabilities by machines, be it software or hardware. AI enables these machines to mimic human cognitive functions such as perception, learning, problem-solving, and decision-making. Essentially, AI is the process of training machines to understand and respond to requests, identify and learn languages, recognise objects, and improve their decision-making process by learning from examples, sometimes based on trial and error (IBM, n.d.). Hancock, Naaman and Levy (2020), expands this definition to AI-mediated communication (AI-MC) as "mediated communication between people in which a computational agent operates on behalf of a communicator by modifying, augmenting, or generating messages to accomplish communication or interpersonal goals" (p. 90).

Machine Learning (ML), a subset of AI, is the process that allows technological systems to learn from experiences and data, improve themselves, and solve problems without explicit programming. ML enables the machine to teach itself, using existing datasets for training or learning from new data. The three dominant approaches for ML are supervised learning, unsupervised learning, and reinforcement learning. All these learning methods involve observing datasets, examples, and instructions. In supervised learning, all data is labelled in advance, while in unsupervised learning, the algorithm identifies the nature of the data. Semi-supervised learning combines these approaches with a small amount of labelled data and a significant amount of unlabeled data. Reinforcement learning, on the other hand, is based on trial and error and constant updates of the algorithms, based on the results of the learning (Jordan & Mitchell, 2015; Mohri, Rostamizadeh & Talwakar, 2018; Van-Engelen & Hood, 2020).

AI and ML technologies, which employ various statistical measures and algorithms, have found applications in fields that were traditionally dependent on human-assisted help, such as customer service and sales. The literature distinguishes between three types of AI: weak AI (or narrow AI), strong AI (or general AI), and super AI (or artificial general intelligence). This deliverable aligns with weak AI, which focuses on simulating human behaviour for a specific task. Strong AI, which exhibits the abilities of human intelligence, and super AI, which goes beyond human cognitive capabilities, are primarily theoretical concepts (Al-Rifaie &, 2015; Liu, 2021; Zifu, 2016).

3.1.2 AI-enabled chatbots in emergencies and disasters

Artificial Intelligence (AI) enabled chatbots, also known as conversational agents, are applications that use AI for text and/or voice-based communication. These chatbots are increasingly being used across various sectors, including customer service, healthcare, and policymaking, due to their ability to support omni-channel communication. They range from basic chatbots that provide answers based on a predefined list of questions to advanced chatbots that can converse with users, understand their queries, provide answers, and decide on the necessary action following a conversation. These abilities are made possible through Machine Learning (ML) approaches and the continuous development of cutting-edge technologies (Almalki & Azeez, 2020; Gupta, Hathwar & Vijayakumar, 2020; Følstad, Skjuve & Brandtzaeg, 2018).

The research leading to these results has received funding from Horizon 2020, the European Union's a Programme for Research and Innovation (H2020/2014-2020) under grant agreement n° 882850.

Despite their significant development in other fields, AI-enabled chatbots have been infrequently used in emergencies and disasters. However, they hold considerable potential in this area. For instance, they can provide information to help citizens and authorities prepare for emergencies, give moral support during emergencies, collect information, and detect environmental changes that require special attention. An example of this is the AI-enabled chatbot "Hakeem", developed by The Norwegian Refugee Council, Microsoft, NetHope and University College of Dublin, which provided education options in areas lacking school and university infrastructure (Hofeditz, Ehnis, Bunker, Brachten & Stieglitz, 2019). Most important, they can offer a set of responses when a human response is not available at all, due to surge in requests following an emergency alert.

However, AI-enabled chatbots still have several limitations. This could be due to a lack of data and implementation, which hinders the learning processes that can improve the AI-enabled chatbots' actions in emergencies and disasters. Another possible reason could be the apprehension of first responders and the public's distrust, preferring a human response during emergencies, if that response is available (Madianou, 2020; Palanica, Flaschner, Tommandram, Li & Fossat, 2019; Abd-Alrazaq, Alajani, Ali, Denecke, Bewick & Househ, 2021; Cheng & Jiang, 2020).

3.1.3 AI-ENABLED CHATBOTS AND SOCIETAL RESILIENCE

The intersection of Artificial Intelligence (AI) technologies, specifically AI-enabled chatbots, and societal resilience is not immediately apparent due to their distinct natures. AI, with its instrumental technological character, is focused on identifying problems and suggesting solutions, while societal resilience is an organic societal phenomenon related to the natural responses and behaviours of the public (Vinuesa, Azizpour, Leite, Balaam, Dignum, Domisch & Nerini, 2020).

The enablement of communication capabilities is what bridges this apparent distinction. AI-enabled chatbots have the potential to transform the communication process between authorities, first responders, and the public during emergencies and disasters. They can provide accurate information, prevent the collapse of communication means during surges of requests, and neutralise false information. However, the adoption process of AI-enabled chatbots should consider both their positive and negative effects on societal resilience (Ignatidou, 2019; Kertysova, 2018; Annoni, Benczur, Bertoldi, Delipertrev, De Prato, Feijoo & Junklewitz, 2018).

The European Commission Joint Research Centre (JRC) suggests a holistic, complex and system view approach when considering the advantages and potential negative implications of AI technologies. The JRC report concludes that "looking at AI from a resilience perspective does not only enrich our understanding of, and preparedness for AI, but also advances the analysis of resilience (Alessi et al., 2018, p. 119). Access to essential information in an emergency can be considered as a societal resilience-enhancing element.

AI-enabled chatbots have been primarily popular in health emergencies and disasters, serving as online self-symptoms tests, information providers, and Q&A tools. The most recent example is their use in providing accurate information about COVID-19 to the public. However, the adoption of AI-enabled chatbots should be done cautiously, considering their potential contributions and limitations in the context of societal resilience (Almalki & Azeez, 2020; Battineni, Chintalapudi & Amenta, 2020).

3.1.4 IMMEDIATE INFORMATION

Rapid dissemination of information is a critical factor contributing to societal resilience, as supported by the findings of Deliverable 1.3. The survey participants prioritised their cognitive needs for information and the need for a fast, unidirectional flow of communication. However, the public's ability to receive immediate answers to their questions presents several challenges, such as unawareness of various information sources and their availability. While social media are quick and adaptable tools for information dissemination, they lack the organisation of websites and offline sources. Social media can lead to a cacophony of messages from sources which can be inaccurate, or deliberately misleading (hybrid warfare), contradictory, confusing, and possibly undermining societal resilience by generating panic in emergency situations. AI-enabled chatbots can address these challenges by better organising information, and scanning various online sources to deliver relevant, official and trustworthy, updated information to users, thereby enhancing societal resilience (Kertysova, 2018; Hasan, Rizvi, Jain & Huria, 2021; Almalki & Azeez, 2020).

3.1.5 Collapsed means of communication during surges

Emergency call centres, while capable of handling a reasonable number of calls during routine times, can collapse under a surge of calls during emergencies and disasters, a situation referred to as "infrastructure resilience" (MacKenzie & Barker, 2013). This overload can also result from technical failures or cyber-attacks. The internet, a more resilient infrastructure and the basic technology of AI-enabled chatbots, continues to function in such scenarios. Some research suggests augmenting or replacing the work of emergency call centres with social media solutions or AI-enabled chatbots, which can handle non-emergency calls for information and distress signals, reducing the workload on call centres and diversifying communication channels (Chaudhry & Yuskal, 2019; Richardson, 2019; Heires, 2017; Imran, Castillo, Lucas, Meier & Viewg, 2014). However, these alternative communication methods should support, not entirely replace, traditional communication channels, as they too are subject to potential collapse and external attacks.

3.1.6 Neutralising false information

False information and rumours pose significant challenges to societal resilience (Kertysova, 2018). The three main types of false information are misinformation, disinformation, and fake news (Shu, Wang, Lee & Liu, 2020; Stahl, 2006). Misinformation is unintentionally incorrect, while disinformation is intentionally misleading. Fake news, which overlaps with disinformation, mimics news media content but lacks editorial norms (Lazer et al., 2018; Rogers & Niederer, 2020). A recent development in false information is "deepfake", where AI and ML technologies manipulate audio-visual content to mislead (Westerlund, 2019; Dolhansky et al., 2020; Frank et al., 2020; Maksutov et al., 2020).

False information impedes societal resilience by preventing accurate information dissemination, misleading people during emergencies, and lowering morale. The EU-funded project "Builders" found that those who do not use multiple news sources and are less skilled are most vulnerable to online misinformation (DSR01 Deliverable 1.4). AI-enabled technologies, including chatbots, can both detect and remove false information (Aphiwongsophon & Chongstitvatana, 2018; Mahabub, 2020; Patwa et al., 2021) and potentially spread it if misused or poorly designed (Ahmen et al., 2021; Coeckelbergh, 2020; Sample et al., 2020).

The issue of false information is complex, with technical and ethical challenges. Misclassifications can occur due to bias (Figueira & Oliveira, 2017; Hakak et al., 2020), and ethical dilemmas arise in deciding what is true or false for society (Stroud, 2019). Fact-checking projects, often manual, are globally popular for identifying false information (Graves, 2018). However, Kertysova (2018) argues for automation due to increasing information volume. Major tech companies invest heavily in developing ML algorithms for this purpose, but these algorithms have been criticised for bias (Sumpter, 2018). The EU has developed ethical guidelines for AI use to address these issues (European Commission, 2018).

The research leading to these results has received funding from Horizon 2020, the European Union's a Programme for Research and Innovation (H2020/2014-2020) under grant agreement n° 882850.

3.2 Users' perspectives of AI-enabled chatbots

AI-enabled chatbots have seen widespread use, particularly during the COVID-19 pandemic, indicating a potential level of trust in these technologies. However, Youn & Jin (2021) argue that the field is underexplored despite the high use of AI-enabled technologies in user interactions and customer relationship management (CRM). Their experiment showed that AI-enabled chatbots tend to build virtual assistantship relationships rather than virtual friendships, receiving lower scores for para-social interaction, satisfaction, trust, and behavioural intentions.

There are significant limitations to AI-enabled chatbots when a human alternative is realistic. They are often less preferred when a human alternative is available, have lower trust scores than human sources, and meet fewer user needs (Lei, Shen & Ye, 2021; Aoki, 2020; Ashfaq, Yun & Loureiro, 2020; Brandzaeg & Følstad, 2018; Laumer, Maier & Gubler, 2019). However, when compared to human services, AI-enabled chatbots have several advantages. They are perceived as a good tool for basic information, particularly for younger populations, and offer a positive, friendly interaction (De Cicco, Silva & Alparone, 2020; Thies, Menon, Magapu, Subramony & O'neill, 2017). They are also seen as more available due to their lack of operating hours (Van-Wezel, Croes & Antheunis, 2020), and user rates tend to be higher if a "fallback" option to a human operator is available (Hill, Ford & Farreras, 2015).

Despite the barriers to adoption, most users would be receptive to using AI-enabled chatbots, although hesitancy regarding this technology is likely to compromise engagement (Nadarzynski, Miles, Cowie & Ridge, 2019). Therefore, it is recommended to integrate AI-enabled chatbots moderately to gain public trust. However, it's important to note that while they have a more resilient infrastructure, AI-enabled chatbots can also collapse or be subjected to cyber-attacks (Verma, Chandra & Joshi, 2021; Yamin, Ullah, Ullah & Katt, 2021).

3.3 Previous AI-enabled chatbot blueprint

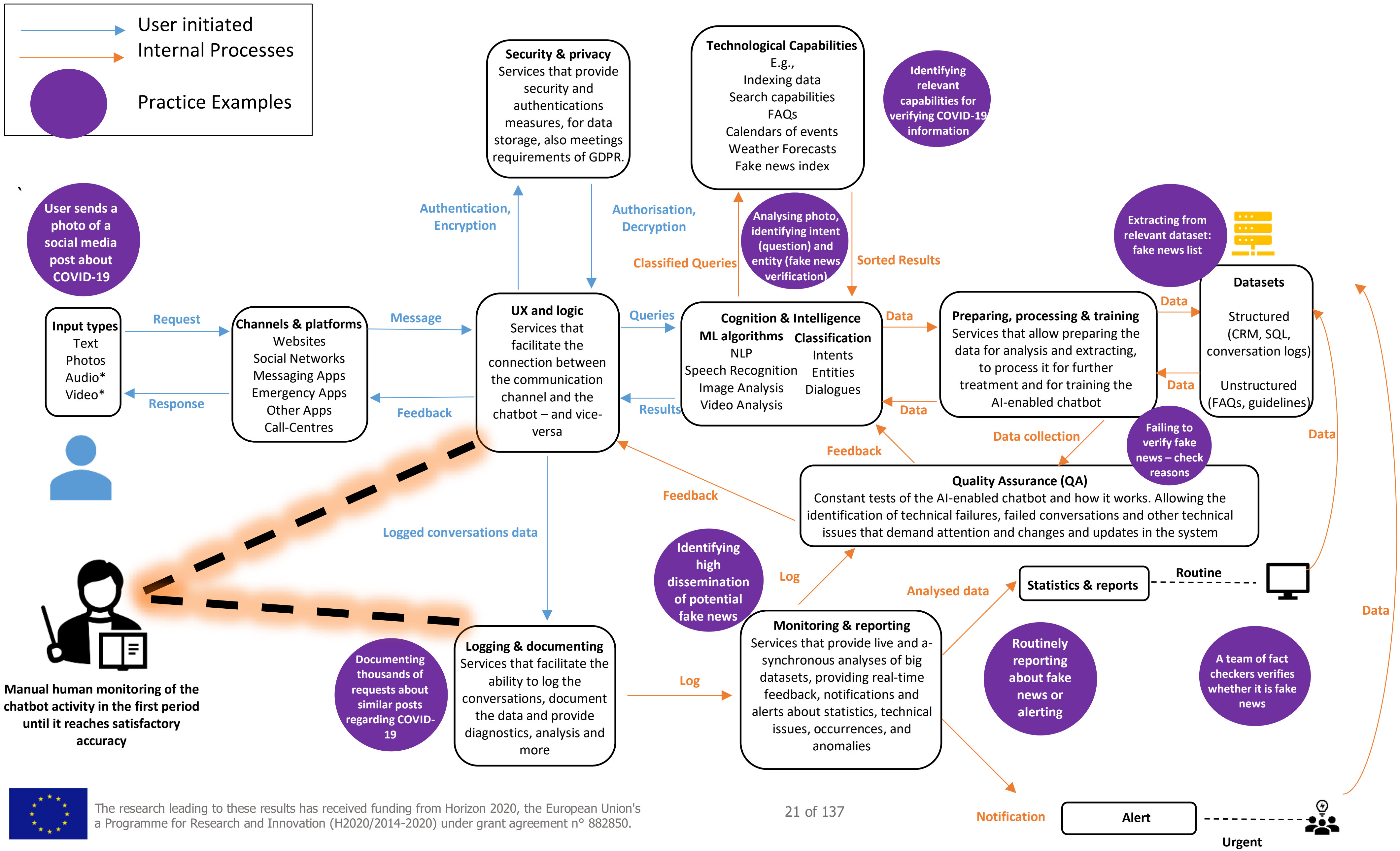
The blueprint for the AI-enabled chatbot for emergencies and disasters, which was presented in D3.2 (Figure 1), is a comprehensive design concept that addresses the design and implementation of the chatbot. It is based on two critical working assumptions. The first assumption is that despite the concerns and barriers expressed by many authorities and first responders, AI-enabled chatbots use a popular technology in the private sector, which answers some of the perceived barriers of the public, and which are not yet common in the emergency management environment. The second assumption is the need to validate that the chatbot is doing what it should do at a satisfactory level. Therefore, the primary initial recommendation is that the chatbot is repetitively tested in simulations and surveys to collect the data to accurately train it, ideally it should collaborate with a human call centre that will inspect that it is functioning correctly, providing accurate answers, maintaining the treatment of false information and more.

The blueprint was designed to address these barriers in several ways. It elaborated on how the technological processes operate, explaining how they can address those concerns, suggested a functioning AI-enabled chatbot, and under which conditions. It also suggested a roadmap and a step-by-step adoption model that allowed authorities and first responders to adopt and adapt to "thinner" solutions. This allows facing one barrier at a time, moving to the second step only after overcoming it.

The original blueprint included various components such as input types, connecting the chatbot to communication channels, the chatbot's logic and UX, security and privacy, logging and documentation, existing platforms, QA, monitoring and reporting, datasets, statistics, alerts, and manual human monitoring. The full description of the components appears in D3.2, and here, will be summarised only for the purpose of the reminder.

The research leading to these results has received funding from Horizon 2020, the European Union's a Programme for Research and Innovation (H2020/2014-2020) under grant agreement n° 882850.

Figure 1. Original blueprint from D3.2.



The research leading to these results has received funding from Horizon 2020, the European Union's a Programme for Research and Innovation (H2020/2014-2020) under grant agreement n° 882850.

3.3.1 Input Types

The "Input Types" section of the blueprint is a crucial part of the AI-enabled chatbot's framework. It serves as the entry point for users and organizations to interact with the chatbot. The chatbot can receive a variety of input types from users, including text, audio messages, photos, and videos. The blueprint emphasizes the importance of allowing multiple methods for the public to report or inquire about emergencies and disasters, as different users may have different communication needs. For instance, a user might need to send a location using their phone's GPS if they can't explain their location verbally. The blueprint also acknowledges that communication might not always be initiated by individuals; in some cases, authorities or first responders might trigger the communication.

3.3.2 Connecting the blueprint into communication channels

The blueprint for the AI-enabled chatbot emphasizes the importance of connecting the chatbot to various official communication channels. This is considered the "body" of the AI-enabled chatbot, linking all the internal and external components of the framework. The ability to interact with authorities and first responders through multiple channels, rather than just one, is perceived to enhance the communication process between the public and these entities. This multi-channel approach is not limited to a single platform, but adding one AI-enabled chatbot can facilitate and support conversations across all platforms. AI-enabled chatbots can be implemented on various platforms, from websites to popular messaging apps or even integrated as part of call centers.

3.3.3 Logic and UX

The blueprint of the AI-enabled chatbot for emergencies and disasters places significant emphasis on the logic and user experience (UX) of the system, the "brain" of the chatbot. This component of the blueprint is crucial as it facilitates the interaction between the official emergency communication channel, the public, and the chatbot, creating a feedback loop. It is responsible for understanding the user's request, processing it, and generating an appropriate response. The logic and UX component also plays a significant role in tailoring the information to individuals and groups. For example, when a citizen calls an emergency service or inquires for information in a call centre of authority, one of the first tasks of the human responder is to understand what the citizen needs. The responder then needs to provide the information and examine whether it answered the citizen's needs. The logic and UX component of the chatbot is designed to mimic this process, understanding the user's needs and providing tailored information accordingly.

3.3.4 Security and privacy

The section on "Security and privacy" in the blueprint emphasizes the importance of securing data and authenticating users, given the heavy reliance of AI-enabled chatbots on large datasets. This is a crucial consideration due to the need to protect information from leaks and to comply with privacy regulations such as the General Data Protection Regulation (GDPR). Authorities and first responders cannot use such technological systems if they do not meet these regulations. The security and authentication components are designed to handle different types of data, including personal and other sensitive data. They help prevent unauthorized access to the system or data that regular users should not access.

The research leading to these results has received funding from Horizon 2020, the European Union's a Programme for Research and Innovation (H2020/2014-2020) under grant agreement n° 882850.

3.3.5 Logging and documentation

The "Logging and Documentation" component of the blueprint is crucial for managing and analyzing the vast amount of data generated by the AI-enabled chatbot. Given the potential for millions of conversations, the chatbot needs a robust system for storing and processing this information. This includes not only the raw data of the conversations themselves but also the details and metadata associated with each interaction.

3.3.6 Cognition & Intelligence

The cognition and intelligence component is the core of the AI-enabled chatbot, responsible for understanding and processing user inputs. It uses machine learning algorithms, natural language processing, and other AI technologies to interpret user requests and generate appropriate responses.

3.3.7 Technological capabilities

The technological capabilities component refers to the various features and functionalities that the chatbot can offer. This could include indexing data, search capabilities, calendars of events, weather forecasts, and more. These capabilities are crucial for providing users with the information they need during emergencies and disasters.

3.3.8 Preparing, processing & training

The preparing, processing, and training component involves the preparation of data for analysis, processing it for further treatment, and training the AI-enabled chatbot. This component ensures that the chatbot is equipped with the necessary data and knowledge to handle a wide range of queries and scenarios.

3.3.9 Quality Assurance (QA)

The Quality Assurance (QA) process is a crucial part of the blueprint for an AI-enabled chatbot. This process involves the analysis, examination, and if necessary, forwarding of the results of the chatbot's actions and decisions to other parts of the system for improvement. The QA process is not only about identifying and correcting errors, mistakes, or wrong decisions, but it also serves to strengthen the successful aspects of the chatbot.

3.3.10 Monitoring and reporting

The monitoring and reporting component of the AI-enabled chatbot blueprint is crucial for maintaining the system's effectiveness and reliability. This component involves analyzing the logged data, identifying patterns, and detecting anomalies. It pinpoints messages that might require special attention or human intervention, such as emergencies, situations posing danger to the user, or cases where the chatbot cannot provide the necessary assistance and needs to transfer the user to a human-assisted conversation. The monitoring component can also connect to specific sources like social networks, discussion boards, and news platforms, providing similar capabilities. For instance, it can identify repeated patterns of information that do not match the official information disseminated by authorities and first responders, such as suspicious misinformation or fake news.

The research leading to these results has received funding from Horizon 2020, the European Union's a Programme for Research and Innovation (H2020/2014-2020) under grant agreement n° 882850.

3.3.11 Datasets

The blueprint of the AI-enabled chatbot for emergencies and disasters places a significant emphasis on the role of datasets. Datasets are the core information that the AI-enabled chatbot relies on to create meaningful and helpful conversations with users. These datasets can include a wide range of information that the organization stores during its activity, such as guidelines, documentation of previous events, logged and transcribed conversations with callers, fact sheets, statistics, and many other data types.

3.3.12 Statistics

The blueprint's "Statistics" section focuses on the importance of data analysis and monitoring in the training and operation of the AI-enabled chatbot. The chatbot's functionality heavily relies on data and information, and thus, data management is a crucial part of the blueprint. This includes extracting data from various types of datasets, defining the type of data to be transferred, the components it should be transferred to, the purposes for the transfer, and the timing.

3.3.13 Alerts

The "Alerts" component of the blueprint is designed to provide real-time feedback, notifications, and alerts about various aspects of the AI-enabled chatbot's operation. This includes statistical data, technical issues, occurrences, and anomalies. Alerts can be triggered by various events, such as high dissemination of potential fake news or repeated patterns of information that do not match the official information disseminated by authorities and first responders.

3.3.14 Manual human monitoring

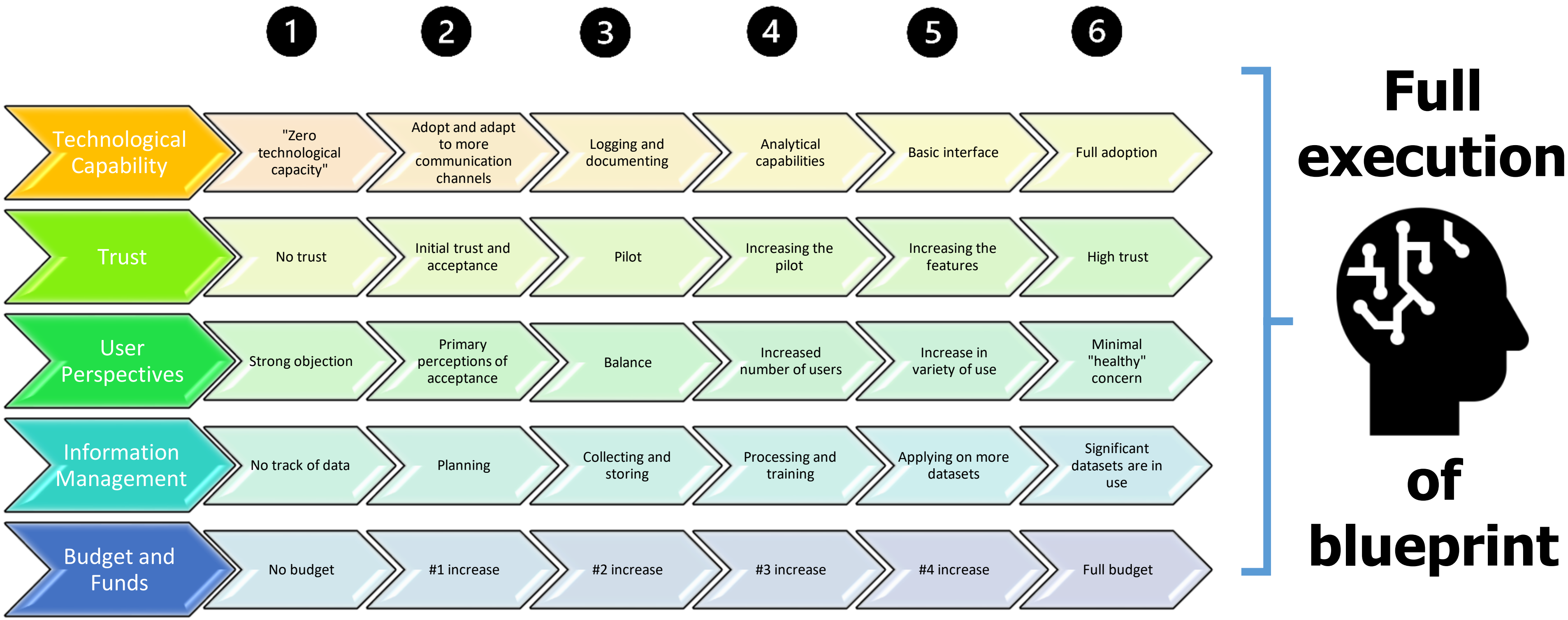
The blueprint of the AI-enabled chatbot for emergencies and disasters also includes a component for manual human monitoring. This component is crucial for ensuring the chatbot's effectiveness and accuracy. The monitoring process involves reviewing the logged data, identifying patterns, discovering anomalies, and pinpointing messages that might require special attention or human observation. This could include emergencies, situations that pose a danger to the user, or cases where the chatbot cannot provide the necessary assistance and needs to transfer the user to a human-assisted conversation. Moreover, the improvement in combining more algorithms, learning approaches, and datasets should be done in steps, with manual human inspection and monitoring for each stage.

3.3.15 Roadmap

The blueprint for the AI-enabled chatbot for emergencies and disasters provides a roadmap for the complete execution of the technology. It acknowledges that while the blueprint represents a complex AI-enabled chatbot with many features and technological capabilities, simpler uses are also possible and can offer substantial value to the public (Figure 2).

The research leading to these results has received funding from Horizon 2020, the European Union's a Programme for Research and Innovation (H2020/2014-2020) under grant agreement n° 882850.

Figure 2. Roadmap from D3.2 for the complete execution of the blueprint.



The research leading to these results has received funding from Horizon 2020, the European Union's a Programme for Research and Innovation (H2020/2014-2020) under grant agreement n° 882850.

3.4 Other social media solutions and open questions

The use of social media-based technologies for communication with the public in emergencies and disasters, and building societal resilience, extends beyond AI-enabled chatbots. Deliverable 3.2 generally reviewed several other technologies that have been identified in the literature that can play a significant role in these contexts. These include social media listening/monitoring (SML), virtual reality (VR), and augmented reality (AR).

These technologies present exciting opportunities for enhancing communication with the public in emergencies and disasters and building societal resilience. However, there are still many open questions about how these technologies can be best utilized in these contexts, which will be discussed further in the deliverable.

3.4.1 Social Media Listening (SML)

Social media listening involves using AI capabilities to monitor social media channels for tags, mentions, topics, and specific keywords. This technology can be particularly useful in emergencies, where social media listening can be conducted to identify tags, keywords, and topics, and employ them with a range of algorithms, such as sentiment analysis, to produce insights that can highlight necessary actions that should be taken (Krüger & Albris, 2020; Rao, 2016). This technology can also be used for false news detection, which is crucial in the context of emergencies and disasters (Barojan, 2021; Luccioni, Pham, Lam, Aylett-Bullock & Luengo-Oroz, 2021).

3.4.2 Use of VR and AR technologies

Virtual Reality (VR) and Augmented Reality (AR) are emerging technologies that can be used in the context of emergencies and disasters. For example, a study conducted in Australia found that the use of VR and AR in online disaster education can increase public awareness and enhance disaster resilience (Kankanamge, Yigitcanlar, & Goonetilleke, 2020). Another study identified VR and AR as critical technologies that can deliver real benefits to the humanitarian supply chain (Argumedo-García, Salas-Navarro, Acevedo-Chedid, Ospina-Mateus, 2021).

The research leading to these results has received funding from Horizon 2020, the European Union's a Programme for Research and Innovation (H2020/2014-2020) under grant agreement n° 882850.

4 PROGRESS AND DEVELOPMENT POST DELIVERABLE 3.2

This chapter discusses the progress and developments made after Deliverable 3.2, including new academic studies, AI and ML solutions, and the development of ChatGPT and its relevancy to this project. These sections focuses on advancements which were made outside of project ENGAGE.

4.1 NEW PUBLISHED ACADEMIC STUDIES

Since the publication of D3.2, several advancements have been identified by the scientific literature in the uses and implications of applying social media innovative technologies in the communication process with the public, in all phases of emergencies and disasters. For start, we begin with the added scientific information, identified in the academic literature.

The advent of artificial intelligence has revolutionized various sectors, including healthcare, disaster management, and emergency reporting. AI-enabled chatbots, in particular, have emerged as a promising tool in these domains especially in the last two years (Kooli, 2023), offering innovative solutions to complex problems. This literature review synthesizes recent advancements documented in academic literature about AI-enabled chatbots in relation to emergencies, focusing on the new information that adds to the existing knowledge base, developed in D3.2.

4.1.1 AI-ENABLED CHATBOTS IN EMERGENCIES

AI-enabled chatbots have shown significant potential in the healthcare domain, particularly in emergency medicine. AI-enabled chatbots have gained a crucial role in patient engagement, clinical support, mental health, health monitoring, patient education, appointment scheduling, symptom checking, chronic disease management, triage, remote monitoring, telemedicine, health coaching, and emergency response (Haque, Chowdhury & Soliman, 2023; Sharma, Kaushal, Kumar & Gainder, 2022).

In addition, several studies show that the COVID-19 pandemic sped up the infiltration of AI-enabled chatbots into the healthcare system, particularly into community health service providers and hospitals. The use of AI-enabled chatbots was able to decentralize many of the activities carried in these institutions, allowing not just self-diagnosis, but also prioritization of requests (e.g., Noble et al., 2022; White, Martin, Dip & White, 2022).

Another example of a recent advancement in AI-enabled chatbots, taken from the tourism industry Abraham & Thazhathethil (2023), highlights how chatbots can store data and assist during disaster management and risk reduction. According to the authors, AI-enabled chatbots can be used as alternative communication tools to allocate resources and help survivors find the emergency services they need. This is particularly relevant in emergency situations, where chatbots can provide travel-related and disaster handling-related information, helping individuals to find information on travel services and emergency conditions.

One example which is related to that issue is dealing with the surge of reports during disasters and the need to prioritize inquiries. A single report of "someone looking suspicious and maybe had a gun somewhere" may not be enough to raise the necessary awareness by a single emergency operator. However, if there were similar reports from multiple citizens in the same area, this could be identified by AI and raise the appropriate attention (Adam et al., 2022; Sharma, Kaushal, Kumar & Gainder, 2022).

The research leading to these results has received funding from Horizon 2020, the European Union's a Programme for Research and Innovation (H2020/2014-2020) under grant agreement n° 882850.

4.1.2 AI-ENABLED CHATBOTS AND THE USER EXPERIENCE

Another significant question which was explored widely since the publication of D3.2, is related to the user experience of individuals from using AI-enabled chatbots, in relation to emergencies and disasters. During the COVID-19 pandemic, AI-enabled chatbots were quickly developed to provide information and answer questions about health issues. However, previous studies, analyzed in D3.2, did not evaluate the user experience of individuals from using these chatbots, but their technological design.

In a systematic review of studies which examined the user experience of AI-enabled chatbots during the COVID-19 pandemic, White, Martin, Dip & White (2022) identified a number of factors impacting adoption and engagement of AI-enabled chatbots, among them content, trust, digital ability and acceptability.

In general, findings shows that the user experience and acceptance of using AI-enabled chatbots increased dramatically in the last two years, with almost 60% of individuals holding positive perceptions and around 85% expressing willingness to use AI-enabled chatbots in the future (Almalki, 2021).

The content of the chatbots was found to be crucial in determining their effectiveness. For instance, the "COVID-19 Preventable" chatbot, developed by the Emergency Operation Center (EOC) and the Department of Disease Control (DDC) in Thailand. It was reported to have answered the users' questions appropriately 64% of the time (Rodsawang et al, 2020). This perceived ability of the chatbot to provide accurate and relevant information was rated as being important. The challenge of managing the amount of content in a fast-moving landscape like the COVID-19 pandemic was also highlighted, with one national government chatbot, developed by the department of public health in San Francisco University, describing the need for partnerships with technical teams to manage the content effectively (Judson et al., 2020).

Another factor identified by White et al., (2022) is trust. Considering trust in the chatbot or in the chatbot provider was commonly reported as being an important factor. For example, two studies by Almalki (2021) and Almalki (2020) found that mass media campaigns, social role model endorsement, or national health authority adoption may help in raising awareness and trust in chatbots. Trust was also linked to the perceived ability of the chatbot to provide accurate information (White et al., 2022). Trust was also an issue of the organization standing behind the chatbot (Dennis et al., 2020).

In addition to those, trust in AI-enabled chatbot is achieved also by strengthening the emotional factor of the chatbot, such as by adding a component of self-disclosure, which positively affected the engagement and perceived warmth, but also other methods of emotional disclosure (Liang, Shi, Oh, Zhang & Yu, 2021).

For example, Jiang et al. (2022) examined an AI-enabled chatbot, called "Replika", based on a human-form digital avatar, to provide emotional support during the COVID-19 pandemic. The study found that "Replika" was able to help relieve overwhelming psychological distress caused by the COVID-19 disruption. The chatbot was perceived by users as having a high level of cognitive empathy and was able to provide empathetic responses. This was particularly beneficial for users who were experiencing a lack of interpersonal support. In terms of emotional disclosure, the study found that "Replika" was able to identify human emotions but had difficulty detecting and interpreting subtle emotional changes. Despite this, users found value in the chatbot's rational and thoughtful responses, which helped them engage in rational thinking and facilitated the development of their resilience. The study also noted that while the chatbot could not share similar experiences to comfort users, its consistent positive inclusivity and energetic support led to users' "willing suspension of disbelief" and admiration for "Replika"'s cognitive empathy.

The research leading to these results has received funding from Horizon 2020, the European Union's a Programme for Research and Innovation (H2020/2014-2020) under grant agreement n° 882850.

Digital ability or perceived IT skills were not reported to have a large effect on chatbot use, as White et al. (2022) identified. Studies by Almalki (2020) and Dennis et al. (2020) found that acceptability was higher for those who already used the internet to seek health information. Interestingly, when the ability of chatbots was perceived to be the same as humans, chatbots were perceived more positively, suggesting that users felt more comfortable with chatbots when discussing socially challenging information.

Last, the theme of acceptability included willingness to use COVID-19 chatbots, perceptions on chatbots in general, and adoption (White et al., 2022). Despite many respondents having little experience with chatbots, most (82.5%) reported being willing to use chatbots to seek general information about health care services and how to prevent the spread of COVID-19 (Almalki, 2021). For example, the "COVID-19 Preventable" chatbot reported that most people (98%) stated that they were likely to continue using the chatbot and 96% said that they would recommend it to others (Rodsawang et al, 2020). Factors such as empathy, emotional sensitivity, and social norms were found to impact chatbot acceptability (Almalki, 2021; White et al., 2022).

4.1.3 AI-ENABLED CHATBOTS DESIGN RECOMMENDATIONS

In the last two years, a part of the technological advancement related to AI-enabled chatbots, recent studies also identified updates recommendations regarding the design and architecture of AI-enabled chatbots, in relation to emergencies and disasters. Chagas et al. (2021) elaborate on the chatbot's architecture, data management, user interface, multilingual support, security and privacy, scalability, integration with other systems and training ana testing.

Regarding the chatbot architecture, Chagas et al. (2021) highlight that the chatbot should be designed to operate in a distributed manner, with multiple instances running in parallel to ensure high availability and reliability. This is particularly important in emergency situations where a high volume of requests is expected. In addition, the chatbot should be integrated with a robust data management system that can handle large volumes of data and provide real-time updates. This includes integration with emergency services databases and other relevant data sources.

The authors also refer to the user interface, that should be intuitive and easy to use, even for individuals who may be under stress due to the emergency situation. The chatbot should provide clear instructions and responses, and avoid technical jargon, with the last one also referring to multilingual support, given the global nature of many emergencies and disasters.

Moreover, Chagas et al. (2021) summarizes additional factors:

- **Security and Privacy:** The chatbot should be designed with strong security measures to protect sensitive user data. This includes encryption of data at rest and in transit, as well as compliance with relevant data protection regulations.
- **Scalability:** The chatbot should be designed to scale up quickly in response to increased demand during emergency situations. This includes the ability to add more resources (e.g., servers, bandwidth) as needed.
- **Integration with Other Systems:** The chatbot should be designed to integrate seamlessly with other systems, such as emergency response systems, social media platforms, and communication tools. This can help to provide a more comprehensive response to emergencies.
- **Training and Testing:** The chatbot should be regularly trained and tested to ensure it can handle a wide range of emergency scenarios. This includes testing the chatbot's ability to handle high volumes of requests and provide accurate responses.

The research leading to these results has received funding from Horizon 2020, the European Union's a Programme for Research and Innovation (H2020/2014-2020) under grant agreement n° 882850.

Stieglitz et al. (2022) adds six additional factors, based on studies from the last two years:

- **Step-by-Step Implementation:** The authors recommend a gradual approach to implementing chatbots in crisis communication. Starting with a social media chatbot, then gradually connecting the systems of the Emergency Management Agencies (EMAs), and finally implementing within an EMA app might be a good starting point. Subsequently, other smart-home applications such as dissemination via smart speakers (e.g., Amazon Alexa) are also conceivable to reach as many people as possible in crisis situations.
- **Adaptation to Specific Context:** The EMAs should compare their requirements with the requirements identified for the Australian and German EMAs that were focused on in the study, and then determine whether the identified design principles may need to be modified.
- **Proactive Interaction:** The emergency chatbot of the future should not only respond to user-initiated conversations but also purposefully initiate conversations on its own. For example, it may approach social media users who have posted relevant content and ask them for more background information before passing this information to its owners, or it may approach social media users in a specific geographic area with relevant information or requests for information.
- **Situation Awareness:** The emergency chatbot of the future is always fully aware of the current situation by frequent updates of information sources such as databases or systems. One challenge the chatbot needs to face is the assimilation of emerging technologies and online communication channels. The chatbot needs to understand this context. The situational awareness is likely to be enhanced real-time by the questions received by the users, validated and learnt by the machine.
- **Avoiding Outdated Advice:** Depending on the nature of the situation, a chatbot that is giving advice which is outdated, even if only by half an hour, may be more harmful than one that is not giving advice at all.
- **Knowledge Transfer:** A rigorous knowledge transfer is mandatory between managers, disaster relief workers, and developers to further improve collaboration resting upon shared experiences.

4.2 ChatGPT and its relevancy to this project

The introduction of OpenAI's GPT-3.5, further later GPT-4 and ChatGPT has marked a significant shift in the field of AI-enabled chatbots, particularly in the context of emergencies and disasters. Therefore, we review the scientific contribution of Open AI's GPT-4 and ChatGPT, in the scientific literature and in general, in a distinctive section.

As a cutting-edge language model, GPT 4 has the ability to generate human-like text, understand context, and respond to prompts in a way that is coherent and contextually relevant. This has opened up new possibilities for the use of chatbots in emergency situations, where rapid, accurate, and contextually appropriate responses are crucial.

GPT 4's ability to understand and generate language in a nuanced and contextually relevant manner has made it a valuable tool for building societal resilience. In emergency situations, GPT-4 powered chatbots can provide timely and accurate information, guide individuals through crisis situations, and even help in dispelling misinformation. Furthermore, these chatbots can be used to monitor social media and other communication channels, identifying emerging crises, and helping authorities to respond more effectively.

The research leading to these results has received funding from Horizon 2020, the European Union's a Programme for Research and Innovation (H2020/2014-2020) under grant agreement n° 882850.

The role of AI-enabled chatbots, particularly ChatGPT, in emergency and disaster scenarios is significant and continues to evolve. These chatbots are increasingly being used to communicate with the public and foster societal resilience. In querying for the updated scientific literature regarding the integration of OpenAI's ChatGPT into the development and use of AI-enabled chatbots in relation to emergencies and disasters, in building societal resilience, several articles and cases studies were found.

In a study by Sarbay, Berikol & Özturan (2023), the performance of ChatGPT in emergency medicine triage prediction was discussed. In a preliminary, cross-sectional study, conducted with case scenarios generated by the authors, emergency cases were consulted both with emergency medicine experts and ChatGPT. The study found that ChatGPT showed a high performance when predicting high acuity cases, with overall sensitivity of 57.1%, with only 18% under triaged and 22% over triaged. Indicating the potential usefulness of ChatGPT in determining cases requiring critical care.

Webb (2023) presented a novel use of ChatGPT in the field of emergency medicine, specifically for training purposes. The study focused on the context of breaking bad news to patients, a challenging yet essential skill for practicing physicians. A detailed input prompt was designed to outline rules of play and grading assessment via a standardized scale. User inputs (physician role), chatbot outputs (patient role), and ChatGPT-generated feedback were recorded. To ensure the chatbot's proficiency in this context, it was trained on a diverse dataset of patient-physician interactions, including real-world scenarios of breaking bad news, expert guidelines, and best practices in emergency medicine communication. This comprehensive training enabled the chatbot to simulate realistic patient responses and provide constructive feedback. The results showed that ChatGPT was able to set up a realistic training scenario on breaking bad news based on the initial prompt. Active roleplay as a patient in an emergency department setting was accomplished, and clear feedback was provided to the user through the application of the SPIKES framework for breaking bad news. The study concluded that the novel use of AI chatbot technology to assist educators in training scenarios is abundant with potential. The author concluded that by facilitating realistic roleplay scenarios and providing effective feedback, ChatGPT can help improve the communication skills of emergency physicians during their training sessions, thereby enhancing their ability to deliver critical news to patients in emergency situations. This, in turn, can contribute to building societal resilience by improving the quality of patient-physician communication during emergencies and disasters.

Ayers et al. (2023) demonstrated the potential of ChatGPT in providing quality and empathetic responses to patient questions. The study used a public and nonidentifiable database of questions from a public social media forum and found that evaluators preferred chatbot responses to physician responses in 78.6% of the evaluations. Notably, chatbot responses were rated of significantly higher quality and more empathetic than physician responses. This suggests that AI chatbots like ChatGPT could potentially aid in drafting responses to patient queries, thereby reducing the workload of healthcare professionals and potentially lowering clinician burnout.

Similarly, Altamini et al. (2023) investigated the potential of ChatGPT in providing accurate and timely information regarding snakebite prevention, recognition, and management. The researchers posed nine hypothetical questions based on comprehensive snakebite management guidelines to ChatGPT, and the responses were evaluated by clinical toxicologists and emergency medicine physicians. The study found that ChatGPT provided accurate and informative responses related to the immediate management of snakebites, the urgency of seeking medical attention, symptoms, health issues following venomous snakebites, the role of antivenom, misconceptions about snakebites, recovery, pain management, and prevention strategies. However, some limitations were identified, including outdated knowledge, lack of personalization, and inability to consider regional variations and individual characteristics. Despite these limitations, the study concluded that ChatGPT can serve as a valuable resource for preliminary information, education, and triage support in remote or underserved areas, complementing professional medical consultation and enhancing patient education.

Last, Tülübaş et al. (2023) explored the potential of ChatGPT in supporting or generating scientific research. The researchers used ChatGPT to conduct a simultaneous query on five basic themes related to emergency remote teaching (ERT), a topic that gained significant interest during the COVID-19 pandemic. The themes included the definition and emergence of ERT, its appropriateness for different grade levels, a comparison between ERT and online teaching, the possible outcomes of ERT, and the future prospects and uses of ERT. The study found that both versions of ChatGPT were capable of generating accurate information without significant bias, although the responses lacked depth and insight and were somewhat repetitive. As the level of judgment required by the query increased, the performance of ChatGPT-4 was much better, providing clearer and more concise responses with more synthesized and detailed categories of information on ERT. The study concludes that while ChatGPT holds promise in scientific research, human intervention is still necessary to ensure accurate and reliable output from AI-based scientific queries.

However, while ChatGPT's capabilities in the realm of emergencies and disasters are undeniably groundbreaking, it's essential to address some of its documented shortcomings. One of the most discussed issues is the phenomenon of "hallucination," where the model generates information that, while sounding plausible, is not accurate or based on factual data. This can be particularly concerning in emergency situations where misinformation can have dire consequences. For instance, in a rapidly evolving crisis, providing incorrect or misleading information can exacerbate panic or lead to inappropriate responses from the public or authorities (Paul, 2023).

Another concern relates to ChatGPT's accuracy. In an article in "Popular Science", Paul (2023) highlights instances where the model, when tested against human fact-checkers, occasionally provided incorrect answers or failed to flag false information. In the context of emergencies and disasters, such inaccuracies can be detrimental. For instance, if a user seeks information on safety protocols during a natural disaster, an inaccurate response could jeopardize lives. While the model's human-like text generation is commendable, it also means that it can confidently generate incorrect or misleading information, making it challenging for users to discern its reliability.

Furthermore, ChatGPT's reliance on vast amounts of internet text for its training means it can sometimes reproduce biases present in those texts. In emergency scenarios, where clear, unbiased communication is paramount, such biases can skew perceptions and responses. It's crucial for developers and users alike to be aware of these limitations and approach ChatGPT's outputs with a degree of skepticism. Continuous monitoring, validation, and collaboration with human experts are essential to ensure that the information provided is both accurate and contextually relevant, especially in high-stakes situations like emergencies and disasters.

In conclusion, the contribution of ChatGPT to the development of AI-enabled chatbots for emergencies and disasters is significant and continues to evolve. By facilitating communication with the public and fostering societal resilience, ChatGPT is revolutionizing the way we respond to emergencies and disasters. However, it is essential to acknowledge that while ChatGPT holds promise in scientific research and emergency response, there are documented challenges. One of the most discussed issues is the phenomenon of "hallucination," where the model might generate plausible-sounding but inaccurate information. This can be particularly concerning in emergency situations where misinformation can have dire consequences. Furthermore, as highlighted by a recent article from Popular Science, there are concerns about ChatGPT's occasional inaccuracies when compared to human fact-checkers. Such inaccuracies, especially in the context of emergencies, can be detrimental. Human intervention remains crucial to ensure accurate and reliable output from AI-based queries. Future research should focus on addressing these limitations and further exploring the potential of ChatGPT in building societal resilience.

In the context of AI-enabled chatbots for emergencies and disasters, the integration of ChatGPT, including GPT-4 and InstructGPT, represents a paradigm shift that consolidates several distinct components of traditional chatbot blueprints. Unlike conventional systems that require separate modules for technological capabilities, cognition and intelligence, and processing, preparing, and

The research leading to these results has received funding from Horizon 2020, the European Union's a Programme for Research and Innovation (H2020/2014-2020) under grant agreement n° 882850.

training, ChatGPT encapsulates all these functionalities within a unified framework. Through its API, ChatGPT can seamlessly integrate with existing chatbot systems, providing a comprehensive solution that leverages state-of-the-art natural language processing techniques. This integration not only simplifies the development process but also enhances the chatbot's ability to understand, generate, and respond to complex queries in a contextually relevant manner. However, the model's vast training on internet text means it can sometimes reproduce biases or provide overly broad answers, lacking the specificity required in critical situations. By incorporating human-like text generation, contextual understanding, and prompt responsiveness, ChatGPT offers a robust and efficient solution that can significantly contribute to societal resilience in emergency situations. Still, it's vital to recognize that while ChatGPT provides a powerful tool, the collaboration with human expertise remains essential to tailor the system to specific needs and ensure the accuracy and reliability of the responses in various emergency and disaster scenarios.

4.3 New AI and ML Solutions

Apart of AI-enabled chatbots, several other cutting-edge social media and AI solutions were identified, for improving the communication process between authorities and first responders to the public and vice-versa, and within the public. All in relation to emergencies and disasters, before, during and after, in building societal resilience. We focus here mainly on social media listening (SML), virtual reality (VR) technologies and augmented reality (AR) technologies. Here, we review the recent literature about these technologies, while the solutions themselves are reviewed in a later section.

4.3.1 Social Media Listening

Social media listening, also known as social media monitoring, is the process of identifying and assessing what is being said about a company, individual, product, or brand on the internet. It involves tracking various social media platforms and other digital channels to gather information and insights. This process is used in various areas such as marketing, customer service, public relations, and market research (Rao, 2016).

In the context of emergencies and disasters, social media listening can play a crucial role in communication with the public. It can provide real-time updates, disseminate crucial information, and serve as a platform for two-way communication between authorities and the public. It can also facilitate communication among community members, fostering a sense of unity and cooperation (Krüger & Albris, 2020).

Social media listening has been used in various areas related to emergencies and disasters. For instance, it has been used for identifying health concerns, where authorities can monitor social media platforms for mentions of specific symptoms or diseases, allowing for early detection and response (Chon & Kim, 2022). During the COVID-19 pandemic, social media listening played a significant role in tracking the spread of the virus, understanding public sentiment, and combating misinformation (Miao, Last & Litvak, 2022).

De Rosis, Lopreite, Puliga, & Vainieri (2023) provides an example for how social media data, using social media listening technologies, could help tracking the spread of COVID-19 symptoms by monitoring tweets on Twitter. Using a sentiment index, known as polarity, that the authors applied to four main key stakeholders, namely the general public, journalists, politicians, and doctors. They monitored their emotional change in three different groups of Italian regions during the initial pandemic when lockdown forced people to online communication. They analyzed the sentiments of each element and examined their interplay in shaping the positive or negative perception of the crisis during the early phase of the coronavirus in Italy. Specifically, they looked at how these

The research leading to these results has received funding from Horizon 2020, the European Union's a Programme for Research and Innovation (H2020/2014-2020) under grant agreement n° 882850.

sentiments influenced each other in creating, amplifying, or reducing the overall perception. This study's unique contribution lies in providing valuable insights into how different stakeholders influenced public perception and response to the crisis. The authors suggest that their findings could be used to develop more effective communication strategies for managing future health crises.

Similarly, Zachlod, Samuel, Ochsner, & Werthmüller (2022) discuss the use of social media data for various applications, including tracking COVID-19 symptoms, in relation to predicting human behavior. The authors explore the potential of social media as a tool for monitoring and controlling government policies, which could include public health measures related to COVID-19, based on a systematic review of 94 studies, identified between 2017-2020. The authors show that social media listening is highly practical, and not just a source for academic studies, with 45 out of the 94 reviewed articles examining practical uses of social media listening.

In addition, social media listening is also used in crisis communication. Social media listening is used to disseminate better information to the public, address misinformation and engage with the public during emergencies. Chiou, Voegeli, Wilhelm, Kolis, Brookmeyer, & Prybylski (2022) analyzed an example of COVID-19 State of Vaccine Confidence Insight Reporting System developed by the Centers for Disease Control and Prevention (CDC). This system routinely collects and analyzes data on public questions, concerns, frustrations, and circulating misinformation. The data is then used to produce biweekly reports that provide actionable insights to guide communication content and intervention development. The system uses a mixed-methods approach to synthesize multiple primary and secondary data sources, including social media, news media, search engine queries, polling data, scientific literature, and direct inquiries from the public submitted to the CDC. The identified themes are prioritized using a threat matrix and then classified and color-coded by risk based on reach, dissemination, and potential effect on vaccine confidence and uptake. This approach allows for early detection of acute threats to public safety, such as conspiracy theories regarding treatments, and provides an early warning system for public health action.

Social media listening has also facilitated community interaction during emergencies. It has been used to understand community needs, gather feedback, and foster community resilience. Herrera & Gjøsaeter (2022) argue that social media analytics often provide a generalized picture of situational awareness, which can overlook the needs of underrepresented communities. They emphasize that communities are diverse and segmented, and this diversity is often not captured in social media analytics. Their interviews with professionals using social media as a source of information in public service organizations found that these practitioners often did not perceive this segmentation in social media conversations. Those who were aware of this limitation agreed that there was room for improvement and used alternative mechanisms to understand and provide services to these communities in need. The authors suggest that social media analytics tools should be designed to support inclusive public services for all, including persons with disabilities and other disadvantaged groups. They also highlight the importance of strategizing, testing, and training when using social media analytics tools to improve the results and better understand community needs.

In general, the literature points out that despite the different frequency of analysis and reporting of different mechanisms, social listening activities were very frequent two years after the COVID-19 pandemic declaration. It is also mentioned that the digital tools used for social media listening were mainly from North-based corporates and were not always adjusted to the humanitarian response/sector (Utama, Supangkat & Purwarianti, 2021).

Last, in relation to emergencies and disasters in building societal resilience, social media listening is also used in considering cultural and linguistic diversity. This includes developing culturally appropriate communication strategies and addressing the information needs of linguistically diverse populations. In a study by Lohiniva, Sibenberg, Austero & Skogberg (2022), the authors show how social listening, based on web platforms and not the traditional social media, can be used to enhance access to appropriate pandemic information among culturally diverse populations.

The research leading to these results has received funding from Horizon 2020, the European Union's a Programme for Research and Innovation (H2020/2014-2020) under grant agreement n° 882850.

4.3.2 VR and AR

The use of Virtual Reality (VR) and Augmented Reality (AR) in communication during emergencies and disasters has been a topic of interest in recent research. One of the key areas where these technologies have been applied is in the training of emergency responders and healthcare providers.

In a systematic review of the literature published between 2000-2020, Brown, Margus, Hart, & Sarin (2022), show that VR and AR were found to be well-suited for disaster training as they provide a safe, socially distant simulation with a high degree of realism. The review found that participants reported improved confidence and positive satisfaction after the simulations. The researchers suggest that VR or AR can be considered for disaster training in addition to other, more traditional simulation methods, to improve the communication of emergency organizations with the public. However, they also noted that more research is needed to create a standardized educational model to incorporate VR and AR into disaster medicine training and to understand the relationship between disaster simulation and improved patient care.

In another study by Yoo, Drumwright, & Piscarac (2023), the authors explored the use of AR in nonprofit marketing and community participation during the COVID-19 pandemic. The researchers found that AR, through its unique partial immersion experience, could enhance users' understanding of the plight of COVID-19 victims, which in turn could impact their perception of social capital and positively affect community participation intentions. The study also highlighted the importance of media richness and social presence in AR content. Media richness, which includes interactivity and vividness, was found to prompt user empathy, while social presence, the feeling of being in the presence of others, was found to foster perceptions of increased social capital. The researchers suggested that nonprofits could leverage these aspects of AR to create a richer experience for the audience, thereby increasing empathy and enhancing perceptions of social capital.

In conclusion, AR and VR can be effectively used in communication during emergencies and disasters, not only to improve training and management but also to enhance empathy, social capital, and community participation. These technologies can help build societal resilience by fostering a sense of community and promoting prosocial behaviors.

4.3.3 Other social media innovative technologies

As mentioned before, other innovative technologies are used the improve the communication between authorities and first responders with the public, before, during and after disasters, and to improve the communication process within community members, for building societal resilience. The literature lists several of them as:

- Blockchain can be used to create secure, decentralized communication networks that are resilient to disasters. It can ensure the integrity and availability of critical information during emergencies (Javadpour et al., 2023).
- IoT devices can provide real-time data about environmental conditions, infrastructure status, and more. This data can be used to inform emergency response and recovery efforts (Esposito, Palma, Belli, Sabbatini, & Pierleoni, 2023).
- Drones can be used for real-time surveillance, damage assessment, and even delivery of supplies during emergencies. They can provide valuable visual information to both authorities and the public (Al-Wathinani et al., 2023).
- The high speed and low latency of 5G networks, combined with edge computing, can enable real-time data analysis and decision-making during emergencies. They can support the use of advanced technologies like AR, VR, and IoT in emergency management (Taveras Cruz et al., 2023).

The research leading to these results has received funding from Horizon 2020, the European Union's a Programme for Research and Innovation (H2020/2014-2020) under grant agreement n° 882850.

- Advanced geospatial technologies, including Geographic Information Systems (GIS) and Remote Sensing, can provide critical information about disaster-affected areas, helping in effective planning and response (Chan, Abid, Sulaiman, Nazir, & Azam, 2022).

The research leading to these results has received funding from Horizon 2020, the European Union's a Programme for Research and Innovation (H2020/2014-2020) under grant agreement n° 882850.

5 METHODS

The main goal of this deliverable is to delineate directions for innovative communication and social media solutions including improved design directions for communication and social media technology. In order to carry out this main goal, we revised the results of D3.2. This included conducting an evaluation of the AI-enabled chatbot blueprint using an eDelphi consensus technique and collection of additional social media innovate technologies to improve the communication process between authorities and first responders to the public and within the community members.

5.1 REVISING THE BLUEPRINT FOLLOWING THE VALIDATION EXERCISES AND THE DEVELOPMENT OF CHATGPT

After the submission of D3.2, the AI-enabled chatbot's blueprint was incorporated into the validation exercises conducted as part of the validation plan in D4.2. The blueprint was initially revised based on the comments received during the validation process, which included workshops with end-users. Additional revisions were made in light of the developments of GPT-4, InstructGPT, and ChatGPT, which were published following the submission of D3.2 and the initial development of the blueprint.

5.2 EVALUATING THE BLUEPRINT USING EDELPHI CONSENSUS TECHNIQUE

5.2.1 WHAT IS EDELPHI?

The eDelphi process, also known as the Delphi consensus technique, is a structured communication method originally developed as a systematic, interactive forecasting method which relies on a panel of experts. The experts answer questionnaires in two or more rounds. After each round, a facilitator or change agent provides an anonymized summary of the experts' forecasts from the previous round as well as the reasons they provided for their judgments. Thus, experts are encouraged to revise their earlier answers in light of the replies of other members of their panel. It is believed that during this process the range of the answers will decrease and the group will converge towards the "correct" answer. Finally, the process is stopped after a pre-defined stop criterion (e.g., number of rounds, achievement of consensus, stability of results) and the mean or median scores of the final rounds determine the results (de Meyrick, 2003; Turoff, 1975).

Delphi is based on the principle that forecasts (or decisions) from a structured group of individuals are more accurate than those from unstructured groups. The technique is designed to overcome the disadvantages of conventional face-to-face meetings and the problems of group dynamics. The Delphi technique is widely used in various fields of study to generate forecasts about the occurrence of specific events (Landeta, 2006).

The eDelphi process is an online adaptation of the traditional Delphi method. The use of online tools for the Delphi process allows for a larger number of participants, greater geographic diversity, and the convenience of participation at any time. It also allows for automated data collection and analysis, making the process more efficient. However, like any method, the eDelphi process has its limitations. These include the potential for low response rates, the time-consuming nature of multiple rounds, and the potential for bias if the panel is not representative of the broader expert community (Hayness & Shelton, 2018).

The eDelphi process is particularly useful in situations where individual judgments need to be combined into a group consensus, and where face-to-face meetings are impractical due to geographic or time constraints. It is also useful when there is incomplete knowledge about a problem

The research leading to these results has received funding from Horizon 2020, the European Union's a Programme for Research and Innovation (H2020/2014-2020) under grant agreement n° 882850.

or phenomenon, making it necessary to rely on expert opinions rather than established facts or evidence. The eDelphi process has been used in a wide range of fields, including health care, education, technology forecasting, and policy making.

5.2.2 Recruitment of expert participants

The recruitment of experts for the eDelphi process was a critical step in ensuring the validity and reliability of the consensus process. The recruitment process took place between March and June 2023, and the experts were selected based on their experience and knowledge in the fields of Artificial Intelligence (AI), Machine Learning (ML), and chatbots. These experts were drawn from a variety of sectors, including academia, technology, and industry, and represented a diverse range of geographical locations, including Italy, the Netherlands, Israel, Spain, the USA, and Greece. The distribution of the experts is described in table 1.

The recruitment process began with a comprehensive search for relevant experts on professional platforms such as LinkedIn, academic institute websites, technological hubs, and personal connections. Additionally, members of the ENGAGE consortium were consulted regarding the reference for additional relevant experts. Potential participants were contacted via email, which included an introduction to the goals of the eDelphi process and a link to the questionnaire. If necessary, a reminder email was sent to encourage participation.

The selection criteria for the experts were intentionally broad to capture a wide range of perspectives and experiences. However, all participants were required to have experience in AI, ML, or chatbots, either from an industrial, technological, consulting, or academic perspective. This ensured that the participants had the necessary knowledge and expertise to contribute meaningfully to the consensus process. The diverse backgrounds and experiences of the experts were instrumental in providing a comprehensive and nuanced evaluation of the AI-enabled chatbot.

Table 1. The distribution of experts in the eDelphi process.

#	Country	Role	Years with experience	Proficiency working with blueprints
1	Italy	Consultant	3	Average
2	Italy	Consultant	5	Average
3	Italy	Consultant	2	Average
4	Italy	UX designer	1.5	Average
5	Spain	Researcher	6	Average
6	Netherlands	R&D engineer	2	High
7[1]	Greece	Software engineer	4	Average
8	Israel	Researcher	35	Very High
9	Italy	Researcher	5	High
10	Israel	Consultant	1	Average
11	USA	Software engineer	1	Average
12	Israel	Product manager	6	Average
13	Greece	Software engineer	4	Average
14	Israel	Researcher	3	Average
15	Italy	CTO	10	Very High
16	Netherlands	Software engineer	2.5	High

The table shows that six experts were from Italy, four from Israel, two from the Netherlands, two from Greece, one from Spain and one from the USA. Five of the experts were engineers (one R&D engineer and the rest software engineers), four were technological consultants, and another four were researchers. The additional three were UX designer, product manager and a CTO. The roles of expertise represent a diverse panel of participants in the eDelphi process. In addition, most experts had significant experience of years working with either AI in general or chatbots in particular

[1] The experts are listed in the table by the order they completed the first round. Expert #7 completed the first round only, but not the second one6.

The research leading to these results has received funding from Horizon 2020, the European Union's a Programme for Research and Innovation (H2020/2014-2020) under grant agreement n° 882850.

(M=5.69, SD=8.16). 11 of them defined their proficiency working with blueprints as "average", three as "high" and two as "very high".

5.2.3 The questionnaire

The eDelphi questionnaire was designed to gather comprehensive feedback on the initial blueprint of the AI-enabled chatbot. The questionnaire was divided into four parts, each serving a specific purpose in the consensus process. The full questionnaire is attached in the appendices.

The questionnaire began with a brief introduction that reiterated the goals of the eDelphi process. Participants were then asked to provide their consent, using a consent form, for participation and their email address for the second round of the process.

The first part of the questionnaire presented the initial blueprint of the AI-enabled chatbot, which consisted of 12 different components. Participants were asked to provide general comments or suggestions after reviewing the blueprint.

The second part of the questionnaire delved deeper into each of the 12 components of the blueprint. Participants were asked to rate their agreement on a Likert scale regarding the technological contribution, comprehensiveness, and placement of each component within the blueprint. Those who rated a component as having a low or very low extent in any of these aspects were given the opportunity to provide suggestions for improvement.

The 12 components examined were: input types, channels & platforms, UX and logic, security & privacy, logging and documentation, existing platforms and models for training the chatbot, quality assurance, monitoring and reporting, statistics, alerts, datasets, and manual human monitoring. This was the core part of the eDelphi process where consensus was sought.

The third part of the questionnaire, disseminated only in the first round, asked participants to consider the blueprint as a whole. They were asked to rate the appropriateness of the process and interaction between the different components in the blueprint. They were also asked to consider the relevance of the blueprint in light of recent AI developments, specifically referring to OpenAI's ChatGPT. It was needed for statistical purposes and therefore was not included in the eDelphi.

The fourth and final part of the questionnaire collected demographic information and other statistical data, such as the participant's country of origin, role in the organization, years of experience, and proficiency in designing AI-enabled chatbot blueprints.

5.2.4 Analysis of the questionnaire

The eDelphi process was conducted in two rounds, until reaching a satisfactory rate of consensus. In the first round, consensus was defined as at least 70% of the experts agreeing to a high or very high extent with each of the three questions posed for the 12 components. Questions that reached this threshold were not included in the second round.

Between the two rounds, minor adjustments were made to the blueprint based on the feedback gathered from the open-ended questions. Most of these adjustments were clarifications rather than significant changes. For instance, several experts in the first round commented on the placement of the security and privacy component. In response, we added a clarification stating that the security and privacy component applies to all parts of the chatbot, providing multiple layers of security. This not only covers the chatbot itself but also the other platforms it interacts with, thereby addressing potential issues such as unintentional sharing of information by the user or potential data leaks to the host platform.

The research leading to these results has received funding from Horizon 2020, the European Union's a Programme for Research and Innovation (H2020/2014-2020) under grant agreement n° 882850.

In the second round of the eDelphi process, the questionnaire was streamlined to focus solely on the components that had not met the consensus criteria in the first round. Participants were presented with their original responses, the average responses, and the median response for each question. They were then asked to reconsider their answers in light of this information. The results were then analyzed again to assess whether consensus rates had improved. Components that still did not reach consensus after the second round were noted in the blueprint. The second questionnaire is attached in the appendices.

5.2.5 Analysis of the questionnaire

In the analysis of the questionnaire responses, we employed descriptive statistics. We calculated the average and median of the responses for each question, as well as the percentage of respondents who selected each response option: very low extent, low extent, neutral, high extent, or very high extent.

The decision to use the percentage of agreement with high or very high extent as the consensus criteria is grounded in the methodology of the eDelphi technique (e.g., Diamond et al., 2014). The Delphi method is designed to seek consensus among a group of experts. The use of a percentage agreement threshold is a common approach in Delphi studies to determine when consensus is reached. This threshold can vary depending on the study, but a common benchmark is 70% or 75% (e.g., Kelly, Moher & Clifford, 2016; Price, Rushton, Tyros & Heneghan, 2021).

In our study, we chose to define consensus as a high or very high extent of agreement (i.e., a rating of 4 or 5 on our Likert scale) by at least 70% of our experts. This threshold was chosen to ensure that the consensus represents a substantial majority of our expert panel, thereby increasing the validity of our findings. This approach is consistent with the principles of the Delphi method, which values the collective judgment of the expert panel.

It's important to note that the Delphi method does not require absolute agreement among all panel members. Instead, it seeks to identify areas of general agreement or consensus. By using a percentage agreement threshold, we can identify these areas of consensus while also acknowledging and respecting the diversity of opinions within our expert panel.

5.3 Preparation of a prototype for warning systems based on the blueprint

After the preparation of the initial blueprint of the AI-enabled chatbot, a prototype of a chatbot, focused on warning systems, was designed by One2Many which are part of the consortium. The goal of the prototype was to follow the blueprint we developed, to check its feasibility, in a specific test case of a chatbot for emergencies.

5.3.1 Rationale for a public warning chatbot

According to a report by the Body of European Regulators for Electronic Communications (BEREC, 2020) the minimum operational requirements that are expected from a public warning system, is that the Location Based Short Message Service (LB-SMS) and Cell Broadcast (CB) are considered sufficient from legislator point of view[2]. Yet, the report also lists demand for capability to send the message in the language of the user, allow longer messages and include sufficient information on the threat and how to react on it. This is a possible contradiction as both LB-SMS and CB have

[2] https://www.berec.europa.eu/en/document-categories/berec/reports/berec-report-on-the-outcome-of-the-public-consultation-on-the-draft-berec-guidelines-on-how-to-assess-the-effectiveness-of-public-warning-systems-transmitted-by-different-means , BoR (20) 114

The research leading to these results has received funding from Horizon 2020, the European Union's a Programme for Research and Innovation (H2020/2014-2020) under grant agreement n° 882850.

limitation in the message size, being five times 160 7-bit characters and 15 times 93 in Latin alphabet respectively[3].

The vast majority of chatbots listed in the D3.2 are single "crisis" (hazard/domain) chatbots. For example, chatbots providing information on the coronavirus or earthquake preparation. Given these concerns and limitations, a chatbot prototype was conceived that can provide the public with the following high-level capabilities to overcome the above:

- Support a longer message that may include more detailed instructions, where to get more information and scenario specific details.
- Work suitably on dynamic alert information for numerous crisis scenario's.
- Translation and audio service.
- Lessen the load on human emergency channels such as call centre's a surge of calls to emergency call centre's during emergencies and disasters may collapse the infrastructure (Carenzo, Costantini, Greco, Barra, Rendiniello, Mainetti & Cecconi, 2020; Hrabi, 2020).
- Provide access for the public to the alert information using a summary page and allowing the them to query for specific information, communicating with the emergency services.

The chatbot prototype matches most closely the categorization as a Response messenger bot as classified in D3.2. For convenience the chatbot access would be provided as part of the alert message (using a short URL to conserve message space) through an easy accessible link. It should be noted that the chatbot doesn't replace LB-SMS or CB as a warning channel; it is positioned as a complement.

5.3.2 Initial steps for building a public warning chatbot

Following are the steps taken in preparation of building a chatbot and their mapping to the implementation stages described in D3.2.

Discover the requirements for the chatbot, per extension building a user journey to get an overview of the user interactions that are needed. And creating a feature set which incorporates end-user and stakeholders needs, taking into account the scope, technical and human resources[4]: step #1 framework, #2 channels, #4 security and #6 capabilities); Research availability of suitable data sets (step #5 datasets); Research tools appropriate for building the application. Going forward with a Progressive Web App (PWA), for the prototype. Combining features of a mobile application and a website. And also being able to run the application on both mobile, desktop and mobile environments: step 3 the "brain" and partly step #1 and #2.

The project took an iterative approach with short cycles, building parts of the prototype, getting feedback and expanding upon it. As such as the start of the project, we did not have a good and complete set of requirements, especially areas that were lacking as per blueprint were #6 capabilities, #7 monitoring and #8 processes.

The following table shows the prioritized feature set, on which the initial prototype was built. More information on the requirements can be found in the internship report[5].

[3] https://www.gsma.com/mobilefordevelopment/wp-content/uploads/2013/01/Mobile-Network-Public-Warning-Systems-and-the-Rise-of-Cell-Broadcast.pdf

[4] https://koenvogel.net/wp-content/uploads/2023/07/Internship_Report_Everbridge.pdf

[5] https://koenvogel.net/wp-content/uploads/2023/07/Internship_Report_Everbridge.pdf

The research leading to these results has received funding from Horizon 2020, the European Union's a Programme for Research and Innovation (H2020/2014-2020) under grant agreement n° 882850.

Table 2. Prioritised feature set.

Priority	Feature	Description	Included in prototype
1 (Required)	**Automated text chat**	Basic Machine Learning chatbot functionality, and the basis for the project	yes
2 (Required)	**Public Warning Integration**	Attaching the product to the existing public warning dissemination technologies, e.g. through a URL, to connect users with the application	yes
3	**Overview page**	An addition to chatbot functionality, where the most important emergency information and metadata is displayed without needing to explicitly ask for it	yes
4	**CAP (Common Alerting Protocol) support**	Adherence to the CAP standard, for importing and displaying universal emergency metadata	yes
5	**Multi language support**	Translating the content and conversation in multiple language	yes
6	**Embedded multimedia**	Rich content, like embedded interactive maps, video, audio, etc	yes
7	**Voice Input**	Speech-to-Text (STT) functionality for asking questions to the chatbot	no
8	**Analytics**	System for logging user behaviour	no
9	**Narration**	Text-to-Speech (TTS) functionality for reading out incoming messages and the overview page	no

The research leading to these results has received funding from Horizon 2020, the European Union's a Programme for Research and Innovation (H2020/2014-2020) under grant agreement n° 882850.

10	**Visual Accessibility**	Colour-blind and Visual impairments display options to increase accessibility	no
11	**Showcase website**	Static website showcasing the chatbot features, include (technical) documentation & reporting, and possibly embed an interactive prototype	yes

Datasets with information about commonly asked questions from the public on crisis situations were not available in the consortium. The data may be available in emergency call centres. However such information is not provided publicly. More readable available is information on what to do for common crisis situations provided by emergency authorities, could be obtained for example from the FEMA website (https://fema.gov/) or datasets with such information from Kaggle (https://www.kaggle.com).

A chatbot could be perceived as an interactive website, providing not only a summary page and answers questions for a particular alert, but also providing dynamic information. For developing the chatbot we used the open-source Flutter[6] framework by Google which is capable of building a website with dynamic information. For providing the function to understand questions from the public, a Natural Language Processing (NLP) solution is needed for which Google Dialogflow was selected[7].

For providing dynamic hazard information from a Public Warning System (PWS) to the chatbot we used the OASIS Common Alerting Protocol (CAP)[8]. The chatbot prototype takes the CAP message as metadata, to provide answers to the users. This protocol defines the alert message fields such as description, location, timing and severity.

5.4 Revising the blueprint

A crucial step following the eDelphi consensus technique and prototype testing was the revision of the AI-enabled chatbot blueprint. This iterative process is fundamental to the Delphi method and is designed to refine and improve the blueprint based on the collective expertise of the panel members and the empirical evidence gathered during testing.

The revision process was guided by the feedback and suggestions received from the expert panel during the eDelphi rounds, as well as the insights gained from the prototype testing. Each comment, suggestion, and piece of feedback was carefully considered and evaluated for its potential to enhance the blueprint. This process was not merely about making changes, but about making informed, evidence-based improvements that would enhance the functionality, usability, and effectiveness of the AI-enabled chatbot. For example, in D3.2, three different components were developed for technological capabilities, cognition & intelligence, and preparing, processing and training the chatbot. However, the development of Open AI's ChatGPT and the input gathered from the prototype, led to the change of integrating these three components into one – focusing on "existing platforms", incorporating all three capabilities within.

[6] https://flutter.dev/
[7] https://cloud.google.com/dialogflow
[8] http://docs.oasis-open.org/emergency/cap/v1.2/CAP-v1.2.html.

The research leading to these results has received funding from Horizon 2020, the European Union's a Programme for Research and Innovation (H2020/2014-2020) under grant agreement n° 882850.

The rationale behind each revision was rooted in the data collected during the eDelphi rounds and prototype testing. The goal was to ensure that each change was not only supported by the data but also aligned with the overall objectives of the AI-enabled chatbot. This rigorous approach to revision helped to ensure that the final blueprint was not only theoretically sound but also practically applicable and effective in real-world settings.

The revisions made to the blueprint were not seen as final or definitive, but rather as part of an ongoing process of refinement and improvement. The nature of technological innovation and the rapidly evolving field of AI mean that the blueprint will need to be continually updated and revised to stay relevant and effective. This iterative process of revision, therefore, not only improved the current version of the blueprint but also laid the groundwork for future updates and enhancements.

5.5 Collection of new solutions in the fields of SML, VR and AR

Another objective of this deliverable, apart from the revision of the AI-enabled blueprint, was to extend the list of innovative communication and social media solutions to improve communication with the public and within community members. In order to collect new solutions in the field of social media listening, virtual reality, augmented reality, and also, additional and newer AI-enabled chatbots, we followed several collections steps. One step we took is conducting a systematic literature review and an additional review of the relevant gray literature, which was published from 2021, after the submission of D3.2. Second step was collection of solutions through members of the consortium and reference to other experts related to them. In the third step we used common search engines such as Google and Bing to look for relevant solutions.

5.5.1 Systematic review of the academic literature

The systematic literature review was conducted using the several databases to explore the integration of innovative technologies such as chatbots, virtual reality (VR), augmented reality (AR), and social media listening/monitoring in disaster or emergency management: Ebsco, Google Scholar, JSTOR, Medline, Proquest Central, Pubmed, ScienceDirect, Scopus, Sociological Abstracts and Web of Science. The review was carried out in three distinct sets of queries, each focusing on different aspects of technology application in emergencies.

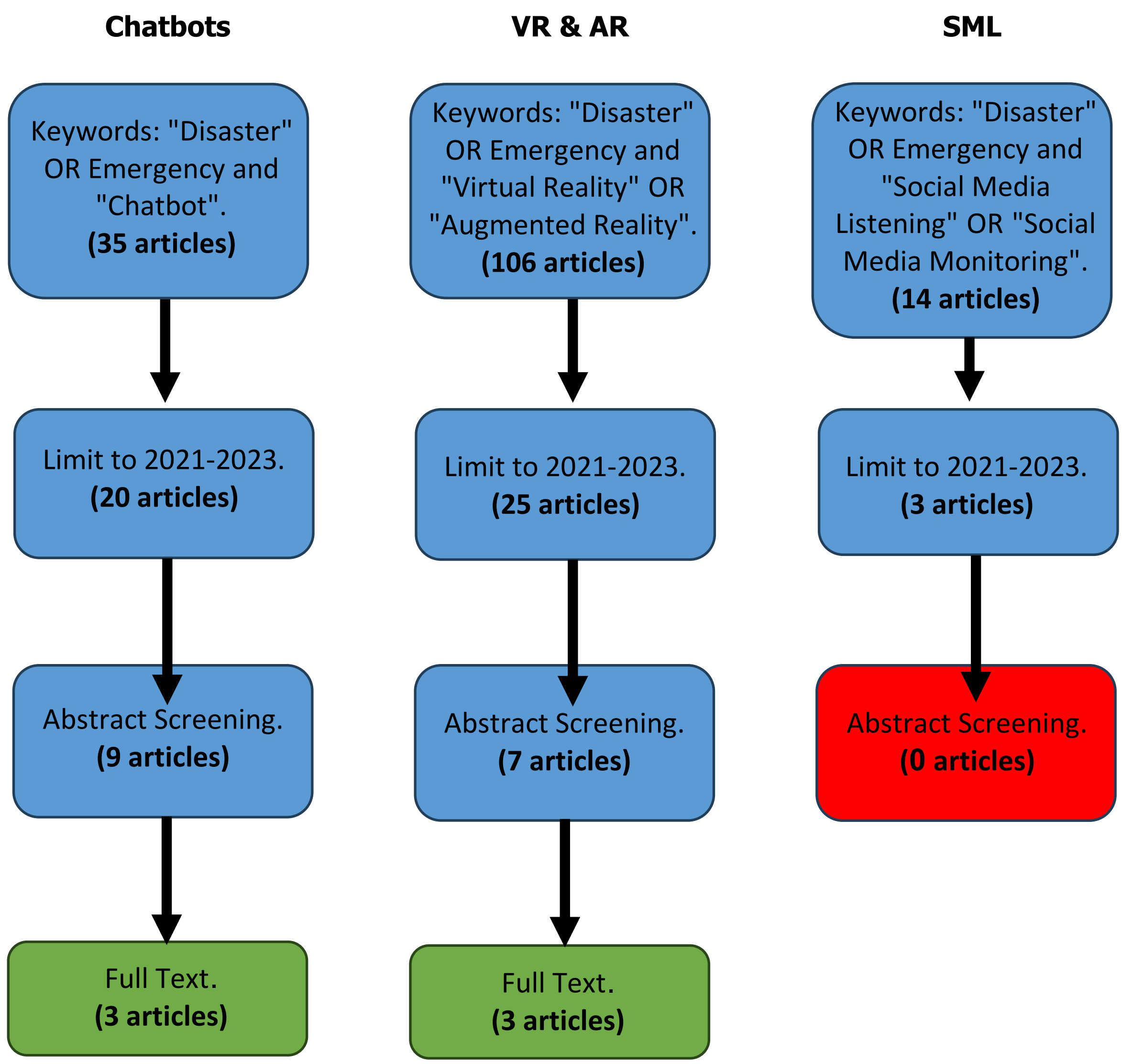
The first set of queries was centered around the following keywords: "disaster" or "emergency" (singular and plural) and "chatbot" (singular and plural). The initial search yielded 35 articles, but when limited to the years 2021-2023, only 20 articles remained. After screening the abstracts, 9 articles were left and finally, three articles remained after the full-text screening.

The second set of queries was centered around these keywords: "disaster" or "emergency" (singular and plural) and "social media listening" or "social media monitoring". The initial search yielded 14 articles, but when limited to the years 2021-2023, only 3 articles remained. Unfortunately, after abstract screening, no relevant articles were found that met the criteria.

The second set of queries explored the relationship between: "disaster" or "emergency" (including plural forms) and "virtual reality" or "augmented reality", in conjunction with "communication" or "communicate". The initial search resulted in 106 articles. By limiting the search to the years 2021-2023, the list was reduced to 25 articles. After abstract screening, 7 articles remained, and following a full-text review, only 3 articles were found to be relevant, one of them is a systematic review of the literature with older articles, but with virtual and augmented reality solutions which were not identified in D3.2.

The research leading to these results has received funding from Horizon 2020, the European Union's a Programme for Research and Innovation (H2020/2014-2020) under grant agreement n° 882850.

Figure 3. Exclusion process of the literature.



5.5.2 Systematic review of the academic literature

We used Google/Bing search engines for the grey literature, with the similar keywords we identified in the previous subsection.

5.5.3 Snowball, ENGAGE partners and KI-CoP organisations

After the literature search, we expanded the list of solutions from ENGAGE partners, Ki-CoP organisations, and other relevant sources. Those were collected through personal emails, meetings, and personal references to other informants. In general, we collected 29 additional solutions as elaborated in the results section. Each solution in the table is analyzed according to its type, and reference criteria of treating false information and management of communication during periods of high information demand and surge of calls.

6 Results of the eDelphi

The eDelphi consensus procedure was carried, as mentioned, between March and June 2023, with the second round taking place in July 2023. The following sections describes the changes that were made in the chatbot before the first round, in light of the validation activities, the emergence of ChatGPT, and the design process of the prototype; the results of the two rounds; areas of consensus and disagreement; and the conclusions.

6.1 Changes made before the first round

Several slight changes were made in the blueprint before passing it to the first round of experts. These changes were a result of the validation process, the emergence of ChatGPT and the discussion around the design of the prototype.

6.1.1 Lessons from the validation activities

The blueprint was validated in a societal resilience hub, in a Ki-CoP workshop which was held on June 14th, 2021.

The validation exercises conducted for the AI-enabled chatbot's blueprint revealed both potential benefits and concerns. End-users recognized the value of the chatbot in managing large volumes of calls to information centres, particularly before and during adversities. This acknowledgment supports the blueprint's direction in enhancing communication channels through AI to build societal resilience and augment the roles of first responders. However, the feedback also highlighted areas for reconsideration and improvement. Concerns were raised about public trust in the information provided by the chatbot, preferences for human interaction over a bot, and the integrity of the information relayed, especially during stressful emergency events. These reservations indicate a need to refine the blueprint to address issues of trust, authenticity, and sensitivity in information delivery. Ensuring that the chatbot can provide reliable and empathetic responses, possibly through integration with human oversight, may be essential in overcoming these challenges.

The lessons learned from the validation process emphasize the importance of balancing technological innovation with human-centric considerations. While the blueprint's focus on AI-enabled solutions offers promising avenues for enhancing societal resilience, careful attention must be paid to the human factors that influence the acceptance and effectiveness of these solutions. The feedback from end-users provides valuable insights that can guide the further development and refinement of the blueprint, ensuring that it aligns with real-world needs and concerns while leveraging the capabilities of AI and machine learning.

6.1.2 Lessons learnt from the development of ChatGPT

Immediately after the launch of ChatGPT, based on GPT 3.5, in November 2022, and later on GPT 4, several concerns were raised regarding the blueprint and its relevancy. The introduction of ChatGPT, with its advanced capabilities in generating human-like text and understanding context, posed questions about the existing blueprint's alignment with this cutting-edge technology. Concerns were centered around the potential redundancy of certain components in the blueprint, given that ChatGPT integrates several technological capabilities, cognition, intelligence, and processing as an external source through its API. There were also apprehensions about the blueprint's adaptability to the rapidly evolving landscape of AI and machine learning, particularly in the context of emergencies and disasters. The need to reassess the blueprint's design, implementation strategy, and integration

The research leading to these results has received funding from Horizon 2020, the European Union's a Programme for Research and Innovation (H2020/2014-2020) under grant agreement n° 882850.

with ChatGPT became evident, ensuring that it leverages the full potential of this new technology while remaining aligned with the project's goals and the real-world needs of societal resilience.

The development of ChatGPT, while revolutionizing certain aspects of AI-enabled chatbots, does not negate the importance of other essential components outlined in the blueprint. Input types, connecting the chatbot to various communication channels, defining the chatbot's logic and user experience (UX), ensuring security and privacy, logging and documentation, integration with existing platforms, quality assurance (QA), monitoring and reporting, handling datasets, statistics, alerts, and manual human monitoring are all integral to the overall functionality and effectiveness of a chatbot system. ChatGPT primarily enhances the language processing and contextual understanding capabilities of the chatbot, but these other components are vital for the chatbot's seamless integration into existing communication infrastructures, user interaction, data protection, continuous improvement, and real-time responsiveness. They ensure that the chatbot not only communicates effectively but also operates securely, efficiently, and in alignment with organizational goals and user needs.

The lessons learned from the integration of ChatGPT with the existing blueprint components underscore the importance of a holistic approach to chatbot development. While advancements in natural language processing (NLP) and AI technologies like ChatGPT can significantly enhance a chatbot's capabilities, they must be harmoniously integrated with other essential elements that govern the chatbot's operation. The experience has shown that focusing solely on technological advancements without considering the broader ecosystem of functionalities can lead to an imbalanced system that may not fully meet the needs of end-users or align with organizational objectives. The integration of ChatGPT with the existing blueprint has highlighted the necessity of continuous evaluation, adaptation, and alignment of all components, ensuring that technological innovations are leveraged to their fullest potential without compromising other critical aspects of the system. It emphasizes the need for a comprehensive and adaptable blueprint that can evolve with the rapidly changing landscape of AI and machine learning, maintaining relevance and effectiveness in various applications and contexts.

This was also supported by the evaluation of the experts. We asked them the following question: "In November 2022, OpenAI launched ChatGPT (Generative Pre-trained Transformer), based on large language models, with both supervised and reinforcement learning techniques. How relevant (7) or irrelevant (1) do you think that the last blueprint, that we presented you, is in light of the release of ChatGPT?" The experts believed that despite the advancement of ChatGPT, the blueprint was still relevant, with a median of 6 out of 7 (M=5.88, SD=1.025). None of them believed that the blueprint is irrelevant. 2 (12.5%) said it was neither irrelevant nor relevant, 3 (18.8%) that it is somewhat relevant, 6 (37.5%) that it is relevant and 5 (31.3%) that it is highly relevant (68.8% relevant or highly relevant).

Those who rated the blueprint to be either relevant or irrelevant were asked whether they could elaborate on their answer. Since all of them rated the blueprint at least as neither irrelevant or relevant, only explanation for the relevancy of the blueprint were recorded:

- *There are already plenty of new services leveraging on chatGPT for different purposes. It is relevant to consider how chatGPT can be used and included in new possible services and how to use it in a better ethical, safe way.*
- *I answer this question also in light of the decision of the Italian Data Protection Authority that formally banned ChatGPT within its jurisdiction. I am skeptical about some of the arguments proposed by the Authority, and from an operative point of view, I think that this kind of AI-powered solutions can greatly benefit individuals and society as a whole. However, the complete unsupervised automation that characterized this solution - and the endangerment this represents for human autonomy - is one the most controversial issues that EU and non-EU authorities (e.g. US FTC, BEUC) are facing approaching these technologies. This is the reason why I think the role of humans (intended both as the*

The research leading to these results has received funding from Horizon 2020, the European Union's a Programme for Research and Innovation (H2020/2014-2020) under grant agreement n° 882850.

autonomy of users and oversight by operators in training, operationalisation and recovery) in the functioning and supervision of these systems should be considered in a more holistic and comprehensive way over the development of the blueprint framework.

- *Why it is not? You can simply benefit from it and add your own features.*
- *Any chatbot system should be using ChatGPT as an optional text generator.*
- *since you can leave the blueprint and take chatgpt as one of the components there, or alternatively use a few llm's (language models – N.S.) and make your platform more robust.*

6.1.3 Lessons learnt from the Prototype

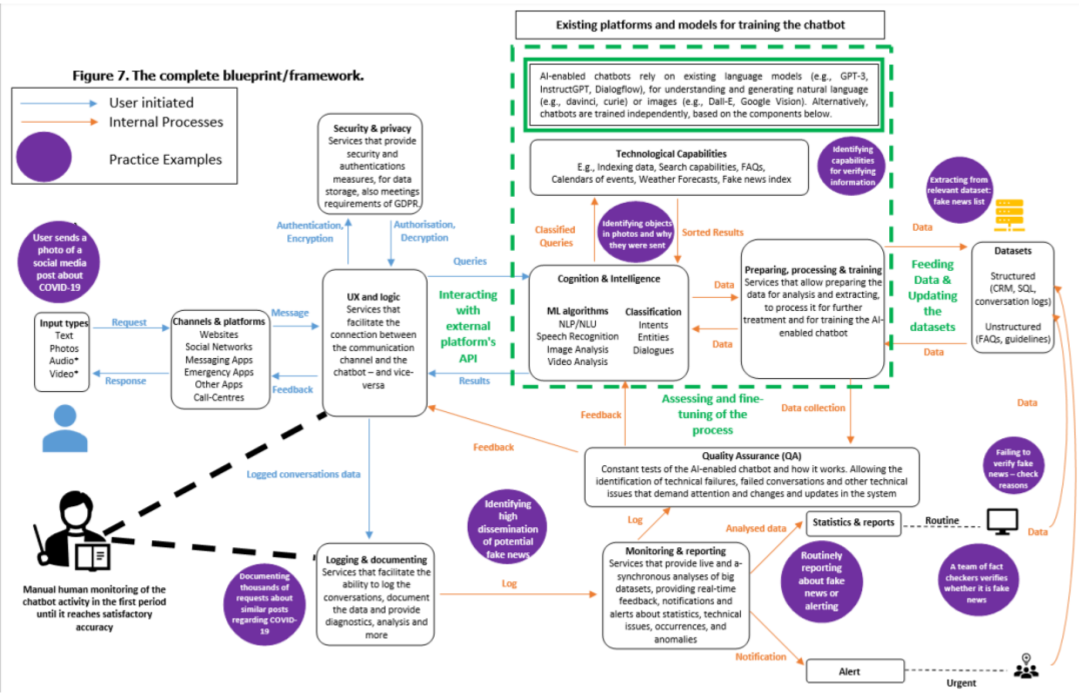
The prototype served as the first attempt to design a chatbot based on the initial blueprint. As will be elaborated in a later section, two options were tested. The first was based on Google's Dialogflow, which identified intents and constructed predefined dialogues but required datasets to train the chatbot. The second involved integration with InstructGPT, leveraging the capabilities of the GPT-3.5 and 4 language models.

The comparison between these two methods reinforced the initial assumption that the area in the blueprint most affected by the emergence of ChatGPT's technology revolves around cognition & intelligence, technological capabilities, and preparing, processing, and training. These could still be considered as three distinct components, but they can also be integrated into one component labeled as "existing platforms."

6.1.4 Revision of the blueprint before the first round of the eDelphi

Following the three processes which were mentioned above, we made the first slight changes in the blueprint, as described in figure 4, and was presented to the experts.

Figure 4. First revision of the blueprint.



The figure shows that the area of cognition & intelligence, technological capabilities and preparing, processing & training, was grouped together in the blueprint under the title of "existing platforms and models for training the chatbot". The description, which was described to the experts, was the following:

"The component of existing platforms and models for training the chatbot refers to three processes: cognition and intelligence, technological capabilities and preparing, processing & training.

AI-enabled chatbots rely on existing language models (e.g., GPT-3, InstructGPT, Dialogflow), for understanding and generating natural language (e.g., davinci, curie) or images (e.g., Dall-E, Google Vision). This component sets the guidelines on how the chatbot can interact with such platforms' APIs in order to be able to use these models. It allows the connection of the existing knowledge (e.g., datasets) of the chatbot with external capabilities (e.g., the language models).

Alternatively, chatbots can be trained independently, without the connection to external platforms, in a self-developed process, based on the three components that such platforms, as abovementioned, are using:

Cognition and Intelligence

The component of cognition and intelligence refers to two sub-categories. One is the relevant algorithms that the chatbots employ, such as NLP, STT, TTS, speech recognition, image and video analysis and more. The decision about the relevant algorithms results from other chatbot components (e.g., input types). In addition, cognition and intelligence also refer to questions of classifications, such as what intents should be covered in the chatbot. Who are the relevant entities? What is the structure of dialogues that the chatbot is expected to conduct? And more. Here, there are relevant issues to consider regarding the training, development and how the AI learns and improves. Questions such as is the AI self-developed or using public cloud AI should be answered.

Technological capabilities

The technological capabilities support the chatbot's functionality, and they depend on its goals. They can include capabilities such as indexing data, searchability, generating automatic or manual FAQs, creating a calendar of events and more.

Preparing, processing and training

Since raw data is mainly unstructured, services are needed to allow the AI-enabled chatbot to train and develop. Training and development of the chatbot are relevant for the backend improvement of the system. They include considerations regarding the chatbot's relevant data and analysis and extraction of information. This component highlights the importance of integrating these services in the design process of the chatbot".

In brief, this component of the blueprint was first introduced to the experts through its integration capabilities with other innovative and cutting-edge platforms and models, such as Google's Dialogflow and OpenAI's ChatGPT. However, the option to dismantle this component into its original three parts was still presented as a possibility.

The research leading to these results has received funding from Horizon 2020, the European Union's a Programme for Research and Innovation (H2020/2014-2020) under grant agreement n° 882850.

6.2 First round results

6.2.1 General evaluation of the blueprint by the experts

Before they were asked about the 12 components of the blueprints, the experts were asked about their general evaluation of the blueprint. In addition, they were asked for one concluding questions at the end. Those two questions were not repeated in the second round. In the first questions, the experts were asked, after viewing the general structure of the blueprint, if they want to add something. Three of them chose yes (18.8%), while the other 13 marked no (81.3%). Since additions were permitted in subsequent sections of the questionnaire and in the second round of the eDelphi, there was no need to repeat this question in later rounds, especially since it had already reached consensus.

From those who wanted to add, one commented on the ethics of the blueprint:

"From a legal perspective, one of the main issues experienced during the Covid-19 pandemic concerned the protection of fundamental rights related to the information lifecycle (e.g. freedom of expression, freedom to access information, freedom to be informed, freedom not to be informed). Moreover, there were considerable doubts about the fair application of necessity and proportionality principles for the restrictions and limitations of these rights due to the emergency. At a glance, the blueprints of the ENGAGE AI-powered chatbot are more focused on the technical aspects of the process. From my point of view (legal and ethics expertise) the possible social and legal consequences of its operationalisation are not directly addressed. I saw from the project website you are inspired by the use of apps such as ChatGPT. However, the adoption of similar solutions is not free of problems and contradictions. Who should be responsible for the design of the chatbots? And who should answer for their substantial use on a large scale for resilience purposes? A public authority/agency or even private entities? And in both cases, how can they be accountable for the selections of information operated by the chatbots? Who should ensure citizens access to official information? And how may their rights be affected by the mediation of a chatbot? I am sure the quality of information may generally benefit from a such solution. However, I would not neglect the aspects related to accessibility, marginalization and non-discrimination potentially experienceable by minorities and vulnerable groups. I thus suggest considering also the legal/regulatory dimensions within the proposed framework, as a premise to the profitable use of the chatbot. Especially if used by public authorities/agencies, this integration could consolidate the legitimate use of the tool, framing from the early stage of the design the possible mitigation measures for the unwanted consequences on fundamental rights".

Another answer referred to the explainability of algorithms: "I think that it could be add an explainability component to explain algorithms decisions"

The third answer focused on the manual part of the blueprint: "Instead of having a manual human monitoring, an algorithm can be implemented as a checker to the chatbot, which can serve also as a tester and improver/enhancer which will work with way higher through put than a human".

The last concluding questions focused, again, on the blueprint as a whole. The first question asked the experts to what extent they agree or disagree that the process and the interaction between the different components in the blueprint (i.e., the arrows) are appropriate. 2 of them answered "neutral" (12.5%) while the other 14 (87.5%) agreed to "a high extent".

In addition, the experts could add a free-text answer to suggest other components and/or interactions and/or something else they would like to add the the blueprint.

One comment referred to the training process of the chatbot and the intent classification:

"Evaluation method of which metric is the goal during training? I would expect an iterative process somewhere aiming to optimize this or these metric(s). If there is somewhere and I missed it, I would

The research leading to these results has received funding from Horizon 2020, the European Union's a Programme for Research and Innovation (H2020/2014-2020) under grant agreement n° 882850.

perhaps suggest to highlight more. Also, I think intent classification would be very relevant and important - in which step will this take place, will it generate alarms on its own, will it be monitored initially?"

The second comment referred to the choice of training sources:

"A very important part of the training process is the selection of the training sources. In our covid-related chatbot, we choose to include only information provided from the official government websites in order to exclude the possibilities to spread fake news. In an emergency related chatbot, it's mandatory to choose an official source of truth in order to spread legitimate and accurate information (that could be different for other types of chatbots)".

6.2.2 Distribution of answers among the experts to the consensus questions

Table 3 summarises the distribution of answers of the experts to the three questions regarding each component, regarding its technological contribution to the goals of the AI-enabled chatbot, whether it includes all relevant considerations that they would expect from an AI-enabled chatbot's blueprint, and whether it is located in the correct place in the process of the blueprint.

Table 3. Distribution of answers in the first round of the eDelphi.

Component	To what extent do you think that this component	1 Very low	2 Low	3 Neutral	4 High	5 Very high	M (SD)	MD	"High" and "Very High" %
Input types	Technological contribution	1 (6.3%)	1 (6.3%)	0 (0%)	10 (62.5%)	4 (25%)	3.94 (1.063)	4	14 (87.5%)
	Relevant considerations	0 (0%)	0 (0%)	4 (25%)	8 (50%)	4 (25%)	4 (.73)	4	12 (75%)
	Located correctly	0 (0%)	0 (0%)	2 (12.5%)	7 (43.8%)	7 (43.8%)	4.31 (.704)	4	14 (87.6%)
Channels & Platforms	Technological contribution	1 (6.3%)	2 (12.5%)	2 (12.5%)	8 (50%)	3 (18.8%)	3.63 (1.147)	4	11 (68.8%)
	Relevant considerations	0 (0%)	0 (0%)	6 (37.5%)	8 (57.1%)	2 (12.5%)	3.75 (.683)	4	10 (62.5%)
	Located correctly	0 (0%)	0 (0%)	2 (12.5%)	9 (56.3%)	5 (31.3%)	4.19 (.655)	4	14 (87.5%)
UX & logic	Technological contribution	1 (6.3%)	2 (12.5%)	3 (18.8%)	8 (50%)	2 (12.5%)	3.5 (1.095)	4	10 (62.5%)
	Relevant considerations	0 (0%)	2 (12.5%)	7 (43.8%)	6 (37.5%)	1 (6.3%)	3.38 (.806)	3	7 (43.8%)
	Located correctly	0 (0%)	0 (0%)	3 (18.8%)	9 (56.3%)	4 (25%)	4.06 (.68)	4	13 (76.3%)
Security and privacy	Technological contribution	0 (0%)	3 (18.8%)	2 (12.5%)	8 (50%)	3 (18.8%)	3.69 (1.014)	4	11 (68.8%)
	Relevant considerations	0 (0%)	2 (12.5%)	5 (31.3%)	8 (50%)	1 (6.3%)	3.5 (.816)	4	9 (56.3%)
	Located correctly	1 (6.3%)	3 (18.8%)	2 (12.5%)	9 (56.3%)	1 (6.3%)	3.38 (1.088)	4	10 (62.6%)
Logging & documentation	Technological contribution	1	0	5	5	5	3.81	4	10

The research leading to these results has received funding from Horizon 2020, the European Union's a Programme for Research and Innovation (H2020/2014-2020) under grant agreement n° 882850.

Component	To what extent do you think that this component	1 Very low	2 Low	3 Neutral	4 High	5 Very high	M (SD)	MD	"High" and "Very High" %
		(6.3%)	(0%)	(31.3%)	(31.3%)	(31.3%)	(1.109)		(62.5%)
	Relevant considerations	0 (0%)	1 (6.3%)	6 (37.5%)	4 (25%)	5 (31.3%)	3.81 (.981)	4	9 (56%)
	Located correctly	0 (0%)	0 (0%)	5 (31.3%)	8 (50%)	3 (18.8%)	3.88 (.719)	4	11 (68.8%)
Existing platforms	Technological contribution	0 (0%)	0 (0%)	0 (0%)	4 (25%)	12 (75%)	4.75 (.447)	5	16 (100%)
	Relevant considerations	0 (0%)	3 (18.8%)	2 (12.5%)	6 (37.5%)	5 (31.3%)	3.81 (1.109)	4	11 (68.8%)
	Located correctly	0 (0%)	0 (0%)	2 (12.5%)	8 (50%)	6 (37.5%)	4.25 (.663)	4	14 (87.5%)
QA	Technological contribution	0 (0%)	1 (6.3%)	3 (18.8%)	7 (43.8%)	5 (31.3%)	4 (.894)	4	12 (75%)
	Relevant considerations	0 (0%)	1 (6.3%)	5 (31.3%)	3 (18.8%)	7 (43.8%)	4 (1.033)	4	10 (62.5%)
	Located correctly	0 (0%)	1 (6.3%)	4 (25%)	5 (31.3%)	6 (37.5%)	4 (.966)	4	11 (68.8%)
Monitoring & reporting	Technological contribution	0 (0%)	1 (6.3%)	2 (12.5%)	10 (62.5%)	3 (18.8%)	3.94 (.772)	4	13 (81.3%)
	Relevant considerations	0 (0%)	0 (0%)	4 (25%)	11 (68.8%)	1 (6.3%)	3.81 (.544)	4	12 (75%)
	Located correctly	0 (0%)	0 (0%)	4 (25%)	12 (75%)	0 (0%)	3.75 (.447)	4	12 (75%)
Datasets	Technological contribution	0 (0%)	0 (0%)	4 (25%)	4 (25%)	8 (50%)	4.24 (.856)	5	12 (75%)
	Relevant considerations	0 (0%)	0 (0%)	7 (43.8%)	5 (31.3%)	4 (25%)	3.81 (.834)	4	9 (56.3%)
	Located correctly	0 (0%)	1 (6.3%)	3 (18.8%)	5 (31.3%)	7 (43.8%)	4.13 (.957)	4	12 (75%)
Statistics	Technological contribution	0 (0%)	3 (18.8%)	2 (12.5%)	11 (68.8%)	0 (0%)	3.5 (.816)	4	11 (68.8%)
	Relevant considerations	0 (0%)	0 (0%)	8 (50%)	8 (50%)	0 (0%)	3.5 (.516)	4	8 (50%)
	Located correctly	0 (0%)	1 (6.3%)	4 (25%)	9 (56.3%)	2 (12.5%)	3.75 (.775)	4	11 (68.8%)
Alerts	Technological contribution	1 (6.3%)	1 (6.3%)	4 (25%)	8 (50%)	2 (12.5%)	3.56 (1.031)	4	10 (62.5%)
	Relevant considerations	0 (0%)	0 (0%)	6 (37.5%)	9 (56.3%)	1 (6.3%)	3.69 (.602)	4	10 (62.5%)
	Located correctly	0 (0%)	2 (12.5%)	2 (12.5%)	10 (62.5%)	2 (12.5%)	3.75 (.856)	4	12 (75%)
Manual human monitoring	Technological contribution	1 (6.3%)	1 (6.3%)	4 (25%)	7 (43.8%)	3 (18.8%)	3.63 (1.088)	4	10 (62.5%)

The research leading to these results has received funding from Horizon 2020, the European Union's a Programme for Research and Innovation (H2020/2014-2020) under grant agreement n° 882850.

Component	To what extent do you think that this component	1 Very low	2 Low	3 Neutral	4 High	5 Very high	M (SD)	MD	"High" and "Very High" %
	Relevant considerations	1 (6.3%)	2 (12.5%)	3 (18.8%)	6 (37.5%)	4 (25%)	3.63 (1.204)	4	10 (62.5%)
	Located correctly	0 (0%)	2 (12.5%)	2 (12.5%)	7 (43.8%)	5 (31.3%)	3.94 (.998)	4	12 (75%)

The results of the table reveal that two out of the 12 components of the blueprint met the consensus criteria in all three questions: input types and monitoring & reporting. Two other components achieved consensus in two out of three questions: existing platforms and datasets, both excluding the question regarding the inclusion of all relevant considerations, which did not reach a consensus. Five components reached consensus in only one question: QA in the question of technological contribution, and channels & platforms, UX & logic, alerts, and manual human monitoring, only in their location within the blueprint. Security & privacy, logging & documentation, and statistics did not reach a consensus in any question. Generally, the median scores for all questions, including those that did not reach consensus, were high, with most falling between 4 (high) and 5 (very high). Only one question received a median score of 3, pertaining to the relevant considerations in the component of UX & logic. All questions that did not reach consensus proceeded to the second round of the eDelphi, while those that did reach consensus were omitted from the second version.

In addition to the close-ended questions, the experts who rated questions in a "very low" or a "low extent" received additional questions, asking them what they would change or suggest, as experts. These are the answers which were given for the various components, in places they were given:

- Security & Privacy
    - From my point of view, security and compliance requirements should be addressed before. From the users' perspective, they can voluntarily or unintentionally share personal information also in the previous interactions with the tool and its interactions. First of all, this information can be subject to data protection law and thus compliance (especially protection by design) could/should be addressed before. From the chatbot 'management' perspective, this data can be relevant for the correct functioning of the tool (accuracy and relevance of the feedback provided). Therefore, adequate security measures by design and by default should be addressed before.
    - More as a general component, assuring privacy and security at multiple layers (manual operator, platform for developing the chatbot, etc.).
    - Privacy and security are different issues. It is hard to say at this abstraction level.
- Logging & Documentation
    - Scale, privacy by design.
- Existing platforms
    - Explainability of AI models.
    - Wouldn't an evaluation method should be added here, and to which I would add what is the metric we try to optimize in this process?
- QA
    - I marked a low rate due to the legal and ethical concerns related to human oversight. In the big picture, I saw you considered the presence of a team of fact-checkers, but this is located in the following step of the flowchart. If feasible, I

think it is advisable to anticipate this phase and modulate the possibility of human intervention in the previous ones too (at least as a recovery measure).

  - QA needs to be a cross function per component.

- Datasets
  - Would the datasets also feed the quality checks? Unseen data to the methods used should be kept separately to assess their performance (perhaps some similarity evaluation can be used, or something along the lines).
- Statistics & Alerts
  - Maybe it would be good to extract statistics from the external methods used as well, so from the internal processes, and not only input-output, so that when what is returned turns out to be weird, there is a better insight.
- Manual human monitoring
  - maybe it would be good to extract statistics from the external methods used as well, so from the internal processes, and not only input-output, so that when what is returned turns out to be weird, there is a better insight.
  - What about content moderation on all the data that is coming from the user not just the bot feed.

6.2.3 Revision of the blueprint before the first round of the eDelphi

Following the comments of the experts to the open questions, several clarifications and slight changes were made in the blueprint, to address the concerns of the experts and improve the structure of the blueprints. The clarifications and changes were as follows[9]:

- **Security & Privacy**. We added a clarification that the security and privacy component is applied to all parts of the chatbot, providing multiple layers of security, covering not just the chatbot itself but also the other platforms it interacts with. Therefore, addressing also questions of unintentionally sharing information by the user or potential "data leaks" to the host platform.
- **Logging & Documentation.** We added a clarification that the chatbot's blueprint, especially in the logging part, is following privacy by design principles.
- **QA.** We added a clarification that QA is achieved both by humans (e.g., fact-checkers, QA workers) and algorithms.
- **Datasets.** Datasets are also used for manual checks and are separated into seen data, revealed to those who evaluate the chatbot, and unseen data, which serves to assess the evaluation process.
- **Statistics.** The component of statistics is applied not just to the internal data of the chatbot but also to external methods and sources.
- **Alerts.** The component of alerts is applied not just to the internal data of the chatbot but also to external methods and sources.
- **Manual human monitoring.** Manual human monitoring occurs in addition to algorithm monitoring and not instead of it.

[9] The changes are documented in the second questionnaire which is attached in the appendices.

The research leading to these results has received funding from Horizon 2020, the European Union's a Programme for Research and Innovation (H2020/2014-2020) under grant agreement n° 882850.

6.3 SECOND ROUND RESULTS

In the second round of the eDelphi, only questions who did not reach a consensus were presented to the experts, along with the previous evaluation, mean score, and median answers. In addition, they were presented with the changes we applied in the blueprint. Then, they were asked if they wish to reevaluate their answer on each question. Table 4 presents the distribution of results in the second round. In this round, only 15 out of the original 16 experts replied.

Table 4. Distribution of answers in the second round of the eDelphi.

Component	To what extent do you think that this component	1 Very low	2 Low	3 Neutral	4 High	5 Very high	M (SD)	MD	"High" and "Very High" %
Input types	Technological contribution								
	Relevant considerations								
	Located correctly								
Channels & Platforms	Technological contribution	0 (0%)	1 (6.7%)	3 (20%)	9 (60%)	2 (13.3%)	3.8 (.774)	4	11 (73.3%)
	Relevant considerations	0 (0%)	0 (0%)	2 (13.3%)	12 (80%)	1 (6.7%)	3.93 (.457)	4	13 (86.7%)
	Located correctly								
UX & logic	Technological contribution	0 (0%)	0 (0%)	5 (33.3%)	1 (6.7%)	1 (6.7%)	3.73 (.593)	4	10 (66.7)
	Relevant considerations	0 (0%)	0 (0%)	7 (43.8%)	8 (50%)	0 (0%)	3.53 (.516)	4	8 (53.3%)
	Located correctly								
Security and privacy	Technological contribution	0 (0%)	1 (6.7%)	0 (0%)	11 (73.3%)	3 (20%)	4.07 (.704)	4	14 93.3%
	Relevant considerations	0 (0%)	0 (0%)	3 (20%)	8 (53.3%)	4 (25%)	4.07 (.704)	4	12 (80%)
	Located correctly	1 (6.7%)	0 (0%)	3 (20%)	10 (66.7%)	1 (6.7%)	3.67 (.9)	4	11 (73.4%)
Logging & documentation	Technological contribution	0 (0%)	1 (6.7%)	1 (6.7%)	10 (66.7%)	3 (20%)	4 (.756)	4	13 (86.7%)
	Relevant considerations	0 (0%)	0 (0%)	3 (20%)	8 (53.3%)	4 (26.7%)	4.07 (.704)	4	12 (80%)
	Located correctly	0 (0%)	0 (0%)	1 (6.7%)	12 (80%)	2 (13.3%)	4.07 (.458)	4	14 (93.3%)
Existing platforms	Technological contribution								
	Relevant considerations	0 (0%)	0 (0%)	3 (20%)	9 (60%)	3 (20%)	4 (.655)	4	12 (80%)
	Located correctly								
QA	Technological contribution								

Component	To what extent do you think that this component	1 Very low	2 Low	3 Neutral	4 High	5 Very high	M (SD)	MD	"High" and "Very High" %
	Relevant considerations	0 (0%)	0 (0%)	4 (26.7%)	5 (33.3%)	6 (40%)	4.13 (.834)	4	11 (73.3%)
	Located correctly	0 (0%)	0 (0%)	1 (6.7%)	9 (60%)	5 (33.3%)	4.27 (.594)	4	14 (93.3%)
Monitoring & reporting	Technological contribution								
	Relevant considerations								
	Located correctly								
Datasets	Technological contribution								
	Relevant considerations	0 (0%)	0 (0%)	3 (20%)	8 (53.3%)	4 (26.7%)	4.07 (.704)	4	12 (80%)
	Located correctly								
Statistics	Technological contribution	0 (0%)	1 (6.7%)	3 (20%)	11 (73.3%)	0 (0%)	3.67 (.618)	4	11 (73.3%)
	Relevant considerations	0 (0%)	0 (0%)	0 (0%)	15 (100%)	0 (0%)	4 (0)	4	15 (100%)
	Located correctly	0 (0%)	0 (0%)	1 (6.7%)	13 (86.7%)	1 (6.7%)	4 (.378)	4	14 (93.4%)
Alerts	Technological contribution	1 (6.7%)	0 (0%)	4 (26.7%)	9 (60%)	1 (6.7%)	3.6 (.91)	4	10 (66.7%)
	Relevant considerations	0 (0%)	0 (0%)	2 (13.3%)	12 (80%)	1 (6.7%)	3.93 (.458)	4	13 (86.7%)
	Located correctly								
Manual human monitoring	Technological contribution	0 (0%)	0 (0%)	5 (33.3%)	8 (53.3%)	2 (13.3%)	3.8 (.676)	4	10 (66.7%)
	Relevant considerations	0 (0%)	0 (0%)	3 (20%)	8 (53.3%)	4 (26.7%)	4.07 (.704)	4	12 (80%)
	Located correctly								

Components which reached consensus in the first round appear in dark green, without numbers. Components which reached consensus in the second round appear in light green with the updated statistics. Components which did not reach consensus have no background color.

The results of the second round of the eDelphi indicate that nearly all components reached full consensus, except for three. Alerts and manual human monitoring achieved consensus regarding the relevant considerations, in addition to their correct location in the blueprint, where they had reached consensus in the first round. However, the experts still did not reach a consensus regarding their technological contribution, with a marginal 10 out of 15 experts (66.7%) agreeing to a "high" or "very high extent" that these components have technological contribution (M=3.6, SD=.91 and M=3.8, SD=.676). The UX & logic component, while reaching consensus regarding its correct location in the blueprint, failed to reach consensus in the second round concerning its technological contribution (10/15, 66.7% rating it "high" or "very high", with M=3.73, SD=.593) and relevant considerations (8/15, 53.3% rating it "high" or "very high", with M=3.53, SD=.516).

The research leading to these results has received funding from Horizon 2020, the European Union's a Programme for Research and Innovation (H2020/2014-2020) under grant agreement n° 882850.

6.4 Areas of consensus vs. continues disagreements

Table 5 compares the results of the first and second rounds of the eDelphi.

Table 5. A comparison of the two rounds of the eDelphi.

Component	To what extent do you think that this component	M (1st round) (SD)	M (2nd round) (SD)	MD (1st round)	MD (2nd round)	"High" and "Very High" % (1st round)	"High" and "Very High" % (2nd round)	Improvement
Input types	Technological contribution	3.94 (1.063)	Consensus	4	Consensus	14 (87.5%)	Consensus	Consensus
	Relevant considerations	4 (.73)	Consensus	4	Consensus	12 (75%)	Consensus	Consensus
	Located correctly	4.31 (.704)	Consensus	4	Consensus	14 (87.6%)	Consensus	Consensus
Channels & Platforms	Technological contribution	3.63 (1.147)	3.8 (.774)	4	4	11 (68.8%)	11 (73.3%)	0 (4.5%)*
	Relevant considerations	3.75 (.683)	3.93 (.457)	4	4	10 (62.5%)	13 (86.7%)	3 (24.2%)
	Located correctly	4.19 (.655)	Consensus	4	Consensus	14 (87.5%)	Consensus	Consensus
UX & logic	Technological contribution	3.5 (1.095)	3.73 (.593)	4	4	10 (62.5%)	10 (66.7)	0 (5.2%)*
	Relevant considerations	3.38 (.806)	3.53 (.516)	3	4	7 (43.8%)	8 (53.3%)	1 (9.5%)
	Located correctly	4.06 (.68)	Consensus	4	Consensus	13 (76.3%)	Consensus	Consensus
Security and privacy	Technological contribution	3.69 (1.014)	4.07 (.704)	4	4	11 (68.8%)	14 93.3%	3 (24.5%)
	Relevant considerations	3.5 (.816)	4.07 (.704)	4	4	9 (56.3%)	12 (80%)	3 (23.7%)
	Located correctly	3.38 (1.088)	3.67 (.9)	4	4	10 (62.6%)	11 (73.4%)	1 (10.8%)
Logging & documentation	Technological contribution	3.81 (1.109)	4 (.756)	4	4	10 (62.5%)	13 (86.7%)	3 (24.2%)
	Relevant considerations	3.81 (.981)	4.07 (.704)	4	4	9 (56%)	12 (80%)	3 (23.7)
	Located correctly	3.88 (.719)	4.07 (.458)	4	4	11 (68.8%)	14 (93.3%)	3 (24.5%)
Existing platforms	Technological contribution	4.75 (.447)	Consensus	5	Consensus	16 (100%)	Consensus	Consensus
	Relevant considerations	3.81 (1.109)	4 (.655)	4	4	11 (68.8%)	12 (80%)	1 (11.2%)
	Located correctly	4.25 (.663)	Consensus	4	Consensus	14 (87.5%)	Consensus	Consensus
QA	Technological contribution	4 (.894)	Consensus	4	Consensus	12 (75%)	Consensus	Consensus

The research leading to these results has received funding from Horizon 2020, the European Union's a Programme for Research and Innovation (H2020/2014-2020) under grant agreement n° 882850.

Component	To what extent do you think that this component	M (1st round) (SD)	M (2nd round) (SD)	MD (1st round)	MD (2nd round)	"High" and "Very High" % (1st round)	"High" and "Very High" % (2nd round)	Improvement
	Relevant considerations	4 (1.033)	4.13 (.834)	4	4	10 (62.5%)	11 (73.3%)	1 (10.8%)
	Located correctly	4 (.966)	4.27 (.594)	4	4	11 (68.8%)	14 (93.3%)	3 (24.5%)
Monitoring & reporting	Technological contribution	3.94 (.772)	Consensus	4	Consensus	13 (81.3%)	Consensus	Consensus
	Relevant considerations	3.81 (.544)	Consensus	4	Consensus	12 (75%)	Consensus	Consensus
	Located correctly	3.75 (.447)	Consensus	4	Consensus	12 (75%)	Consensus	Consensus
Datasets	Technological contribution	4.24 (.856)	Consensus	5	Consensus	12 (75%)	Consensus	Consensus
	Relevant considerations	3.81 (.834)	4.07 (.704)	4	4	9 (56.3%)	12 (80%)	3 (23.7%)
	Located correctly	4.13 (.957)	Consensus	4	Consensus	12 (75%)	Consensus	Consensus
Statistics	Technological contribution	3.5 (.816)	3.67 (.618)	4	4	11 (68.8%)	11 (73.3%)	0 (4.5%)*
	Relevant considerations	3.5 (.516)	4 (0)	4	4	8 (50%)	15 (100%)	7 (50%)
	Located correctly	3.75 (.775)	4 (.378)	4	4	11 (68.8%)	14 (93.4%)	3 (24.5%)
Alerts	Technological contribution	3.56 (1.031)	3.6 (.91)	4	4	10 (62.5%)	10 (66.7%)	0 (4.2%)*
	Relevant considerations	3.69 (.602)	3.93 (.458)	4	4	10 (62.5%)	13 (86.7%)	3 (24.2%)
	Located correctly	3.75 (.856)	Consensus	4	Consensus	12 (75%)	Consensus	Consensus
Manual human monitoring	Technological contribution	3.63 (1.088)	3.8 (.676)	4	4	10 (62.5%)	10 (66.7%)	0 (4.2%)
	Relevant considerations	3.63 (1.204)	4.07 (.704)	4	4	10 (62.5%)	12 (80%)	2 (17.5%)
	Located correctly	3.94 (.998)	Consensus	4	Consensus	12 (75%)	Consensus	Consensus

* The change of percentage is due to the decrease of the number of experts from 16 to 15 between the rounds.

The comparison reveals that the most significant improvement in the consensus score occurred in the component of statistics, specifically concerning the question of relevant considerations – the agreement on a "high" or "very high extent" increased from 50% to 100% (all of which were rated as "high extent"). Regarding the technological contribution of alerts and manual human monitoring, which did not reach consensus in the second round, the data also shows that the number of experts who agreed on the "high" or "very high" extent of the technological contribution remained unchanged. A similar pattern is observed concerning the technological contribution question of the UX & logic component, while the change in the relevant consideration question was only by one.

The research leading to these results has received funding from Horizon 2020, the European Union's a Programme for Research and Innovation (H2020/2014-2020) under grant agreement n° 882850.

6.5 Conclusions from the eDelphi

The results of the eDelphi process provide valuable insights for the revised blueprint. Components that achieved consensus in both rounds can be confidently integrated into the blueprint, ensuring that they align with expert opinions. For components that did not achieve full consensus, especially in areas like technological contribution, further discussions were necessary. These areas of disagreement highlight the complexities and challenges in developing an AI-enabled chatbot blueprint for emergencies. They underscore the need for a balanced approach that considers both technological feasibility and the practical implications of each component.

One component that did not reach a consensus was the one of "alerts", in the context of technological contribution to the goals of the AI-enabled chatbot. This lack of consensus indicates the complexity and variability of opinions on how alerts should be integrated and managed within the chatbot system. There are several explanations for this.

Alerts, in many digital systems, are a well-established feature. The technology behind sending alerts is relatively mature and straightforward compared to other, more advanced components of a chatbot system. As such, experts might view it as a basic feature rather than a cutting-edge technological contribution. Additionally, while the concept of an alert is simple, its effective implementation in a crisis communication context can be complex. Ensuring that alerts reach the right people at the right time with the right information is challenging. Experts might believe that while the technological aspect of sending an alert is simple, the broader system in which it operates (ensuring accuracy, timeliness, and relevance) might not be as technologically advanced. Other explanations could include:

- **Over-reliance Concerns**: There might be concerns about over-relying on automated alerts without human verification, especially in critical situations. Misinformation or false alarms can have severe consequences, and experts might be wary of placing too much trust in an automated system for such crucial tasks.
- **Integration Challenges**: Alerts need to be integrated with various other systems to be effective. This includes data sources, verification systems, and communication channels. The technological challenges in achieving seamless integration might make experts skeptical about the high technological contribution of the alert component in isolation.
- **User Fatigue**: There's a well-documented phenomenon where users become desensitized to frequent alerts, leading to alert fatigue. If alerts are not managed and curated effectively, they can lose their impact. Experts might be considering the broader implications of how alerts are received and acted upon by users, rather than just the technology behind sending them.
- **Comparative Evaluation**: When evaluating the technological contribution of various components in a system, experts might be comparing the "alert" component to other, more technologically advanced or innovative components. In such a comparative scenario, alerts might seem less technologically significant.

The manual human monitoring component was another area that garnered attention. While it reached a consensus in its location within the blueprint and the relevant considerations, it did not achieve a unanimous agreement in the aspect of technological contribution. This highlights the challenges and varying perspectives on how human intervention should be incorporated into automated systems. Manual human monitoring is crucial to ensure the accuracy and appropriateness of chatbot responses, especially in critical situations. However, the exact implementation and

The research leading to these results has received funding from Horizon 2020, the European Union's a Programme for Research and Innovation (H2020/2014-2020) under grant agreement n° 882850.

integration of this component require careful consideration, balancing automation with human oversight. Similarly, there are several explanations for this.

By its very nature, "human manual monitoring" emphasizes human intervention over automated processes. Experts might view this component as more of a procedural or operational element rather than a technological innovation. The reliance on human judgment, while crucial, might not be seen as a cutting-edge technological contribution.

Additionally, manual monitoring, in various forms, has been a part of many systems long before the advent of advanced AI and machine learning. As such, its inclusion in a chatbot system might be seen as a continuation of an established practice rather than a novel technological advancement.

Moreover, one of the significant advantages of automated systems is their scalability. Manual human monitoring, by contrast, doesn't scale as efficiently. As the volume of interactions or data increases, the need for human monitors grows linearly, which might be seen as a technological limitation. Other explanations could include:

- **Subjectivity and Consistency**: Human monitors, while invaluable for their judgment and expertise, can introduce variability. Different monitors might make different decisions given the same data or situation. This potential lack of consistency might be viewed as a technological drawback compared to more deterministic automated systems.
- **Comparative Evaluation**: In the context of a chatbot system with various components, experts might be comparing the "human manual monitoring" component to others that leverage more advanced technologies, such as machine learning algorithms or natural language processing. In such a comparative scenario, manual monitoring might seem less technologically innovative.
- **Integration Challenges**: While the act of human monitoring might not be technologically advanced, integrating human decisions seamlessly into an automated system can be challenging. However, this integration challenge might be seen more as a system design issue rather than a technological innovation.
- **Evolution of AI**: As AI and machine learning technologies advance, there's a growing belief in their potential to handle tasks traditionally reserved for humans. While we're not there yet, the trajectory might lead some experts to view human intervention as a temporary solution until AI reaches a certain level of sophistication.

Therefore, and since the lack of experts consensus related only to the question of technological contribution, we decided to leave these components of "alerts" and "manual human monitoring" unchanged.

The User Experience (UX) and logic component of the chatbot blueprint underwent significant scrutiny. The results revealed that while there was consensus on its location within the blueprint, there was no unanimous agreement on the inclusion of all relevant considerations for this component and the technological contribution. This suggests that while experts recognize the importance of UX and logic in chatbot design, there are differing opinions on what constitutes "relevant considerations" for this component or "technological contribution". The challenge lies in designing a user interface that is intuitive and user-friendly while ensuring the underlying logic is robust and efficient. There are a few possible explanations.

User Experience (UX) design principles have been around for a long time and are applied across various digital platforms, not just chatbots. Experts might view UX as a foundational aspect of any digital tool rather than a standout technological feature. It's a given that any system should have a user-friendly interface, so it might not be seen as a unique technological contribution. Moreover, what constitutes a good user experience can be highly subjective and varies based on individual preferences, cultural contexts, and specific use cases. This inherent subjectivity might make it challenging to reach a consensus on its technological contribution or its relevant considerations.

The research leading to these results has received funding from Horizon 2020, the European Union's a Programme for Research and Innovation (H2020/2014-2020) under grant agreement n° 882850.

Hence, while UX is crucial for the effective use of technology, it's often seen as a separate discipline from the core technological development. The design of an interface (UX) and the underlying algorithms and data structures (logic) might be developed by different teams with different expertise. This separation might lead experts to undervalue the technological contribution of UX.

Additionally, the logic behind chatbots, while complex, might be based on established algorithms and decision trees that have been in use for some time. Unless there are innovative machine learning or AI components introduced, experts might view the logic as a standard feature rather than a cutting-edge technological advancement. A part of the "existing platforms" section in the blueprint. Therefore, in the broader context of a chatbot system, experts might be comparing the "UX and logic" component to other components that leverage more advanced technologies. In such a comparative scenario, UX and logic might seem foundational rather than innovative.

In technology circles, there's sometimes an overemphasis on functionality and features at the expense of user experience. Experts might prioritize components that introduce new capabilities over those that improve usability, leading to a lower valuation of the UX and logic component's technological contribution.

Finally, unlike other components that can be evaluated based on metrics like accuracy, speed, or scalability, the success of UX design is harder to quantify. This lack of clear metrics might make it challenging to evaluate its technological contribution objectively.

Based on that, several revisions were made to the "UX & Logic" component as will be elaborated in the next section.

The research leading to these results has received funding from Horizon 2020, the European Union's
a Programme for Research and Innovation (H2020/2014-2020) under grant agreement n° 882850.

7 REVISED BLUEPRINT

Based on the results of the eDelphi consensus technique, we validated the core of the AI-enabled chatbot's blueprint for communication with the public in relation to disasters. In light of the technological developments achieved post the creation of the initial blueprint (e.g., ChatGPT and GPT 4), and based on the evaluation of the experts, we suggested several minor revisions in the final blueprint.

7.1 AI-ENABLED BLUEPRINT V2.0

7.1.1 FINAL BLUEPRINT

Figure 5 presents the revised version of the blueprint.

7.1.2 WHAT REMAINS THE SAME IN THE BLUEPRINT?

The following components were left in the blueprint, as in the original suggestion:

- **Input Types**: This component is foundational for any chatbot system. It determines how users can interact with the chatbot, whether through text, voice, images, or other means. The consensus on this component underscores its fundamental importance in ensuring that users can communicate with the chatbot in diverse and effective ways.
- **Channels and Platforms**: The platforms on which the chatbot operates and the channels it uses for communication are crucial for its accessibility and reach. Retaining this component in the blueprint ensures that the chatbot can be accessed by a wide audience across various platforms, enhancing its utility during emergencies.
- **Security and Privacy**: Given the sensitive nature of emergencies and the potential for personal data exchange, security and privacy are paramount. The consensus on this component highlights the unanimous understanding among experts about the importance of safeguarding user data and ensuring secure interactions. This component applies to all parts of the chatbot, providing multiple layers of security, covering not just the chatbot but all the platforms it interacts with.
- **Logging and Documenting**: Keeping a record of interactions can be invaluable for post-emergency analysis, refining the chatbot's responses, and ensuring accountability. The retention of this component emphasizes the importance of transparency and continuous improvement. All following the privacy by design principles.
- **Existing Platforms**: Leveraging existing platforms, such as Open AI's GPT-4 and ChatGPT, midjourney and Google Vision, can expedite the chatbot's development and deployment. By building on established systems, the chatbot can benefit from their features and user base, ensuring a quicker and more effective response during emergencies.
- **Quality Assurance**: Ensuring that the chatbot provides accurate and reliable information is crucial, especially during emergencies when stakes are high. The consensus on this component underscores the need for rigorous testing and refinement to ensure the chatbot's reliability. QA is conducted both by humans (e.g., fact checkers, QA workers) and by algorithms.

The research leading to these results has received funding from Horizon 2020, the European Union's a Programme for Research and Innovation (H2020/2014-2020) under grant agreement n° 882850.

- **Monitoring and Reporting**: Continuous monitoring allows for real-time adjustments, ensuring that the chatbot remains effective during dynamic emergency situations. Reporting, on the other hand, provides insights into the chatbot's performance, facilitating ongoing improvements.
- **Statistics and Reports**: Data-driven insights can be invaluable for emergency response teams. By analyzing user interactions and generating reports, the chatbot can provide a clearer picture of the situation on the ground, aiding decision-making processes. This is based on internal statistics, but also on external sources and methods.
- **Alerts**: Timely alerts can save lives during emergencies. The alerts component also refers to identifying discourse pattern or any other irregularities and alerting the emergency organizations.
- **Datasets**: The quality of a chatbot's responses is only as good as the data it's trained on. Retaining the datasets component ensures that the chatbot has access to relevant and accurate information, enhancing its effectiveness during emergencies. Data is divided into seen data, used for evaluation of the chatbot, and unseen data, serves for assessing the evaluation process.
- **Manual human monitoring**: This component emphasizes the importance of human oversight in an AI-enabled chatbot system, especially during emergencies. While AI can process vast amounts of data quickly and provide automated responses, human judgment is invaluable in nuanced or complex situations. By retaining this component in the blueprint, it underscores the recognition among experts that while automation is beneficial, human intervention remains crucial to ensure the accuracy, appropriateness, and sensitivity of chatbot responses.

7.1.3 What revisions were made?

The final blueprint presents several minor revisions and add-ons.

7.1.3.1 "Universal Web Integration Plugin (UWIP)

One add-on to the blueprint is in integration with the "channels and platforms" component. The "Universal Web Integration Plugin (UWIP)" is a natural extension of the "Channels & Platforms" component. While the "Channels & Platforms" component emphasizes the chatbot's accessibility across various platforms, the UWIP ensures that this accessibility is not just limited to major platforms but can be extended to any website globally. This plugin acts as a bridge, seamlessly integrating the chatbot into diverse web environments, ensuring that the chatbot's reach is truly universal in a large scale.

The UWIP is a component designed to amplify the reach and accessibility of the AI-enabled chatbot for emergencies. Recognizing the diverse digital landscapes across countries and the myriad of websites people frequent, the UWIP ensures that the chatbot can be effortlessly integrated into any website, regardless of its origin or nature. This not only democratizes access to the chatbot but also ensures that crucial emergency information and interactions are just a click away for users, no matter where they are online. By providing a simple and efficient integration process, the UWIP ensures that the chatbot becomes a ubiquitous presence, ready to assist and guide users during emergencies on a truly global scale.

7.1.3.2 Dynamic Data Integration (DDI) for databases

Another add-on is the Dynamic Data Integration (DDI) for databases. This is an advanced enhancement of the "Datasets" component. While the original "Datasets" component ensures the chatbot has access to relevant and accurate information, the DDI goes a step further by ensuring this data is always up-to-date and comprehensive. By connecting directly to the data sources of organizations, the DDI ensures real-time access to a plethora of information, from regulations to scenario documentation.

The DDI represents a paradigm shift in how chatbots access and utilize data. Traditional chatbots rely on periodic training sessions, where they are updated with new information. However, this approach has limitations, especially in dynamic environments like emergency management where information is continuously evolving. The DDI addresses this challenge by establishing live connections with the data repositories of organizations. Whether it's the latest regulations, newly documented emergency scenarios, or any other pertinent information, the chatbot has real-time access to it. This not only ensures that the chatbot's responses are always based on the most recent and accurate data but also eliminates the time and resources typically spent on repeated training sessions. In essence, the DDI transforms the chatbot into a constantly evolving entity, always in sync with the organization's knowledge base, ready to provide users with timely and accurate information during emergencies.

7.1.3.3 Meta Prompts

The "Meta Prompts" component acts as a mediator between the "UX & Logic" component and the "Existing Platforms" component. While the "Existing Platforms" provides the foundational structure and capabilities of the chatbot, and the "UX & Logic" defines its interaction style, the "Meta Prompts" ensures that these interactions are tailored specifically for emergencies and disasters. It serves as a guiding layer, instructing the chatbot on the nuances of communication, especially in the sensitive context of emergencies.

In the vast realm of AI chatbot interactions, ensuring specificity and relevance is a challenge. The introduction of "Meta Prompts" addresses this challenge head-on. These prompts are essentially high-level instructions that guide the chatbot's behavior, ensuring it remains focused on its primary purpose: assisting users during emergencies and disasters.

While the chatbot has the capability to answer a wide range of questions, the "Meta Prompts" limit this scope, ensuring that the chatbot doesn't venture into unrelated territories. This is crucial to maintain the chatbot's credibility and to ensure users receive relevant and timely assistance during crises.

But the role of "Meta Prompts" goes beyond just limiting the chatbot's scope. It also defines the manner in which the chatbot interacts. Emergencies are high-stress situations, and the tone, empathy, patience, and calmness exhibited by the chatbot can make a significant difference in user experience. The "Meta Prompts" ensure that the chatbot's responses are not just accurate but also emotionally attuned to the user's state, providing comfort and clarity during trying times.

Here are three examples for meta prompts and how they can be used in the chatbot:

1. **Emergency Relevance Prompt**:
   1.1. Meta Prompt: "Prioritize queries related to emergency situations. If a query is unrelated to emergencies or disasters, gently redirect the user to the main topic."
   1.2. User Query: "What's the weather like tomorrow?"
   1.3. Chatbot Response (guided by the meta prompt): "I'm here to assist with emergency situations. If you have concerns related to a potential emergency or disaster, please let me know. For general weather updates, please check a reliable weather website or app."
2. **Empathy and Assurance Prompt**:
   2.1. Meta Prompt: "Always respond with empathy and assurance during distressing situations. Provide comfort and clear guidance."
   2.2. User Query: "I'm scared. There's a storm warning in my area."

The research leading to these results has received funding from Horizon 2020, the European Union's a Programme for Research and Innovation (H2020/2014-2020) under grant agreement n° 882850.

2.3. Chatbot Response (guided by the meta prompt): "I understand your concern. Please stay calm and safe. Here are the recommended safety measures for storms in your area..."

3. **Information Verification Prompt**:

3.1. Meta Prompt: "Always verify the accuracy of emergency information before providing it. If unsure, guide the user to official sources."

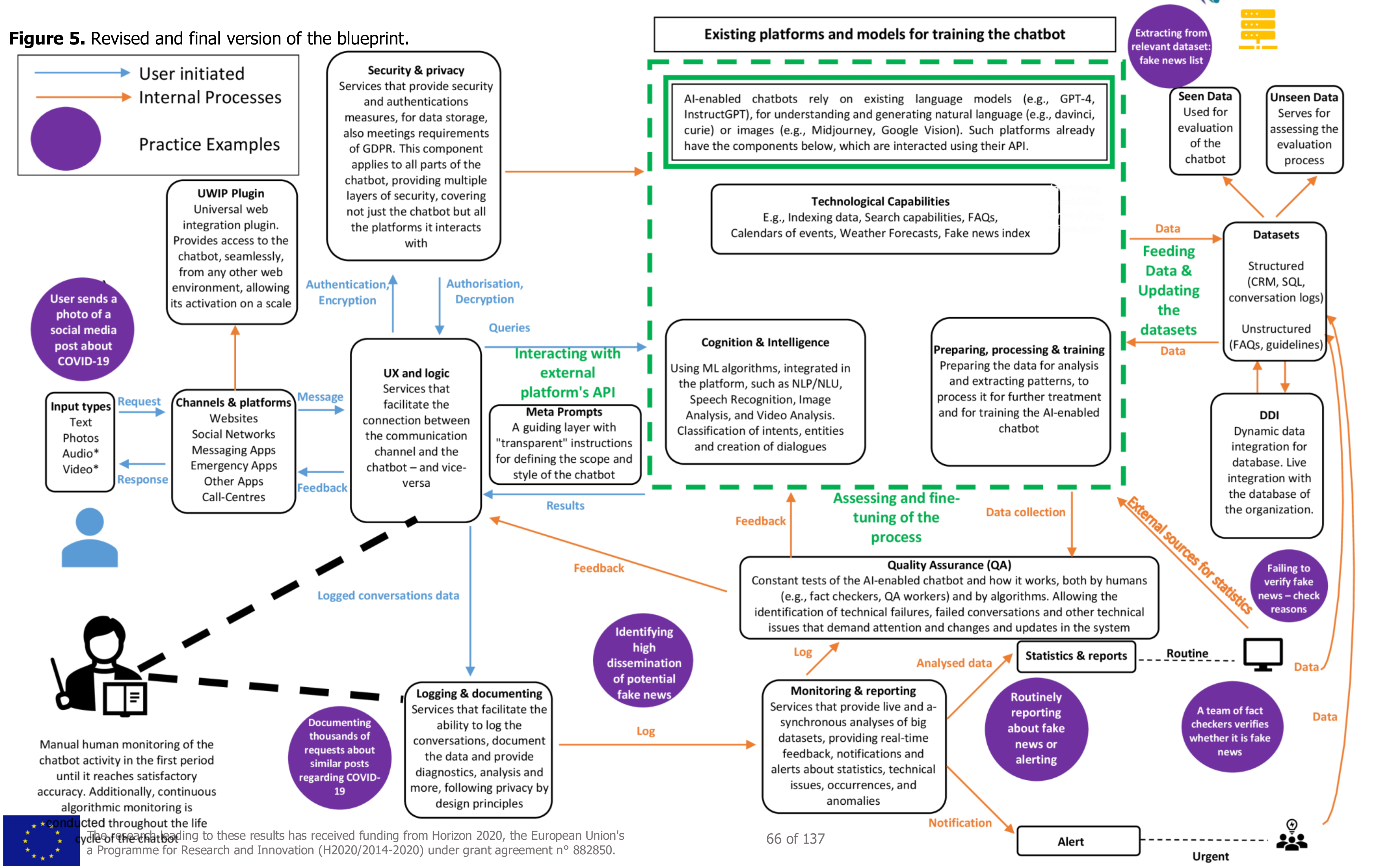
3.2. User Query: "I heard there's a flood warning in my city. Is it true?"

3.3. Chatbot Response (guided by the meta prompt): "Let me check the latest official updates for you. In the meantime, please ensure you're in a safe location and refer to local authorities or news channels for immediate information."

The addition of the "Meta Prompts" component was driven by the realization that while technology can provide answers, the manner and context in which these answers are delivered are equally important. The lack of consensus on the "UX & Logic" component indicated a need for clearer guidelines on chatbot behavior. "Meta Prompts" fill this gap, providing a framework that ensures the chatbot remains a focused, empathetic, and effective conversational agent during emergencies and disasters. It ensures that the chatbot's interactions are not just technologically sound but also emotionally resonant, catering to the unique needs of users during crises.

The research leading to these results has received funding from Horizon 2020, the European Union's a Programme for Research and Innovation (H2020/2014-2020) under grant agreement n° 882850.

Figure 5. Revised and final version of the blueprint.



The research leading to these results has received funding from Horizon 2020, the European Union's a Programme for Research and Innovation (H2020/2014-2020) under grant agreement n° 882850.

7.2 ChatGPT's role in AI-enabled chatbots for emergencies

Since the establishment of GPT-4 architecture and ChatGPT, allowing integration of existing chatbots with advanced NLP capabilities, we particularly examined the application of the AI-enabled chatbot's blueprint with it.

ChatGPT, powered by the GPT-4 architecture, offers a transformative approach to AI-enabled chatbot systems, especially in the context of emergencies. Its advanced natural language processing capabilities allow for nuanced, context-aware interactions that can be tailored to the specific needs of emergency situations. Integrating ChatGPT into the proposed blueprint enhances the chatbot's conversational capabilities and ensures seamless integration with various components, especially existing platforms. This is pivotal for organizations aiming to develop AI-enabled platforms for emergency response. By leveraging the capabilities of ChatGPT and the GPT-4 architecture, organizations can expedite the development process, ensuring that their chatbots are both technologically advanced and contextually relevant. The adaptability of ChatGPT means it can be molded to fit the unique requirements of different organizations, making the blueprint a versatile foundation for a wide range of AI-enabled emergency response platforms.

The following can be initial steps of using Open-AI's GPT-4 architecture for emergency organizations[10]:

API Setup: Begin by setting up the GPT-4 API. OpenAI provides comprehensive documentation to get started. Ensure you have the necessary API keys and access permissions.

Database Integration: To integrate organizational databases with ChatGPT, establish a middleware layer that can fetch data from these databases in real-time. This middleware will communicate with the GPT-4 API, sending prompts that might require data and receiving responses which can then be relayed to the user.

Prompt Design: Designing effective prompts is crucial. Given the emergency context, prompts should be clear and specific. For instance, instead of a generic "Provide information," use "Provide emergency evacuation procedures for [specific scenario]."

Meta Prompts: As discussed earlier, meta prompts guide the chatbot's behavior. With ChatGPT, these can be integrated as initial instructions in each interaction, ensuring the chatbot's responses align with the desired tone, context, and scope.

Continuous Training: While GPT-4 is pre-trained on vast datasets, fine-tuning it for emergency scenarios is beneficial. Use real-world data, scenarios, and feedback to train the model, ensuring its responses are tailored to emergency contexts.

Integration with Existing Channels and Platforms: ChatGPT can be integrated into existing platforms using the GPT-4 API. Whether it's a web platform, mobile app, or any other digital interface, ensure that the user interface is designed to facilitate seamless interactions with ChatGPT.

Security and Privacy: When integrating databases, especially in the context of emergencies, ensure that all data transfers are encrypted. Also, implement measures to anonymize user data, ensuring privacy.

Monitoring and Feedback Loop: Establish a system to monitor ChatGPT's interactions. This not only helps in ensuring accuracy but also provides data for continuous improvement. Feedback from users can be used to refine prompts and improve the chatbot's overall effectiveness.

[10] Based on the overview section on Open AI's website: https://platform.openai.com/overview.

The research leading to these results has received funding from Horizon 2020, the European Union's a Programme for Research and Innovation (H2020/2014-2020) under grant agreement n° 882850.

Scalability: Given the potential high demand during emergencies, ensure that the infrastructure supporting ChatGPT is scalable. This might involve optimizing API call rates, ensuring server capacity, and implementing load balancers.

The research leading to these results has received funding from Horizon 2020, the European Union's a Programme for Research and Innovation (H2020/2014-2020) under grant agreement n° 882850.

8 Warning Systems Chatbot Prototype

8.1 Prototype feature implementation and mapping to blueprint

The prototype chatbot is hosted as a website (as the only channel) and currently supports text and suggestion chips as input. The chatbot is hosted on Amazon Web Services (AWS) and runs in a virtual EC2 (Elastic Compute Cloud) instance. However, it could be hosted on any cloud or private hosting platform as no specific functionality from AWS was used. For building the User Experience (UX) and logic part, the Dart language was chosen because of the choice to create the web application with the Flutter framework. Flutter was suitable because it allows application development from just one codebase, which can run on any web browser making it very efficient effort-wise and inclusive.

The UX and logic receives the CAP alert message in electronic format and presents the landing page with details about the emergency, for instance the severity, location and expiration date. Based on the blueprint analysis on UX we also incorporated in the chatbot suggestion chips and in the design, indications of how to come into contact with a human when needed.

Another essential element which the blueprint brought forward was the use of a channel which is known and trusted. For this we made considerable effort to come up with a strong identity for the chatbot. Although the chatbot is integrated in the 'regular' Public Warning System, users need to be able to recognise and trust the chatbot in order to use it. Efforts made to reach this goal were: a concise and short domain name was chosen, crisis.chat, which is used in the public warning alert messages and a consistent colour schema was adapted conform UI and UX guidelines[11].

For the Cognition & Intelligence part, the first NLP solution chosen was Google Dialogflow. The initial implementation of the prototype called the AI platform to identify intents in user questions and mapped the identified intents by Google Dialogflow back into answers. The actual answer was created by the logic mapping intent back to answer, together with the information retrieved from the CAP message.

A limitation imposed by this way of mapping is that the amount and quality/content of answers given is limited by the amount and completeness of the data fields in the alert message, as no other information sources are used. The opposite of this limitation can also be perceived as a benefit: all provided answers (and contents) are thus fully in control of the alert issuing agency.

8.2 Development iterations and the chatbot roadmap

Development of the prototype occurred in several iterations.

- Initial working version included Google Dialogflow as NLP solution.
- Second version that added OpenAI ChatGPT as 2nd engine.
- Third version where the AI engine and intent mapping was separated as driver and logging.
- Planned fourth version with additional user location based information and creation of database with informative resources from authoritative organisations alongside CAP message.

[11] https://koenvogel.net/wp-content/uploads/2023/07/Internship_Report_Everbridge.pdf

The research leading to these results has received funding from Horizon 2020, the European Union's a Programme for Research and Innovation (H2020/2014-2020) under grant agreement n° 882850.

The iterations are explorative in nature where we get familiar with the technology stack, experiment with implementing new features, then test and receive feedback, which then serve as input for the next iteration.

After this first version of the chatbot prototype was built, OpenAI released the ChatGPT chatbot in November 2022. The project then explored how OpenAI's language models could be beneficial and incorporated into the chatbot. The first implementation used the Davinci 'text-davinci-003' model[12], as a fallback method, in case Dialogflow could not provide a suitable intent mapping. A prompt created for the GPT-3 model based on the alert message allowed steering the answer to the hazard in question. All of this was still implemented within the UX and logic.

With possibly more suitable NLP solutions being introduced (that could be utilized) and all AI related handling residing within the UX and logic component, this part would become complicated to maintain and extend. Supporting additional channels like voice and SMS would also be difficult. Thus, the decision was made to move all AI related logic to a separate component. This challenge was defined as a separate project which was given to the Saxion University in Deventer.

This resulted in the development of a separate AI driver. This driver implements all logic particular to calling an NLP solution, including in our case the Dialogflow intent mapping to answer and engineering the prompt required by the OpenAI API and the different NLP API's.

The UX and logic now only has one interface to the AI driver and over this interface queries are issued, and answer responses are returned, mapping well into the blueprint's Cognition & Intelligence component. An additional benefit of the separate AI driver is that this component is now generally re-usable, allowing also other channels to make use of AI functionality.

Last functionality added which complies with the blueprint, is the logging and documenting feature to the chatbot. The logging functionality logs the combination of timestamp, alert identifier, question, and response. This logging function has been primarily intended to manually analyze what questions the public asks and if appropriate answers are given by the solution. The logging is done in Amazon OpenSearch instance whilst Kibana is used for ad-hoc reporting.

At this time of writing, it is planned to include datasets for handling instructions on common crisis scenario's and including also location data as this would significantly increase the capability of the chatbot to provide specific answers like nearest evacuation sites and evacuation routes. This would extend the chatbot exposure in the blueprint technical capabilities.
In these iterations we expanded the chatbot according to the blueprint roadmap on Technical Capability, we finally added logging to the chatbot and by having separated the AI logic from the presentation logic there is now possibility to allow multiple channels to the chatbot and multiple AI engines. The logging could facilitate (automated) monitoring but such logic has not been planned yet. The logging could also facilitate in applying the fourth step in the roadmap in adopting more analytical models and ML use-cases such as finetuning keywords for intents or providing answers to the public based on previously asked questions for identical crisis.

On the other parts of the roadmap the prototype did not advance since no public tests were done and no big datasets were used.

8.3 Preliminary tests

In the initial stages usability testing was not feasible due to the lack of time and datasets. Instead, evaluation was done on the prototype in an iterative process during design and prototype

[12] https://platform.openai.com/docs/models

The research leading to these results has received funding from Horizon 2020, the European Union's a Programme for Research and Innovation (H2020/2014-2020) under grant agreement n° 882850.

presentations. Various ENGAGE members and developers from One2Many provided feedback during these moments, which resulted in a continuous integration of enhancements of the prototype[13].

8.4 Testing the prototype

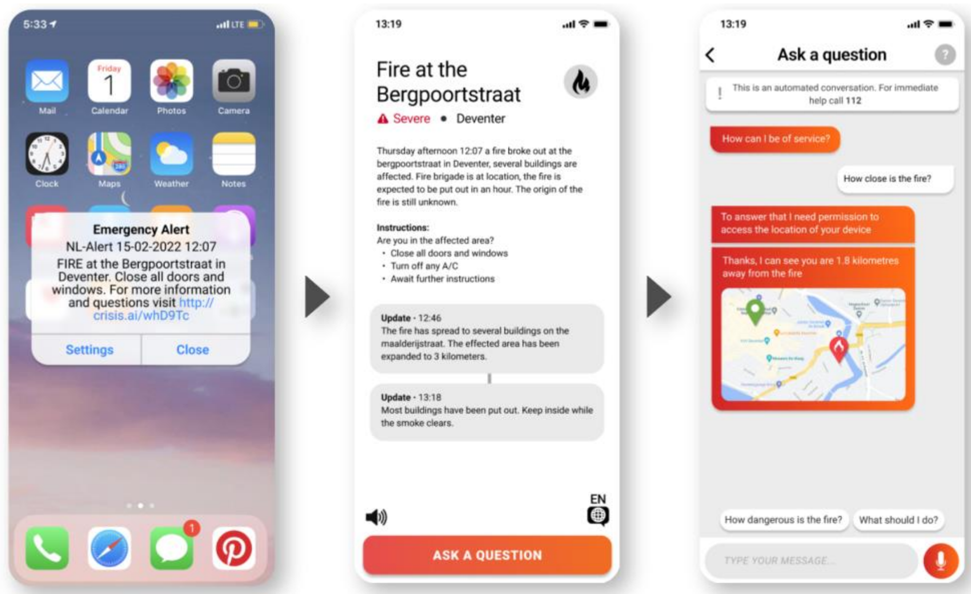
The chatbot prototype is integrated in the Public Warning System of One2Many, and therefore end-to-end system testing is done in the One2Many lab which includes a Cell Broadcast Center, a 4G & 5G network and various mobile phones. The access to the chatbot by a URL link in the public warning text message seems to be easy and fast for users.

Moreover, the second version of the chatbot prototype has been evaluated by the ENGAGE consortium and people working in the field of disaster risk management during an exercise in Romania in May 2023. Respondents were able to test the chatbot from their mobile phones and give feedback afterwards in a questionnaire. We discussed the results of the questionnaire and gathered more useful feedback with the respondents during a group conversation. This feedback gave us insight into which features were crucial to the chatbot, what kind of information is useful to a citizen, and how to proceed development.

A short summary from these results regarding what they liked the most: it is a user-friendly interface, it provides a map which allows users to adequately plan their next moves in a crisis situation, it can provide multiple languages and the accessibility is easy from the smartphone. Summarizing what respondents disliked the most about the chatbot prototype: the chatbot can only provide limited information and incomplete answers, it gives a lot of additional information and not a direct answer to the user's question, and it is not inclusive enough for all groups in society.

These results have led us to work on providing additional information by setting up a database in the future where more information on hazard type will be stored and consulted in case the CAP message does not suffice. Moreover, we will also work on location-based information by calculating the distance of a user to the emergency, and possibly near shelters and provide other useful information. Figure 6 shows screenshots from the prototype.

Figure 6. Screenshots of the warning-systems' prototype.



[13] https://koenvogel.net/wp-content/uploads/2023/07/Internship_Report_Everbridge.pdf

The research leading to these results has received funding from Horizon 2020, the European Union's a Programme for Research and Innovation (H2020/2014-2020) under grant agreement n° 882850.

8.5 Conclusions regarding future AI-enabled chatbots

The chatbot prototype has not been part of a wide scale test, and therefore more testing needs to be done under the population for further development on the chatbot and it's usability.

One key finding is that the provision of accurate and useful data is crucial for AI-enabled chatbots, leaving no room for any deficiencies in this aspect. With the rapid advancement of highly sophisticated chatbots such as ChatGPT, users expect similar features and information delivery from AI-enabled chatbots. Therefore, the software of future AI-enabled chatbots must remain up-to-date and flexible, incorporating efficiently new AI engines or relying on powerful and capable language models.

Another crucial finding is that users prefer concise answers that are neither overly general nor excessively detailed, a requirement that the current chatbot failed to meet adequately. Respondents become irritated when presented with large blocks of information that are imprecise and lack precision. Therefore, it is essential to ensure that the chatbot delivers concise responses that are both accurate and to the point.

Moreover, the use case of a chatbot during emergencies seems to be highly dependent on the cultural context and background of users. Emergencies regarded in one country, are a regular occurrence in another. Also, it was mentioned by some respondents that the use of such a chatbot would be more effective in extended crises, such as earthquakes and floods instead of short duration crisis events.

The research leading to these results has received funding from Horizon 2020, the European Union's a Programme for Research and Innovation (H2020/2014-2020) under grant agreement n° 882850.

9 Additional SML, VR and AR solutions

While previous sections reviewed in general the literature about these solutions, in this part we report about the systematic collection of solutions.

9.1 SML solutions

As mentioned before, in today's digital age, social media platforms have become indispensable tools for communication, especially during emergencies. Social media listening, or monitoring, refers to the process of identifying and analyzing conversations on these platforms to gather insights and act on them. For emergency organizations, this means understanding public sentiment, tracking the spread of information (or misinformation), and gauging the effectiveness of their communication strategies.

Our collection of social media listening solutions yielded seven types of uses and solutions, both for improving the interactions with the citizens, and also used by emergency organizations to improve the management and information gathering from SML: Comprehensive social media management, brand monitoring and insights, real-time monitoring, social media management with AI, conversation analysis, and social media management with scheduling. Table 6 presents the list of SML tools, divided by their type.

Table 6. The list of SML tools, divided by their type.

Type of tools	Examples	Aims & Description	Relevancy for emergencies
Comprehensive Social Media Management	Hotsuite, Sprout Social	Platforms for scheduling posts, interacting with followers, and analyzing social media performance. Offer analytics and the ability to monitor multiple social media accounts.	Provide real-time communication channels and analytics to gauge public sentiment during emergencies. Enables efficient tracking of emergency-related conversations across platforms.
Brand Monitoring and Insights	Brandwatch, Crimson Hexagon, Meltwater	Give insights about online brand presence and sentiments on social media. Analyze audiences and tracking brand perceptions.	Monitor public sentiment and reactions to emergency announcements or situations. Use AI to gauge public reactions and sentiments during crises. Keeps emergency agencies informed about public discussions and potential misinformation.
Real-time Monitoring	TweetDeck, Mention	Monitors mentions across the web and social media in real-time.	Offer real-time tracking of emergency-related social media content and hashtags.
Social Media Management with AI	SocialBakers	Use AI to optimize content, identify influencers, and provide ad intelligence.	Leverage AI to optimize emergency communications and identify key influencers during crises.
Conversation Analysis	TalkWalker	Provide insights into conversations happening around a brand or topic.	Analyze public conversations to understand sentiments and concerns during an emergency.
Social Media Management with Scheduling	AgoraPulse	Scheduling posts, managing social inboxes, and analyzing performance.	Ensure timely communication and track public engagement during crises.

9.1.1 Comprehensive Social Media Management

Comprehensive Social Media Management tools, like Hootsuite and Sprout Social, are platforms designed to streamline the process of managing multiple social media accounts. They offer features such as scheduling posts, monitoring mentions and interactions, and providing analytics on social media performance. These tools are essential for brands, businesses, and organizations aiming to maintain a consistent and effective online presence across various platforms.

In the realm of disaster management and emergencies, these tools are invaluable. They allow emergency agencies to schedule timely alerts, monitor public sentiment in real-time, and gauge the effectiveness of their communication strategies. For instance, during an impending hurricane, agencies can schedule a series of warnings, monitor public reactions, and adjust their communication based on the feedback received. This ensures that the public is well-informed, prepared, and that any misconceptions or rumors are promptly addressed.

A notable case study is the use of Hootsuite by government agencies during Hurricane Sandy in 2012 (Steward & Wilson, 2016). The platform was used to disseminate real-time information, monitor public concerns, and coordinate response efforts. By centralizing their social media communication, agencies were able to provide consistent updates, address queries, and manage the flow of information effectively, ensuring that residents had the most up-to-date and accurate information during the crisis.

9.1.2 Brand Monitoring and Insights

Brand Monitoring and Insights tools, such as Brandwatch, Crimson Hexagon, and Meltwater, are designed to provide deep insights into how a brand or topic is perceived online. They analyze mentions, sentiments, and trends across various online platforms, from social media to blogs and news sites. These tools are crucial for businesses and organizations aiming to understand their online reputation and adjust their strategies accordingly.

When it comes to disaster management, understanding public sentiment is crucial. These tools can help emergency agencies gauge public reactions to their announcements, measure the effectiveness of their communication campaigns, and identify potential areas of concern or misinformation. For example, during a public health crisis, agencies can monitor sentiments around vaccination campaigns, identify sources of misinformation, and tailor their communication to address public concerns effectively.

For example, during the COVID-19 pandemic, Brandwatch was used by health organizations to monitor public sentiments around lockdown measures, mask mandates, and vaccination campaigns (Al-Ramahi & Elnoshokaty, 2021; Berdahl, Fischer & Marcelino, 2022; Mogo et al., 2022). By analyzing millions of online conversations, the tool provided insights into public concerns, misconceptions, and the effectiveness of various communication strategies. This data-driven approach allowed health agencies to adjust their campaigns, address public concerns, and combat misinformation more effectively.

9.1.3 Real-time Monitoring

Real-time Monitoring tools, like TweetDeck and Mention, are specifically designed to provide instant notifications and insights into online mentions and conversations. They allow users to track specific keywords, hashtags, or mentions across various platforms and respond in real-time. These tools are essential for brands and organizations that need to stay on top of online conversations and respond promptly.

The research leading to these results has received funding from Horizon 2020, the European Union's a Programme for Research and Innovation (H2020/2014-2020) under grant agreement n° 882850.

In emergency situations, the speed of response can make a significant difference. Real-time monitoring tools enable emergency agencies to instantly capture public reactions, concerns, and queries related to an ongoing crisis. Whether it's a natural disaster, a terrorist attack, or a public health emergency, these tools ensure that agencies are always in the loop and can address public concerns promptly.

For instance, during the Australian bushfires in 2019-2020, Mention was used by local agencies to track real-time conversations and concerns from residents (Daume, Galaz & Bjerser, 2023). This allowed them to address rumors, provide timely updates on evacuation procedures, and ensure that residents in affected areas had the most recent and accurate information. The ability to monitor and respond in real-time was crucial in managing public sentiments and ensuring effective communication during the crisis.

9.1.4 Social Media Management with AI

Social Media Management tools with AI capabilities, such as SocialBakers, leverage artificial intelligence to optimize content, identify influencers, and provide ad intelligence. These platforms not only streamline the process of managing social media but also use AI to predict trends, analyze sentiments, and optimize content delivery. They represent the next generation of social media tools, combining traditional management features with the power of machine learning.

In the context of disaster management, AI-driven tools can be game-changers. They can predict public reactions based on past data, optimize the delivery of emergency alerts to ensure maximum reach, and even identify key influencers in the community who can aid in disseminating crucial information. For instance, during a flood alert, these tools can analyze which content formats (videos, infographics, text updates) are most engaging for the audience and adjust the communication strategy in real-time.

A practical application of this can be seen in how SocialBakers was utilized during the California wildfires (Guawong & Jankananon, 2015). The platform's AI capabilities helped local agencies identify key community influencers who were then collaborated with to spread safety information and evacuation procedures. By targeting these influencers, the reach of the emergency alerts was significantly amplified, ensuring that a larger portion of the community received timely and crucial updates.

9.1.5 Conversation Analysis

Conversation Analysis tools, like Talkwalker, delve deep into online conversations to provide insights and analytics. They don't just monitor mentions; they analyze them to understand sentiments, trends, and the broader context. These tools can dissect vast amounts of online chatter to extract meaningful insights, making them invaluable for brands and organizations aiming to truly understand their audience.

For emergency management, understanding the nuances of public conversation is vital. It's not just about knowing how many people are talking about an emergency, but understanding their concerns, fears, and misconceptions. For example, during a disease outbreak, these tools can help health agencies understand public sentiments around vaccination, identify prevalent myths, and tailor communication strategies to address these specific issues.

During the Zika virus outbreak, Talkwalker was employed to analyze public sentiments around the disease (Hoernke et al., 2021). The tool helped health agencies identify prevalent myths (like misconceptions about transmission) and adjust their communication campaigns to address these

The research leading to these results has received funding from Horizon 2020, the European Union's a Programme for Research and Innovation (H2020/2014-2020) under grant agreement n° 882850.

specific concerns. By understanding the nuances of the public conversation, agencies were better equipped to address fears and ensure that accurate information was being disseminated.

9.1.6 Social Media Management with Scheduling

Social Media Management tools with Scheduling capabilities, such as AgoraPulse, are designed to automate and streamline the process of content delivery. They allow agencies to plan, schedule, and automate posts across multiple platforms, ensuring consistent communication. These tools are crucial for organizations that need to maintain a regular online presence without the constant manual effort.

In the sphere of emergencies, timely and consistent communication is paramount. These tools ensure that emergency agencies can schedule alerts, reminders, and updates to go out at optimal times, ensuring maximum reach. For instance, in the lead-up to a hurricane, agencies can schedule a series of reminders about evacuation procedures, ensuring that residents are consistently reminded and informed.

The research leading to these results has received funding from Horizon 2020, the European Union's a Programme for Research and Innovation (H2020/2014-2020) under grant agreement n° 882850.

9.1.7 Summary

Table 7 examines all the tools presented here, in the criteria developed in D2.2, for analyzing formal solutions.

Table 7. The analysis of SML solutions according to the criteria of D2.2.

Type of tools	Tools	Improve top-down communication	Improve bottom-up communication	Enhance risk awareness	Facilitate resource allocation	Enhance preparedness	Capitalize social networks & relationships	Improve health and mental outlook	Empower in governance and leadership	Promote efficient response	Facilitate quick recovery	Before the emergency	During the emergency	After the emergency
Comprehensive Social Media Management	Hootsuite Sprout Social	Yes	Yes	Yes	Potential	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes
Brand Monitoring and Insights	Brandwatch Crimson Hexagon Meltwater	Yes	Yes	Yes	No	Yes	Yes	No	Yes	Potential	Potential	Yes	Yes	Potential
Real-time Monitoring	TweetDeck Mention	Yes	Yes	Yes	No	Yes	Yes	No	Potential	Yes	Potential	Yes	Yes	Yes
Social Media Management with AI	SocialBakers	Yes	Yes	Yes	Potential	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes
Conversation 0Analysis	TalkWalker	Yes	Yes	Yes	No	Yes	Yes	No	Yes	Yes	Potential	Yes	Yes	Yes
Social Media Management with Scheduling	Agora Pulse	Yes	Yes	Yes	No	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes

The research leading to these results has received funding from Horizon 2020, the European Union's a Programme for Research and Innovation (H2020/2014-2020) under grant agreement n° 882850.

The table provides a comprehensive overview of the diverse tools available, ranging from comprehensive social media management platforms like Hootsuite to AI-driven solutions like SocialBakers. A consistent trend across these tools is their ability to enhance top-down communication, ensuring that authoritative and accurate information reaches the public swiftly. Many of these platforms, especially those with real-time monitoring capabilities, also excel in facilitating bottom-up communication, capturing public sentiment, concerns, and feedback during crises. Tools with AI capabilities, such as SocialBakers, offer predictive insights, optimizing content delivery based on past data and ensuring maximum engagement. Furthermore, conversation analysis tools like Talkwalker delve deep into online discussions, providing nuanced insights into public sentiments, fears, and misconceptions. This enables agencies to tailor their communication strategies more effectively. In terms of preparedness and response, scheduling tools like AgoraPulse ensure timely and consistent dissemination of alerts and updates. Overall, the integration of these tools can significantly enhance an agency's ability to communicate during emergencies, fostering better public understanding, preparedness, and trust.

9.2 VR AND AR SOLUTIONS

Virtual Reality (VR) and Augmented Reality (AR) technologies have emerged as transformative tools in the realm of emergency response and management. These technologies offer a unique blend of immersive experiences and real-time data overlay, enabling emergency organizations to communicate more effectively with the public, especially during critical situations. For instance, during natural disasters, VR can simulate the aftermath, allowing individuals to understand the severity and take necessary precautions. AR, on the other hand, can overlay evacuation routes on a user's smartphone, guiding them to safety. Such applications not only bridge the gap in collapsed communication channels during surges but also play a pivotal role in neutralizing the spread of false information.

Several real-world examples underscore the efficacy of these technologies. In Japan, VR simulations have been used to train citizens on earthquake preparedness, resulting in heightened awareness and faster response times during actual events (Nakanishi, Ono, Kawano & Sakashita, 2023). Similarly, AR apps have been developed to show floodwater levels in real-time, assisting residents in making informed decisions about evacuation (de Andrade, Padilha, Vahldick & de Oliveira, 2022). These tools, backed by accurate data, have significantly reduced panic and the spread of misinformation during emergencies.

The mapping of VR and AR solutions for crisis communication, yielded 19 tools, with actual uses or potential uses before, during and after emergencies and disasters. Table 8 divides them into their different categories.

Table 8. The list of VR and AR tools divided for crisis communication, divided by their type.

Type of tools	Examples	Aims & Description	Relevancy for emergencies
High-end AR/VR Headsets	Vive Pro Eye, Magic Leap One, Microsoft HoloLens	High-quality immersive experiences	Provides real-time, immersive simulations for emergency training and situational awareness.
Platform Development Kits	ARKit by Apple, Google ARCore, A-Frame	Development platforms for AR/VR applications	Enables the creation of tailored emergency response applications for various platforms.
Social VR Platforms	VRChat, AltspaceVR, Rec Room	Virtual spaces for social interaction	Facilitates virtual community gatherings and discussions during emergencies.
AR Tools	| Spark AR, Zapper, Blippar	Augmented reality experiences typically for mobile	Enhances real-world views with critical information overlays during disasters.

The research leading to these results has received funding from Horizon 2020, the European Union's a Programme for Research and Innovation (H2020/2014-2020) under grant agreement n° 882850.

Type of tools	Examples	Aims & Description	Relevancy for emergencies
VR Platforms for Meetings	Spatial, Glue VR, MeetinVR, Mozilla Hubs	Virtual meeting spaces for collaboration	Provides a virtual space for emergency teams to coordinate and strategize.
VR Headsets	Oculus	Immersive VR experiences	Offers immersive training and awareness programs related to emergencies.
Other	fly 4k Camera, Rapid VR	Various purposes (e.g., capturing VR content)	Enables capturing and sharing of on-ground VR content for situational awareness.

9.2.1 High-end AR/VR Headsets

High-end AR/VR headsets, such as the Vive Pro Eye, Magic Leap One, and Microsoft HoloLens, are designed to provide immersive experiences that blur the lines between the physical and digital worlds. These devices are equipped with advanced sensors, high-resolution displays, and spatial audio capabilities, enabling users to interact with virtual objects and environments in a lifelike manner. The primary aim of these headsets is to offer a high-quality, immersive experience, whether it's for gaming, professional applications, or social interactions.

In the context of emergencies and building societal resilience, high-end AR/VR headsets can play a pivotal role. They can be used for training first responders in simulated emergency scenarios, allowing them to practice and hone their skills in a risk-free environment. Additionally, these headsets can be employed for public awareness campaigns, enabling community members to experience potential emergency situations virtually, thereby fostering a better understanding and preparedness. The immersive nature of these devices can also facilitate better communication among community members, as they can virtually "walk through" emergency evacuation routes, understand the locations of safe zones, and even participate in virtual town hall meetings to discuss emergency preparedness strategies.

While there are numerous applications of AR/VR in various sectors, one relevant example in the context of emergencies is the use of the Microsoft HoloLens in medical training (Palumbo, 2022). Medical professionals can use the HoloLens to simulate emergency medical scenarios, allowing them to practice procedures and make decisions in a controlled environment. This kind of training can be invaluable in preparing medical teams for real-world emergencies, ensuring they have the skills and knowledge needed to respond effectively. Another potential application, as mentioned in the scholarly results, is the use of AR in remote medical care, where augmented reality can enhance the ability to gather observational and medical information, serving diverse applications in remote health care delivery and education.

9.2.2 Platform development kits

Platform Development Kits (PDKs) for AR and VR, such as ARKit by Apple, Google ARCore, and A-Frame, serve as foundational tools for developers to create augmented and virtual reality applications. These platforms offer a suite of tools, libraries, and frameworks that simplify the development process, allowing for the creation of immersive experiences without the need to build everything from scratch. They often come with features like spatial tracking, environmental understanding, and user interaction capabilities, enabling developers to focus on the content and functionality of their applications.

PDKs play a crucial role in fostering communication and societal resilience, especially in the context of emergencies. By leveraging these development kits, applications can be created to simulate

The research leading to these results has received funding from Horizon 2020, the European Union's a Programme for Research and Innovation (H2020/2014-2020) under grant agreement n° 882850.

emergency scenarios, educate the public about potential risks, and provide real-time information during crises. For instance, an AR application could overlay evacuation routes on a user's real-world view during a natural disaster, guiding them to safety. Similarly, VR applications can be used for community training sessions, allowing residents to virtually experience and prepare for potential emergencies. The interactive and immersive nature of AR and VR applications developed using PDKs can enhance community engagement, awareness, and preparedness, ultimately contributing to societal resilience.

One relevant example in the domain of emergencies using A-Frame (Morgado, 2023). This platform utilized advanced technologies to decrease the communication time of alarm locations from an average of 3.1 minutes to a second-level system query. Such rapid communication can be crucial during emergencies, ensuring timely responses and potentially saving lives. By integrating AR/VR capabilities, such platforms can further enhance their effectiveness, providing visual cues and immersive experiences to aid in emergency response and management.

9.2.3 Social VR Platforms

Social VR platforms, such as VRChat, AltspaceVR, and Rec Room, are virtual spaces designed to foster social interaction in an immersive environment. These platforms allow users to create avatars, interact with others, participate in events, and even design their own virtual worlds. The primary aim is to provide a platform where users can socialize, collaborate, and share experiences in a way that transcends the limitations of physical space.

Social VR platforms have the potential to revolutionize communication, especially in the context of emergencies and societal resilience. These platforms can serve as virtual gathering places where community members can come together to discuss, plan, and coordinate during crises. For instance, in the event of a natural disaster, a virtual town hall could be organized in a social VR space, allowing residents to receive updates, ask questions, and share resources without the need for physical presence. This can be particularly useful when physical movement is restricted. Moreover, these platforms can be used for public awareness campaigns, simulations, and training sessions related to emergency preparedness, enabling community members to virtually experience potential emergency situations and learn how to respond effectively.

One example of possible use is highlighted in the use of social VR platforms for older adults, showcasing how avatar-mediated communication in platforms like AltspaceVR supports introverted users and offered advantages when discussing sensitive topics (Baker et al., 2021). This suggests that in emergency situations, where stress and anxiety levels are high, social VR platforms can provide a safe and supportive environment for individuals to communicate, share their concerns, and seek help. Another example, in the case of VRChat, demonstrates its potential for collaborative experiments, suggesting its utility for community-based emergency planning and simulations (Kelley, 2021). While these studies don't directly address emergency scenarios, they underscore the potential of social VR platforms in fostering communication and collaboration, which are crucial during crises.

9.2.4 AR Tools

Augmented Reality (AR) tools, such as Spark AR, Zapper, and Blippar, are designed to overlay digital information onto the real world, typically through the lens of a smartphone or AR glasses. These tools allow users to interact with virtual objects and information seamlessly integrated into their physical environment. The primary aim of AR tools is to enhance the user's perception of reality by adding contextually relevant digital content, making information more accessible and interactive.

AR tools can significantly enhance communication and societal resilience, especially during emergencies. By providing real-time, contextually relevant information, AR can guide individuals

The research leading to these results has received funding from Horizon 2020, the European Union's a Programme for Research and Innovation (H2020/2014-2020) under grant agreement n° 882850.

during crises, whether it's navigating evacuation routes, identifying safe zones, or accessing vital emergency services. For instance, in the event of a flood, an AR application could overlay water levels on streets, helping residents understand the severity and navigate safely. Furthermore, AR tools can be used for public awareness campaigns, allowing community members to visualize potential threats in their environment, fostering better understanding and preparedness. The interactive nature of AR can also facilitate community engagement, enabling residents to access and share information easily, ultimately contributing to societal resilience.

One example of use is the development of a framework for the affordances offered by AR-enabled geographic information systems (GIS) tools and their impact on the effectiveness of disaster response (Yang, Kim, Lee, Hwang & Choi, 2022). This underscores the potential of AR tools in providing real-time spatial information during emergencies, aiding in more efficient and effective disaster response. Another example is of perceptions about Augmented Reality in remote medical care. For example, by enhancing the ability to gather observational and medical information by AR, especially in remote health care delivery and education. This suggests that in emergency situations, AR tools can play a pivotal role in facilitating remote medical consultations and assessments, ensuring timely and effective care.

9.2.5 VR Platforms for Meetings

VR platforms for meetings, such as Zoom VR, Spatial, and ENGAGE, are designed to facilitate virtual interactions in a more immersive environment compared to traditional video conferencing. These platforms allow users to create avatars, participate in virtual meetings, and interact with 3D content, offering a more engaging and collaborative experience. The primary aim of these platforms is to bridge the gap between physical meetings and traditional online meetings, providing a space where users can feel more present and connected with their colleagues or peers.

In the context of emergencies and societal resilience, VR platforms for meetings can serve as crucial tools for communication and coordination. When physical gatherings are restricted or unsafe, such as during natural disasters or pandemics, these platforms can provide a space for emergency response teams, community leaders, and residents to come together, discuss, and plan. The immersive nature of VR can enhance the sense of presence, making discussions more engaging and ensuring that crucial information is effectively communicated. For instance, community leaders can use these platforms to hold virtual town halls, updating residents about ongoing emergency situations, evacuation plans, or recovery efforts.

During the COVID-19 pandemic, online meetings became a norm for daily teamwork across various sectors. A study by Bonfert et al. (2023), explored the opportunities and challenges of meeting in virtual reality compared to traditional videoconferences. The study was conducted over 4 months, where weekly team meetings of a human-computer interaction research lab were held on various online meeting platforms. The findings revealed that VR meetings offered a unique spatial aspect, improved meeting atmosphere, and enhanced expression of emotions compared to traditional video calls. This suggests that in emergency situations, where face-to-face interactions might be limited, VR platforms can provide a more effective and engaging medium for communication, ensuring that teams remain coordinated and informed.

9.2.6 Other uses

Drones equipped with high-resolution cameras, like the fly 4k Camera, and other VR/AR tools, such as Rapid VR, are revolutionizing the way we capture and experience content. These tools allow for the acquisition of aerial footage, 360-degree videos, and immersive content that can be integrated

The research leading to these results has received funding from Horizon 2020, the European Union's a Programme for Research and Innovation (H2020/2014-2020) under grant agreement n° 882850.

into virtual and augmented reality platforms. Drones, in particular, offer a unique vantage point, capturing expansive landscapes, inaccessible terrains, and providing real-time aerial surveillance.

Drones and cameras for VR/AR can play a pivotal role in emergency communication and building societal resilience. In the event of natural disasters, drones can be deployed to assess damage, locate stranded individuals, and provide real-time data to emergency response teams. The footage captured can be integrated into VR/AR platforms to create immersive simulations for training, awareness campaigns, and public briefings. For instance, after a flood or earthquake, drones can capture the extent of the damage, and this data can be used to create VR simulations to train rescue teams or inform the public about affected areas and safe routes.

A study by Stanga, Banfi & Roascio (2023), showcases the potential of drones in capturing detailed aerial imagery. In this study, UAV (Unmanned Aerial Vehicle) photogrammetry was employed to capture detailed images of the Claudius Anio Novus aqueduct in Tor Fiscale Park, Rome. The data acquired was then processed to generate orthophotos, drawings, and historic building information modeling (HBIM) of the aqueduct. This information was further used to create a virtual reality experience, allowing users to explore and interact with the archaeological site. Such applications underscore the potential of drones and VR/AR tools in preserving cultural heritage, educating the public, and preparing for potential threats or emergencies.

The research leading to these results has received funding from Horizon 2020, the European Union's a Programme for Research and Innovation (H2020/2014-2020) under grant agreement n° 882850.

9.2.7 Summary

Table 9 examines all the tools presented here, in the criteria developed in D2.2, for analyzing formal solutions.

Table 9. The analysis of the VR and AR solutions according to the criteria of D2.2.

Type of tools	Tools	Improve top-down communication	Improve bottom-up communication	Enhance risk awareness	Facilitate resource allocation	Enhance preparedness	Capitalize social networks	Improve health and mental outlook	Empower in governance and ...	Promote efficient response	Facilitate quick recovery	Before the emergency	During the emergency	After the emergency
High-end AR/VR Headsets	Vive Pro Eye Magic Leap One Microsoft HoloLens	Yes	Potential	Yes	Potential	Yes	No	Potential	Potential	Yes	Potential	Yes	Yes	Potential
Platform Development Kits	ARKit by Apple Google ARCore A-Frame	Potential	Potential	Yes	No	Yes	No	No	No	Potential	No	Yes	Potential	Potential
Social VR Platforms	VRChat AltspaceVR Rec Room	No	Yes	No	No	No	Yes	Yes	No	No	No	Potential	Yes	Yes
AR Tools	Spark AR Zapper Blippar	Potential	Yes	Yes	No	Yes	Yes	No	No	Potential	No	Yes	Yes	Potential
VR Platforms for Meetings	Spatial Glue VR MeetinVR Mozilla Hubs	Yes	Yes	No	Yes	No	Yes	No	Yes	Yes	No	Potential	Yes	Potential
VR Headsets	Oculus	Yes	Potential	Yes	Potential	Yes	No	Potential	Potential	Yes	Potential	Yes	Yes	Yes
Other	fly 4k Camera Rapid VR	No	No	Yes	No	No	No	No	No	No	No	Yes	Yes	Yes

The research leading to these results has received funding from Horizon 2020, the European Union's a Programme for Research and Innovation (H2020/2014-2020) under grant agreement n° 882850.

In the table above, yes means that the tools are used to answer this goal, described in the column. Potential means that they have this potential, but we did not document such a use. No means that they cannot be used for this goal.

The table provides a comprehensive analysis of various AR/VR tools and platforms, evaluating their potential contributions to different facets of disaster communication. High-end AR/VR headsets, such as the Vive Pro Eye, Magic Leap One, and Microsoft HoloLens, emerge as pivotal assets, offering real-time immersive simulations that can significantly enhance top-down communication, risk awareness, and preparedness. Platform Development Kits, including ARKit by Apple and Google ARCore, serve as foundational tools for creating tailored emergency response applications, emphasizing risk awareness and pre-emergency preparedness. Social VR platforms, like VRChat and AltspaceVR, uniquely capitalize on social networks and relationships, fostering bottom-up communication and providing emotional support during crises. AR tools, including Spark AR and Zapper, enhance real-world views with critical information, aiding in both top-down and bottom-up communication. VR platforms designed for meetings, such as Spatial and Glue VR, facilitate collaborative strategies, emphasizing efficient response during emergencies. The Oculus VR headset stands out for its immersive experiences, enhancing risk awareness and preparedness. Lastly, tools categorized under "Other," like the fly 4k Camera, primarily focus on capturing real-time content for situational awareness. In essence, while each tool offers distinct capabilities, their combined potential can revolutionize disaster communication, addressing various needs from preparedness to recovery.

The research leading to these results has received funding from Horizon 2020, the European Union's a Programme for Research and Innovation (H2020/2014-2020) under grant agreement n° 882850.

10 TRANSFERRING RESEARCH RESULTS TO THE PUBLIC

The true value of findings often lies not just in their discovery but in their dissemination. Especially when it comes to topics that have a direct impact on the public, such as the development of a blueprint for an AI-enabled chatbot used in emergency communication, the prototype which is based on it, and the integration of Social Media Listening (SML), Virtual Reality (VR), and Augmented Reality (AR) solutions in emergency communication. The importance of ensuring that these findings are communicated effectively to the general public cannot be overstated.

In order to allow a true impact on the field of societal resilience, we outline our strategies for effectively communicating our research findings to the public. The goal is to ensure that our work does not remain confined within inner circles, but reaches the people who can benefit from it the most - the general public, directly and through emergency authorities and organizations. We discuss in this section various methods of public engagement. We aim to present our findings in a clear, accessible manner, helping to increase public understanding of the issues at hand and promoting informed decision-making.

10.1 Strategies for public engagement

Based on the dissemination plan, which is portrayed in the deliverables of WP5, we suggest the following strategies for transferring the research results to the public: public workshops and seminars, interactive online platforms, collaboration with the media, educational outreach and feedback mechanisms.

10.1.1 Public workshops and seminars

Public workshops and seminars refer to organizing events where the public and emergency professionals can directly interact with experts, ask questions, and get a firsthand understanding of the findings. This aligns with the exploitation activities outlined in WP5, such as workshops with academia, end-user partners, and industry/SME partners.

The development process of the blueprint and the eDelphi results can be presented in a workshop format, with members of the public and emergency professionals, both are different types of end-users to the suggested solution. The emergency professionals, as representatives of the organizations that will implement the solution and the general public as those who will benefit from its affordances. Attendees can be walked through the stages of the blueprint's creation, the feedback loops, and the iterative improvements made based on expert opinions. A seminar can be dedicated to the prototype, allowing attendees to interact with it and understand its functionalities, raise potential barriers and suggest how these models can be tailored to their types of organization.

This direct interaction demystifies the research process and allows for immediate feedback, questions, and clarifications.

10.1.2 Interactive online platforms

Interactive online platforms involve creating web platforms where the research is presented in interactive formats, allowing users to engage with the data and understand its implications. The ENGAGE website and the knowledge platform, including the catalogue of solutions, are several examples for it. This is in line with the continuous review and update of communications and dissemination actions to support exploitation as mentioned in WP5.

The research leading to these results has received funding from Horizon 2020, the European Union's a Programme for Research and Innovation (H2020/2014-2020) under grant agreement n° 882850.

The collection of SML, VR, and AR solutions, including the blueprint of the AI-enabled chatbot and the prototype example, are part of the catalogue of solutions and the knowledge platform. The revised model, for instance, can be presented as an interactive flowchart or diagram, allowing users to click on different components for detailed explanations.

Such platforms cater to a global audience and can be accessed at convenience, ensuring wider dissemination of research findings.

10.1.3 Collaborations with the media

Collaborations with the media include working with media outlets to present the findings in layman's terms, ensuring a wider reach.

Media outlets will be provided with press releases detailing the key findings of the research, especially the significance of the blueprint in developing AI-enabled chatbots as part of the strategy of disaster management. Interviews with key researchers can be organized, where they discuss the importance of neutralizing false information during emergencies and the role of the prototype in achieving this.

Media has the power to reach a vast audience quickly, making it a crucial tool for public engagement.

10.1.4 Educational outreach

Educational outreach takes into consideration collaboration with educational institutions to integrate the findings into curricula, ensuring that the next generation is well-informed. This mirrors WP5's focus on workshops with academia partners and the continuous recruitment of KI-CoP members to facilitate exploitation.

The research's findings, especially the importance of SML, VR, AR, and AI-enabled chatbots in disaster management, can be integrated into courses related to public administration, disaster management, and communication. Workshops can be conducted in educational institutions, where students are introduced to the prototype and its functionalities.

This ensures that the next generation of professionals is equipped with the latest knowledge and tools in the field.

10.1.5 Feedback mechanisms

Feedback mechanisms allow to establish channels where the public and emergency professionals can provide feedback on the research, which can be invaluable for further studies.

After presenting the revised model and the prototype, feedback forms can be distributed among the emergency professionals and experts. These forms can gather opinions on the model's comprehensiveness, the prototype's user-friendliness, and suggestions for improvement. Online platforms showcasing the SML, VR, and AR solutions can have integrated feedback sections for users to share their experiences and insights.

Feedback provides a direct line to the end-users, ensuring that the research remains aligned with public needs and expectations.

The research leading to these results has received funding from Horizon 2020, the European Union's a Programme for Research and Innovation (H2020/2014-2020) under grant agreement n° 882850.

10.2 Strategies for public engagement – Challenges and opportunities

The integration of AI-enabled chatbots, SML, VR, and AR solutions in the context of emergency communication presents a unique set of challenges and opportunities. For instance, during times of crisis, communication channels can become overwhelmed, leading to a breakdown in the dissemination of critical information. SML tools can help authorities monitor public sentiment in real-time, allowing them to address concerns and counteract misinformation promptly. VR and AR, on the other hand, offer immersive experiences that can be used for public awareness campaigns, training, and even during actual emergencies to guide the public.

However, while these technologies offer immense potential, their effective integration requires a deep understanding of public perceptions, cultural nuances, and technological challenges. For instance, in some cultures, there might be a mistrust of AI-driven tools, while in others, there might be challenges related to technology access. By ensuring that our research findings are widely disseminated and understood, we can work towards a future where these tools are integrated seamlessly into emergency communication strategies, leading to safer, more resilient communities.

By integrating the research findings into these public engagement strategies, which are part of the dissemination plan of WP5, we ensure that the knowledge is not just disseminated but is also understood, appreciated, and applied in real-world scenarios.

10.3 AI-enabled tools for disseminating scientific knowledge

AI chatbots, like the blueprint developed in this project, present a unique opportunity to disseminate scientific knowledge related to risks or prevention measures. For instance, in the context of a pandemic like COVID-19, an AI chatbot can be programmed to provide accurate information about viruses, barriers, and vaccination. This ensures that the public receives timely and accurate information, reducing the spread of misinformation.

To effectively transfer research results to citizens, the AI chatbot can be integrated with popular communication channels such as social media. This allows for a wider reach, ensuring that critical information is accessible to a broader audience. By leveraging the chatbot's capabilities, research findings can be translated into bite-sized, easily digestible content that resonates with the general public.

Moreover, the AI chatbot can be tailored to convey the project's solutions to specific groups within the population, based on their unique needs and expectations. This personalized approach ensures that the information is relevant and actionable, leading to better outcomes in terms of public safety and resilience.

Training the chatbot to achieve this requires a multi-faceted approach. First, the chatbot should be fed with comprehensive and up-to-date research data. This data should then be processed and translated into simpler terms, ensuring that complex scientific jargon is converted into everyday language. For example, based on other texts or items which are known as successful "translators" of research findings to the public. Regular feedback loops with diverse focus groups can help in refining the chatbot's responses, ensuring that the information is presented in a manner that's easily understood by the public. Additionally, the chatbot can be trained to recognize and adapt to various cultural nuances and linguistic preferences, ensuring inclusivity and broader acceptance.

However, it's essential to recognize the challenges associated with using AI chatbots for this purpose. Cultural nuances, public perceptions, and technological barriers can impact the effectiveness of the chatbot. For instance, in some cultures, there might be a mistrust of AI-driven tools, while in others, there might be challenges related to technology access. By understanding these challenges and addressing them proactively, we can ensure that the AI chatbot becomes a valuable tool for public education and engagement.

The research leading to these results has received funding from Horizon 2020, the European Union's a Programme for Research and Innovation (H2020/2014-2020) under grant agreement n° 882850.

Additionally, by leveraging SML, researchers can identify common misconceptions or knowledge gaps among the public. Once identified, targeted information campaigns, whether through AI chatbots or in other channels, can be launched to address these gaps. Furthermore, SML provides an avenue to understand the language and terminologies the public uses, ensuring that research results are communicated in relatable and understandable terms. By actively engaging on social media platforms and responding to public queries or concerns, the research community can foster trust and ensure that scientific findings are both accessible and accepted by the broader population.

Moreover, VR and AR technologies can also be used to transfer research results to the public in more engaging and interactive ways. For instance, VR can be used to create simulations that allow the public to visualize and interact with complex research data, making it easier to comprehend. Imagine a VR experience where users can "walk through" a simulation of a pandemic's spread, understanding the research-backed measures that can mitigate its impact.

Regarding AR, in relevant apps on smartphones, users can point their devices at specific locations or objects and receive real-time research-based information. This could be particularly useful in scenarios like understanding the structural integrity of buildings post an earthquake or visualizing the spread of a disease in a particular area.

Both VR and AR can be integrated with AI chatbots to provide real-time answers to user queries, enhancing the interactive experience. However, it's crucial to ensure that the content delivered through VR and AR is not just technologically impressive but also scientifically accurate and easily comprehensible. Collaborations with educators, communication experts, and technologists can ensure that these tools are effectively utilized to bridge the gap between complex research findings and public understanding.

The research leading to these results has received funding from Horizon 2020, the European Union's a Programme for Research and Innovation (H2020/2014-2020) under grant agreement n° 882850.

11 DISCUSSION

Deliverable 3.4, presented in this report, is a culmination of extensive research efforts to understand and improve the design and implementation of communication and social media technologies. The insights drawn from the research have been instrumental in shaping the blueprint revisions and the prototype that emerged from it, along with the collection of SML, VR and AR additional solutions with the transference strategies of the results to the public.

The transition from D3.2 to D3.4 represents a significant evolution in our understanding and approach to AI-enabled communication technologies for emergencies. D3.2 laid the groundwork, introducing the foundational elements and initial blueprints for the system. It provided a starting point, a preliminary framework that captured the essential components and strategies. However, as with any pioneering endeavor, there was room for refinement and enhancement. D3.4, building upon the insights, feedback, and advancements since D3.2, offers a more comprehensive and nuanced blueprint. It revises and finalizes the initial concepts, integrating expert opinions, real-world challenges, and the latest technological advancements. This iterative process, from D3.2's inception to D3.4's culmination, underscores the commitment to excellence and the importance of continuous adaptation in the face of an ever-evolving technological landscape. D3.4 stands as a testament to this journey, encapsulating the collective wisdom, research, and innovation that has gone into crafting a state-of-the-art communication model for emergencies.

11.1 ANALYSIS OF THE REVISIONS OF THE BLUEPRINT

The blueprint's evolution is a testament to the iterative nature of design, especially in the realm of emergency communication systems. The feedback loop, involving experts and stakeholders, was instrumental in refining the blueprint to its current form. The consensus criteria, as observed, showed that components like input types and monitoring & reporting were well-received and met the expectations across the board. This is in line with the literature which emphasizes the importance of clear input mechanisms and robust monitoring in emergency communication systems (Ignatidou, 2019; Kertysova, 2018; Annoni, Benczur, Bertoldi, Delipertrev, De Prato, Feijoo & Junklewitz, 2018).

However, the lack of consensus on components like security & privacy, logging & documentation, and statistics is intriguing. It underscores the multifaceted challenges in designing such systems. The literature has always highlighted the paramount importance of security in communication technologies (Chagas et al., 2021; Stieglitz et al., 2022). For instance, in emergency scenarios, the integrity and confidentiality of information can be as crucial as the information itself. A breach can lead to misinformation, panic, or even misuse of sensitive data. The feedback from experts, especially on components like security & privacy, resonates with this sentiment. It's a clear indication that while technological advancements can offer solutions, they also bring forth challenges that need to be addressed proactively.

The original blueprint developed in D3.2 was a foundational step. It provided a structured approach to designing AI-enabled chatbots for emergencies. However, the revisions made post feedback have made it more aligned with real-world challenges and requirements. The literature, especially studies focusing on emergency communication, has always emphasized the need for systems to be dynamic, adaptable, and secure (MacKenzie & Barker, 2013). The revised blueprint, with its emphasis on components like security & privacy and monitoring & reporting, is a reflection of these requirements.

The median scores for most questions being high is a positive sign. It indicates a general agreement among experts on the importance and relevance of the components. This is crucial because, in the realm of emergency communication, consensus among experts can lead to more standardized and effective solutions. The literature has numerous instances where the lack of standardized communication protocols led to inefficiencies during emergencies (Kim, Lyu & Gong, 2020; Puildo, Villarejo-Carballido, 2020).

The revisions made to the blueprint, based on expert feedback, have made it more robust and aligned with the challenges of emergency communication. The alignment with the literature further validates the decisions made during the revision process. The blueprint now stands as a comprehensive guide for organizations looking to leverage AI for emergency communication.

11.2 Analysis of the prototype

The transition from a theoretical blueprint to a tangible prototype is a significant leap in the development process. The prototype, in essence, is the manifestation of the blueprint's principles, embodying its core ideas and strategies, even if in a specific case-study. It serves as a bridge between abstract concepts and real-world applications, allowing for a more hands-on understanding of the system's capabilities and limitations.

Drawing from the literature, the importance of prototyping in system development is well-established (Carenzo, Costantini, Greco, Barra, Rendiniello, Mainetti & Cecconi, 2020). Prototypes act as a medium for testing, validation, and iterative refinement. They provide stakeholders with a tangible system to interact with, facilitating a more concrete understanding and feedback mechanism. In the context of the AI-enabled chatbot for emergencies, the literature emphasizes the critical role of prototypes in understanding user interactions, gauging system responsiveness, and ensuring the system's reliability in real-time emergency scenarios (Hrabi, 2020).

Expert insights, especially those highlighting the importance of the explainability of AI models, were integrated into the prototype's design. This is in line with the literature which underscores the significance of transparency and trust in AI systems, especially in high-stakes scenarios like emergencies. Users and stakeholders need to understand and trust the decisions made by the AI, and the prototype was designed keeping this principle at its core.

Furthermore, the emphasis on an evaluation metric, as suggested by experts, resonates with the literature's focus on the importance of continuous monitoring and assessment of AI systems (Alessi et al., 2018; European Commission, 2018; Kertysova, 2018). An effective evaluation metric not only ensures that the system is performing optimally but also provides a benchmark for future improvements.

The original blueprint developed in D3.2 laid the groundwork for the prototype. It provided a structured approach, emphasizing key components and strategies. However, the revised blueprint, enriched by expert feedback, added layers of depth and nuance to this foundation. The prototype, in turn, was a reflection of this enriched blueprint, incorporating its refined strategies while also addressing the practical challenges identified during the eDelphi rounds.

The prototype serves as a crucial link between the theoretical constructs of the blueprint and the practical demands of real-world emergency communication. Drawing from both the original and revised blueprints, and enriched by expert feedback and literature insights, the prototype stands as a testament to the iterative and collaborative nature of design in the realm of emergency communication technologies.

11.3 Improved design directions for communication and Social Media technologies

The research findings have significantly advanced our understanding of communication and social media technologies, especially in the context of disaster management. The emphasis on "privacy by design" and the clear distinction between privacy and security underscore the importance of

The research leading to these results has received funding from Horizon 2020, the European Union's a Programme for Research and Innovation (H2020/2014-2020) under grant agreement n° 882850.

safeguarding user data while ensuring efficient communication. This aligns with the broader literature which has consistently advocated for a user-centric approach to design (Nadarzynski, Miles, Cowie & Ridge, 2019). Such an approach ensures that technologies are not only efficient but also respect user privacy and security.

Furthermore, the expansion from D3.2 to D3.4 has broadened the scope of our understanding. While D3.2 laid the foundation by introducing the blueprint, D3.4 delves deeper into the intricacies of communication models. The integration of feedback, expert opinions, and real-world challenges has enriched the blueprint, making it more robust and adaptable. This iterative process of refinement, rooted in both empirical data and theoretical frameworks, has been instrumental in shaping a comprehensive communication model.

The literature review section in the provided document further reinforces these findings. It highlights the growing importance of communication technologies in disaster management and the need for a holistic approach that encompasses various facets of communication. The literature suggests that while technology plays a crucial role, the human element – understanding user needs, preferences, and behaviors – is equally important (Lei, Shen & Ye, 2021; Aoki, 2020; Ashfaq, Yun & Loureiro, 2020; Brandzaeg & Følstad, 2018; Laumer, Maier & Gubler, 2019). The integration of these insights into the expanded D3.4 offers a more nuanced and effective communication model, bridging the gap between theory and practice.

11.4 Integration of SML, VR and AR tools

The integration of SML, VR, and AR tools into communication technologies, especially in the context of disaster management, represents a paradigm shift in how information is disseminated and consumed. The literature underscores the transformative potential of these tools (Barojan, 2021; Luccioni, Pham, Lam, Aylett-Bullock & Luengo-Oroz, 2021) . For instance, VR and AR can create immersive environments that simulate real-world disaster scenarios, providing both emergency responders and the public with a tangible sense of the challenges and decisions they might face (Argumedo-García, Salas-Navarro, Acevedo-Chedid, Ospina-Mateus, 2021) . Such experiential learning can significantly enhance preparedness and response strategies.

Moreover, SML tools, with their ability to process vast amounts of data and discern patterns, can be instrumental in real-time decision-making during emergencies. By analyzing social media feeds, news sources, and other digital channels, these tools can provide timely alerts, helping authorities make informed decisions. The literature emphasizes the importance of harnessing the power of SML to predict, monitor, and respond to emergencies more effectively (Krüger & Albris, 2020; Rao, 2016).

While D3.2 introduced the foundational elements, D3.4, enriched by the insights from the literature, emphasizes the integration of cutting-edge tools like SML, VR, and AR. This holistic approach ensures that the communication technologies suggested are not only advanced but also user-centric, catering to the diverse needs of the public during emergencies. The iterative process of refinement, as seen in the transition from D3.2 to D3.4, underscores the importance of continuous learning and adaptation, especially in the ever-evolving landscape of communication technologies.

11.5 Accountability

Accountability, in its essence, refers to the obligation or willingness to accept responsibility or to account for one's actions. In the realm of technology, and more specifically in the domain of Artificial Intelligence (AI), the concept of accountability becomes increasingly complex and multifaceted. As AI systems, including chatbots, become more integrated into our daily lives and critical decision-making processes, the question of who or what is accountable for their actions or decisions becomes paramount.

The research leading to these results has received funding from Horizon 2020, the European Union's a Programme for Research and Innovation (H2020/2014-2020) under grant agreement n° 882850.

In traditional systems, accountability is straightforward. Humans make decisions, and they are held responsible for their outcomes. However, with AI, especially deep learning models, the decision-making process can be opaque. This "black box" nature of AI models means that even experts can't always explain why a particular decision was made. This lack of transparency can be particularly concerning in high-stakes situations, such as emergencies, where understanding the rationale behind decisions can be crucial.

For example, if a human police officer provides incorrect directions during an emergency, they can be questioned, and they can provide an explanation for their decision. In contrast, an AI-driven chatbot might not offer a clear rationale for its guidance. This difference isn't just about transparency but also about the fundamental understanding of decision-making. Humans base decisions on a combination of knowledge, experience, emotions, and even biases. AI, on the other hand, relies on data and algorithms.

The comparison to self-driving cars further underscores the complexity of accountability in AI. If a self-driving car makes a wrong decision leading to an accident, who is responsible? The software developer? The car manufacturer? The owner of the car? Or the AI itself? These questions don't have straightforward answers, but they highlight the need for a robust framework for AI accountability.

In the context of emergencies, where AI-driven decisions can have life-altering consequences, establishing clear lines of accountability is even more critical. While AI offers immense potential in enhancing emergency responses, its integration should be accompanied by rigorous checks and balances. This might include:

- **Transparency Protocols**: Developing AI models that, while complex, have layers of interpretability built-in, allowing for some level of insight into their decision-making.
- **Human-in-the-loop Systems**: Ensuring that critical decisions always have a human oversight, where AI provides recommendations, but humans make the final call.
- **Regulatory Frameworks**: Establishing clear guidelines and regulations that define accountability in scenarios where AI is deployed, ensuring that there's a responsible entity in case of failures.

Therefore, while AI, including chatbots, offers transformative potential in emergency management and other domains, it also introduces new challenges in accountability. Addressing these challenges requires a multi-faceted approach, combining technological advancements with ethical considerations, regulatory frameworks, and public discourse. As AI continues to evolve, so must our understanding and approach to ensuring its responsible and accountable use.

11.6 APPROACHES FOR TRANSFER OF RESEARCH RESULTS TO THE PUBLIC

Transferring research results to the public is a critical aspect of any research endeavor. The feedback and consensus scores from the eDelphi rounds indicate a strong emphasis on transparency, accuracy, and relevance. The public, more than ever, is keen on understanding the intricacies of technologies that influence their daily lives. The literature emphasizes the importance of clear communication, avoiding jargon, and ensuring that the research results are accessible to a broad audience. The feedback from experts on the need for human oversight and the ethical concerns surrounding it further underscores the importance of a transparent and ethical transfer of research results.

The research leading to these results has received funding from Horizon 2020, the European Union's a Programme for Research and Innovation (H2020/2014-2020) under grant agreement n° 882850.

12 Strengths & Limitations

The integration of advanced technologies into disaster management and emergency communication systems offers a plethora of advantages but also comes with its set of challenges. The transition from D3.2 to D3.4 has been instrumental in refining the approach, ensuring that the technologies are not only efficient but also user-centric, respecting user privacy and security.

12.1 Strengths of the considered technologies

12.1.1 AI-enabled chatbots

AI-enabled chatbots, as outlined in the blueprint and exemplified in the prototype, offer real-time, scalable communication solutions. They can handle a vast number of queries simultaneously, ensuring that vital information reaches the public promptly. In the context of emergencies, this immediacy can be life-saving. Furthermore, AI chatbots can be trained to provide accurate and consistent information, reducing the chances of misinformation. Their integration with various data sources ensures that they are always updated with the latest information.

D3.4 further emphasized the importance of AI chatbots by integrating feedback from experts and stakeholders. The iterative process of design and feedback ensured that the chatbot prototype was not only efficient but also user-centric, addressing concerns related to privacy, security, and usability.

12.1.2 SML

Social Media Listening (SML) tools provide authorities with the capability to monitor public sentiment in real-time. This is crucial during emergencies when public sentiment can shift rapidly. By understanding public concerns, authorities can tailor their communication strategies, address misinformation, and ensure that the public remains informed and calm.

D3.4 expanded on the potential of SML tools by highlighting their integration with other communication technologies. The deliverable emphasized the need for a holistic approach, where SML tools work in tandem with other technologies to provide a comprehensive communication solution during emergencies.

12.1.3 VR & AR

Virtual Reality (VR) and Augmented Reality (AR) offer immersive experiences that can be pivotal in emergency training and public awareness campaigns. They allow for the simulation of emergency scenarios, providing both first responders and the public with a risk-free environment to practice and understand emergency protocols. The lifelike simulations can also foster empathy, helping individuals understand the gravity of certain situations.

D3.4 further underscored the importance of VR and AR tools by integrating them into the revised blueprint. The deliverable highlighted potential applications, emphasizing the need for user-centric design and the integration of feedback to ensure that these tools are both innovative and usable.

The research leading to these results has received funding from Horizon 2020, the European Union's a Programme for Research and Innovation (H2020/2014-2020) under grant agreement n° 882850.

12.2 Weaknesses of the considered technologies

12.2.1 AI-enabled chatbots

While AI chatbots offer numerous advantages, they are not without limitations. They are often less preferred when a human alternative is available. Trust in AI chatbots can also be an issue, especially if users have had prior negative experiences. Furthermore, while AI chatbots can provide information, they might not always meet the emotional and psychological needs of users during emergencies.

D3.4 addressed these limitations by emphasizing the need for continuous feedback and iteration. The deliverable highlighted the importance of integrating human expertise with AI capabilities, ensuring that the chatbot remains a reliable and trusted source of information.

12.2.2 SML

SML tools, while powerful, are dependent on the data they receive. Misinformation on social media platforms can skew the insights derived from these tools. Additionally, not all segments of the population might be represented on social media, leading to potential biases in the data.

D3.4 recognized these challenges and emphasized the need for a multi-pronged approach. The deliverable suggested integrating SML tools with other data sources, ensuring that the insights derived are comprehensive and representative.

12.2.3 VR & AR

VR and AR tools require specialized equipment and might not be accessible to everyone. There's also a learning curve associated with these technologies. In high-stress situations, such as emergencies, individuals might find it challenging to navigate VR or AR interfaces.

D3.4 addressed these challenges by emphasizing the need for user-centric design and training. The deliverable highlighted the potential of VR and AR tools when integrated with other communication technologies, ensuring that they remain accessible and usable.

The research leading to these results has received funding from Horizon 2020, the European Union's a Programme for Research and Innovation (H2020/2014-2020) under grant agreement n° 882850.

13 CONCLUSIONS

Deliverable 3.4, building upon the foundation set by Deliverable 3.2, furthered ENGAGE's mission to harness validated actionable knowledge on societal resilience. It aimed to demonstrate the tangible benefits and impact of the project's solutions across various disaster scenarios while also identifying best practices for communication and social media. The refined blueprint of an AI-enabled chatbot for emergencies and disasters, addressing both design and implementation intricacies, offers a roadmap for innovative solutions that present four pivotal contributions.

The first contribution is the validation and refinement of ENGAGE's underlying assumption that AI-enabled technologies can bolster societal resilience. This deliverable delved deep into the challenges this assumption faces, from the hesitancy of authorities, first responders, and the public to technological constraints like biases and inaccuracies. Yet, despite these challenges, we proposed a myriad of solutions and showcased technologies that can mitigate these barriers. Thus, while optimism is warranted, it should be paired with a judicious adoption of the blueprint outlined in this deliverable.

The second major contribution is the blueprint's emphasis on an AI-enabled chatbot that empowers emergency authorities to deliver timely, context-aware responses to the public throughout all phases of emergencies. Rooted in contemporary technologies, the blueprint underscores the importance of real-time information processing and the augmentation of datasets held by authorities and first responders, ensuring immediate and contextually relevant responses.

Thirdly, the deliverable introduces innovative strategies to counteract misinformation during crises. By swiftly detecting and tracking trending falsehoods on social media, the AI-enabled chatbot plays a pivotal role. However, as delineated in the objectives, the chatbot's primary function isn't to actively disseminate counter-messages but to identify potential misinformation surges, analyze them internally, and then engage human fact-checkers for verification. Thus, the conclusion drawn is that AI-enabled chatbots, in their current iteration, should primarily serve as information identifiers and analyzers, rather than autonomous decision-makers.

Lastly, the fourth contribution revolves around the development of novel solutions for citizen engagement and the seamless transfer of research insights to the public. By tailoring the project's recommendations to cater to specific demographic needs and expectations, this deliverable stands as a cornerstone of ENGAGE's knowledge platform. It is poised to be an invaluable resource for authorities and first responders keen on developing AI-enabled chatbot solutions. Furthermore, its integration into D5.4, the website and knowledge platform, and D5.5's knowledge and innovation community, ensures its widespread accessibility and relevance.

The research leading to these results has received funding from Horizon 2020, the European Union's a Programme for Research and Innovation (H2020/2014-2020) under grant agreement n° 882850.

14 References

Abd-Alrazaq, A. A., Alajlani, M., Ali, N., Denecke, K., Bewick, B. M., & Househ, M. (2021). Perceptions and opinions of patients about mental health chatbots: Scoping review. Journal of Medical Internet Research, 23(1), 17828.

Abraham, S. T. & Thazhathethil. (2023). Bindu Vazhakkatte; Tourism Disaster Management through Chatbots as an Alternative Tool of. Communication Revista Turismo & Desenvolvimento, 41, 349–367.

Adam, H., Balagopalan, A., Alsentzer, E., Christia, F., & Ghassemi, M. (2022). Mitigating the impact of biased artificial intelligence in emergency decision-making Communications Medicine 2 1 149. Nature Publishing Group UK.

Alessi, L., Benczur, P., Campolongo, F., Cariboni, J., Manca, A. R., Menyhert, B., & Pagano, A. (2018). The resilience of EU Member States to the financial and economic crisis.

Almalki, M., & Azeez, F. (2020). Health Chatbots for Fighting COVID-19: A scoping review. Acta Informatica Medica, 28(4), 241.

Al-Ramahi, M., Elnoshokaty, A., El-Gayar, O., Nasralah, T., & Wahbeh, A. (n.d.). Public discourse against masks in the COVID-19 era: Infodemiology study of Twitter data JMIR Public Health and Surveillance 7 4 e26780 2021. JMIR Publications Inc.

Al-Rifaie, M. M., & Bishop, M. (2015). Weak and strong computational creativity. In Computational creativity research: Towards creative machines (pp. 37–49). Springer.

Altamimi, I., Altamimi, A., Alhumimidi, A. S., Altamimi, A., & Temsah, M.-H. (n.d.). Snakebite advice and counseling from artificial intelligence: An acute venomous snakebite consultation with ChatGPT Cureus 15 6 2023 Cureus.

Al-Wathinani, A. M., Alhallaf, M. A., Borowska-Stefańska, M., Wiśniewski, S., Sultan, M. A. S., Samman, O. Y., Alobaid, A. M., Althunayyan, S. M., & Goniewicz, K. (2023). Elevating Healthcare. Rapid Literature Review on Drone Applications for Streamlining Disaster Management and Prehospital Care in Saudi Arabia Healthcare, 11 11 1575.

Andrade, Vargas, G., Padilha, V. L., Vahldick, A., & Oliveira. (2022). Francisco Henrique. Towards an Augmented Reality Application to Support Civil Defense in Visualizing the Susceptibility of Flooding Risk in Brazilian Urban Areas International Conference on Computational Science and Its Applications 494-506.

Annoni, A., Benczur, P., Bertoldi, P., Delipetrev, B., Prato, G., Feijoo, C., Macias, E. F., Gutierrez, E. G., Portela, M. I., & Junklewitz, H. (2018). Artificial Intelligence: A European Perspective.

Aoki, N. (2020). An experimental study of public trust in AI chatbots in the public sector. Government Information Quarterly, 37(4), 101490.

Ashfaq, M., Yun, J., Yu, S., & Loureiro, S. M. C. (2020). I, Chatbot: Modeling the determinants of users' satisfaction and continuance intention of AI-powered service agents. Telematics and Informatics, 54, 101473.

Authors Title Publication Volume Number Pages Year Publisher. (n.d.).

Ayers, J. W., Poliak, A., Dredze, M., Leas, E. C., Zhu, Z., Kelley, J. B., Faix, D. J., Goodman, A. M., Longhurst, C. A., & Hogarth, M. (2023). Comparing physician and artificial intelligence chatbot responses to patient questions posted to a public social media forum JAMA internal medicine.

The research leading to these results has received funding from Horizon 2020, the European Union's a Programme for Research and Innovation (H2020/2014-2020) under grant agreement n° 882850.

Baker, S., Waycott, J., Carrasco, R., Kelly, R. M., Jones, A. J., Lilley, J., Dow, B., Batchelor, F., Hoang, T., & Vetere, F. (n.d.). Avatar-mediated communication in social VR: an in-depth exploration of older adult interaction in an emerging communication. Platform Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems 13-Jan 2021.

Barojan, D. (2021). Building Digital Resilience Ahead of Elections and Beyond. In Disinformation and Fake News (pp. 61–73). Palgrave Macmillan.

Battineni, G., Chintalapudi, N., & Amenta, F. (2020). AI chatbot design during an epidemic like the novel coronavirus (Vol. 8, Issue 2, p. 154).

Berdahl, C., Fischer, S., Marcelino, W., Ahluwalia, S., Cohen, C., Armstrong, C., & Mendel, P. (n.d.). 278 Dissemination of Information about Ivermectin on Twitter During the COVID-19 Public Health Emergency Annals of Emergency Medicine 80 4 S121 2022 Elsevier.

Bonfert, M., Reinschluessel, A. V., Putze, S., Lai, Y., Alexandrovsky, D., Malaka, R., & Döring, T. (n.d.). Seeing the faces is so important—Experiences from online team meetings on commercial virtual reality platforms Frontiers in Virtual Reality 3 945791 2023 Frontiers.

Brandtzaeg, P. B., & Følstad, A. (2018). Chatbots: Changing user needs and motivations. Interactions, 25(5), 38–43.

Brown, N., Margus, C., Hart, A., Sarin, R., Hertelendy, A., & Ciottone, G. (2023). Virtual reality training in disaster medicine: A systematic review of the literature Simulation in Healthcare 18 (Vol. 4, pp. 255–261).

Carenzo, L., Costantini, E., Greco, M., Barra, F., Rendiniello, V., Mainetti, M., Bui, R., Zanella, A., Grasselli, G., & Lagioia, M. (2020). Hospital surge capacity in a tertiary emergency referral centre during the COVID-19 outbreak in Italy. Anaesthesia, 75(7), 928–934.

Chan, S. W., Abid, S. K., Sulaiman, N., Nazir, U., & Azam, K. (n.d.). A systematic review of the flood vulnerability using geographic information system Heliyon 2022 Elsevier.

Chatbot as a telehealth intervention strategy in the COVID-19 pandemic: Lessons learned from an action research approach CLEI. (n.d.).

Chaudhry, S., & Yuksel, M. (XXXX). Using Social Media for Crowd-Sourced Public Safety. MMTC Communications-Frontiers.

Cheng, K., Li, Z., Guo, Q., Sun, Z., Wu, H., & Li, C. (2023). Emergency surgery in the era of artificial intelligence: ChatGPT could be the doctor’s right-hand man. International Journal of Surgery, 109 6, 1816–1818.

Cheng, Y., & Jiang, H. (2020). AI-Powered mental health chatbots: Examining users’ motivations, active communicative action and engagement after mass-shooting disasters. Journal of Contingencies and Crisis Management, 28(3), 339–354.

Chiou, H., Voegeli, C., Wilhelm, E., Kolis, J., Brookmeyer, K., & Prybylski, D. (n.d.). The future of infodemic surveillance as public health surveillance Emerging Infectious Diseases 28.

Chon, M.-G., & Kim, S. (2022). Dealing with the COVID-19 crisis: Theoretical application of social media analytics in government crisis management Public relations review 48: Vol. 3 102201. Elsevier.

Cicco, R., Silva, S. C., & Alparone, F. R. (2020). Millennials’ attitude toward chatbots: An experimental study in a social relationship perspective. International Journal of Retail & Distribution Management.

The research leading to these results has received funding from Horizon 2020, the European Union's a Programme for Research and Innovation (H2020/2014-2020) under grant agreement n° 882850.

Clark, A., Fox, C., & Lappin, S. (2013). The handbook of computational linguistics and natural language processing. John Wiley & Sons.

Coeckelbergh, M. (2020). AI ethics. MIT Press.

Crutchfield, A. S., & Harkey, K. A. (2019). A comparison of call volumes before, during, and after Hurricane Harvey. The American Journal of Emergency Medicine, 37(10), 1904–1906.

Cruz, T., J, A., Aybar-Mejía, M., Roque, D., Yobany, C. R., Karla, D., Gabriel, J., Weeks, R., Dinelson, M.-H., Deyslen, H.-C., & Luis. (2020). Implications of 5G Technology in the Management of Power Microgrids. A Review of the Literature Energies, 16 4.

Daume, S., Galaz, V., & Bjersér, P. (563). Automated Framing of Climate Change? The Role of Social Bots in the Twitter Climate Change Discourse During the 2019/2020 Australia Bushfires Social Media+ Society 9. SAGE Publications Sage UK.

Engelen, J. E., & Hoos, H. H. (2020). A survey on semi-supervised learning. Machine Learning, 109(2), 373–440.

Esposito, M., Palma, L., Belli, A., Sabbatini, L., & Pierleoni, P. (n.d.). Recent advances in internet of things solutions for early warning systems: A review Sensors 22 6 2124 2022 MDPI.

Figueira, Á., & Oliveira, L. (2017). The current state of fake news: Challenges and opportunities. Procedia Computer Science, 121, 817–825.

Følstad, A., Skjuve, M., & Brandtzaeg, P. B. (2018). Different chatbots for different purposes: Towards a typology of chatbots to understand interaction design (pp. 145–156).

Ganapathy, S., Chang, S. Y. S., Tan, J. M. C., Lim, C., & Ng, K. C. (2023). Acute paediatrics tele-support for caregivers in Singapore: An initial experience with a prototype Chatbot. UPAL Singapore Medical Journal, 64 5, 335.

Giansanti, D. (n.d.). The Chatbots Are Invading Us: A Map Point on the Evolution, Applications, Opportunities. In And Emerging Problems in the Health Domain Life 13 5 1130 2023 MDPI.

Gunawong, P., & Jankananon, P. (2015). Flood 2.0: Facebook use and reactions during the 2011/2012 flood in Thailand. International Journal of Innovation and Learning, 17 2, 162–173.

Gupta, A., Hathwar, D., & Vijayakumar, A. (2020). Introduction to AI chatbots. Int J Eng Res Technol (IJERT, 9(7), 255–258.

Hakak, S., Khan, W. Z., Bhattacharya, S., Reddy, G. T., & Choo, K. K. R. (2020). Propagation of fake news on social media: Challenges and opportunities. International Conference on Computational Data and Social Networks, 345–353.

Haque, A., Chowdhury, M. N.-U.-R., & Soliman, H. (n.d.). Transforming chronic disease management with chatbots: Key use cases for personalized and cost-effective care 2023 TechRxiv.

Hasan, I., Rizvi, S., Jain, S., & Huria, S. (2021). The AI-enabled Chatbot Framework for Intelligent Citizen-Government Interaction for Delivery of Services (pp. 601–606).

Hayes, J. L., Britt, B. C., Evans, W., Rush, S. W., Towery, N. A., & Adamson, A. C. (2021). Can Social Media Listening Platforms. Artificial Intelligence Be Trusted? Examining the Accuracy of Crimson Hexagon's (Now Brandwatch Consumer Research's, 50(1), 81–91.

Heires, K. (2017). Technology to the Rescue. Risk Management, 64(11), 32–35.

Herrera, L. C. & Gjøsæter. (n.d.). Terje; Community Segmentation and Inclusive Social Media Listening 2022 International Association for Information Systems for Crisis Response and.

Herrera, L. C., Majchrzak, T. A., & Thapa, D. (2021). Ecosystem of Social Media Listening Practices for Crisis Management. Conference on E-Business, e-Services and e-Society, 710–722.

Hill, J., Ford, W. R., & Farreras, I. G. (2015). Real conversations with artificial intelligence: A comparison between human–human online conversations and human–chatbot conversations. Computers in Human Behavior, 49, 245–250.

Hoernke, K., Djellouli, N., Andrews, L., Lewis-Jackson, S., Manby, L., Martin, S., Vanderslott, S., & Vindrola-Padros, C. (n.d.). Frontline healthcare workers' experiences with personal protective equipment during the COVID-19 pandemic in the UK: a rapid qualitative appraisal BMJ open 11 1 e046199 2021. British Medical Journal Publishing Group.

Hrabí, M. (2020). Call centres: Going voice-first in the post-Covid world. Biometric Technology Today, 2020(8), 10–12.

Ignatidou, S. (2019). AI-driven personalisation in digital media: Political and societal implications.

Jiang, Q., Zhang, Y., & Pian, W. (n.d.). Chatbot as an emergency exist: Mediated empathy for resilience via human-AI interaction during the COVID-19 pandemic Information processing & management 59 6 103074 2022 Elsevier.

Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. Science, 349(6245), 255–260.

Karakose, T., Demirkol, M., Yirci, R., Polat, H., Ozdemir, T. Y., & Tülübaş, T. (n.d.). A Conversation with ChatGPT about Digital Leadership and Technology Integration.

Kelley, B. (2021). The rise of the 'quarantine bar simulator': The uses and gratifications of social VR during the COVID-19 pandemic. 4th International Conference on Information and Computer Technologies (ICICT) 216-221 2021 IEEE.

Kertysova, K. (2018). Artificial Intelligence and Disinformation: How AI Changes the Way Disinformation is Produced, Disseminated, and Can Be Countered. Security and Human Rights, 29(1–4), 55–81.

Kohnke, L. (2022). A qualitative exploration of student perspectives of chatbot use during emergency remote teaching. International Journal of Mobile Learning and Organisation, 16 4, 475–488.

Kooli, C. (n.d.). Chatbots in education and research: A critical examination of ethical implications and solutions Sustainability.

Krueger, M., & Albris, K. (2021). Resilience unwanted: Between control and cooperation in disaster response Security dialogue 52 4 343-360. SAGE Publications Sage UK.

Landeta, J. (2006). Current validity of the Delphi method in social sciences Technological forecasting and social change 73 (Vol. 5). Elsevier.

Laumer, S., Maier, C., & Gubler, F. T. (2019). Chatbot acceptance in healthcare: Explaining user adoption of conversational agents for disease diagnosis.

Lazer, D. M., Baum, M. A., Benkler, Y., Berinsky, A. J., Greenhill, K. M., Menczer, F., & Zittrain, J. L. (2018). The science of fake news. Science, 359(6380), 1094–1096.

Lei, S. I., Shen, H., & Ye, S. (2021). A comparison between chatbot and human service: Customer perception and reuse intention. International Journal of Contemporary Hospitality Management.

Linstone, H. A., & Turoff, M. (n.d.). The delphi method 1975 Addison-Wesley.

The research leading to these results has received funding from Horizon 2020, the European Union's a Programme for Research and Innovation (H2020/2014-2020) under grant agreement n° 882850.

Liu, B. (2021). ‘Weak AI’ is Likely to Never Become" Strong AI", So What is its Greatest Value for us?

Lohiniva, A.-L., Sibenberg, K., Austero, S., & Skogberg, N. (2022). Social listening to enhance access to appropriate pandemic information among culturally diverse populations: Case study from Finland JMIR infodemiology 2. JMIR Publications Inc.

Lu, Y., Qi, S., Cheng, K., & Wu. (1097). Haiyang; WHO declares end of mpox global health emergency: First glance from a perspective of ChatGPT/GPT-4. International Journal of Surgery.

Luccioni, A. S., Pham, K. H., Lam, C. S. N., Aylett-Bullock, J., & Luengo-Oroz, M. (2021). In Ensuring the Inclusive Use of Natural Language Processing in the Global Response to COVID-19.

MacKenzie, A. (2004). Infrastructural Resilience: Lessons from the WTC Collapse. E. Nz Magazine: The Magazine of Technical Enterprise, 5(1), 4–13.

MacKenzie, C. A., & Barker, K. (2013). Empirical data and regression analysis for estimation of infrastructure resilience with application to electric power outages. Journal of Infrastructure Systems, 19(1), 25–35.

Madianou, M. (2020). Non-human humanism: When AI for good turns out to be bad. AoIR Selected Papers of Internet Research.

Mahabub, A. (2020). A robust technique of fake news detection using Ensemble Voting Classifier and comparison with other classifiers. SN Applied Sciences, 2(4), 1–9.

Meyrick & Julian. (n.d.). The Delphi method and health research Health education 103 1 16-Jul 2003. MCB UP Ltd.

Miao, L., Last, M., & Litvak, M. (n.d.). Tracking social media during the COVID-19 pandemic: The case study of lockdown in New York State Expert Systems with Applications 187 115797 2022 Elsevier.

Mohri, M., Rostamizadeh, A., & Talwalkar, A. (2018). Foundations of machine learning. MIT press.

Morgado, G. J. F. (2023). Live preview and coding assistance tool for Virtual Reality programming with A-Frame.

Nadarzynski, T., Miles, O., Cowie, A., & Ridge, D. (2019). Acceptability of artificial intelligence (AI)-led chatbot services in healthcare: A mixed-methods study. Digital Health, 5, 2055207619871808.

NAKANISHI, E., ONO, H., KAWANO, T., & SAKASHITA, R. (2023). Why do we not develop a social game about evacuation simulation for disasters preparedness? Health Emergency and Disaster Nursing, 10 1, 46–48.

Natale, S. (2019). If software is narrative: Joseph Weizenbaum, artificial intelligence and the biographies of ELIZA. New Media & Society, 21(3), 712–728.

Noble, J. M., Zamani, A., Gharaat, M., Merrick, D., Maeda, N., Foster, A. L., Nikolaidis, I., Goud, R., Stroulia, E., & Agyapong, V. I. O. (2022). Developing, implementing, and evaluating an artificial intelligence–guided mental health resource navigation chatbot for health care workers and their families during and following the COVID-19 pandemic: Protocol for a cross-sectional study JMIR. Research Protocols, 11 7 e33717.

The research leading to these results has received funding from Horizon 2020, the European Union's a Programme for Research and Innovation (H2020/2014-2020) under grant agreement n° 882850.

Palanica, A., Flaschner, P., Thommandram, A., Li, M., & Fossat, Y. (2019). Physicians' perceptions of chatbots in health care: Cross-sectional web-based survey. Journal of Medical Internet Research, 21(4), 12887.

Palumbo, A. (n.d.). Microsoft HoloLens 2 in medical and healthcare context: State of the art and future prospects Sensors 22.

Patwa, P., Sharma, S., Pykl, S., Guptha, V., Kumari, G., Akhtar, M. S., & Chakraborty, T. (2021). Fighting an infodemic: Covid-19 fake news dataset. In International Workshop on Combating On line Ho st ile Posts in Regional Languages dur ing Emergency Situation (pp. 21–29). Springer.

Rao, N. R. (n.d.). Social media listening and monitoring for business applications 2016 IGI Global.

Rao, N. R. (Ed.). (2016). Social media listening and monitoring for business applications. IGI Global.

Richardson, M. (2016). The role social media played in wake of Orlando nightclub massacre. Bizjournals. Com.

Rodríguez Cardona, D., Werth, O., Schönborn, S., & Breitner, M. H. (2019). A mixed methods analysis of the adoption and diffusion of Chatbot Technology in the German insurance sector.

Rodriguez-Arrastia, M., Martinez-Ortigosa, A., Ruiz-Gonzalez, C., Ropero-Padilla, C., Roman, P., & Sanchez-Labraca, N. (2022). Experiences and perceptions of final-year nursing students of using a chatbot in a simulated emergency situation. A Qualitative Study Journal of Nursing Management, 30 8, 3874–3884.

Rosis, Sabina, L., Milena, P., Michelangelo, V., & Milena. (2023). Analyzing the emotional impact of COVID-19 with Twitter data: Lessons from a B-VAR analysis on Italy Socio-Economic Planning Sciences 87 101610. Elsevier.

Sample, C., Jensen, M. J., Scott, K., McAlaney, J., Fitchpatrick, S., Brockinton, A., & Ormrod, A. (2020). Interdisciplinary lessons learned while researching fake news. Frontiers in Psychology, 11, 2947.

Sangaiah, A. K., Rezaei, S., Javadpour, A., & Zhang, W. (2023). Explainable AI in big data intelligence of community detection for digitalization e-healthcare services Applied Soft Computing 136 110119. Elsevier.

Saravi, S., Kalawsky, R., Joannou, D., Rivas Casado, M., Fu, G., & Meng, F. (2019). Use of artificial intelligence to improve resilience and preparedness against adverse flood events. Water, 11(5), 973.

Sarbay, İ., Berikol, G. B., & Özturan, İ. U. (2023). Performance of emergency triage prediction of an open access natural language processing based chatbot application (ChatGPT): A preliminary, scenario-based cross-sectional study. Turkish Journal of Emergency Medicine, 23 3, 156.

Searle, J. R. (1980). Minds, brains, and programs. Behavioral and Brain Sciences, 3(3), 417–424.

Sharma, D., Kaushal, S., Kumar, H., & Gainder. (n.d.). Shalini; Chatbots in Healthcare: Challenges. Technologies and Applications 2022 4th International Conference on Artificial Intelligence and Speech Technology (AIST) 6-Jan 2022 IEEE.

Smith, S. W., Braun, J., Portelli, I., Malik, S., Asaeda, G., Lancet, E., & Goldfrank, L. R. (2016). Prehospital indicators for disaster preparedness and response: New York City emergency medical services in hurricane sandy. Disaster Medicine and Public Health Preparedness, 10(3), 333–343.

Stanga, C., Banfi, F., & Roascio, S. (n.d.). Enhancing Building Archaeology: Drawing, UAV Photogrammetry and Scan-to-BIM-to-VR Process of Ancient Roman Ruins.

Stewart, M. C., & Wilson, B. G. (n.d.). The dynamic role of social media during Hurricane# Sandy: An introduction of the STREMII model to weather the storm of the crisis lifecycle Computers in Human Behavior 54 639-646 2016 Elsevier.

Stieglitz, S., Hofeditz, L., Brünker, F., Ehnis, C., Mirbabaie, M., & Ross, B. (2022). Design principles for conversational agents to support Emergency Management Agencies. International Journal of Information Management, 63 102469.

Stroud, S. R. (2019). Pragmatist media ethics and the challenges of fake news. Journal of Media Ethics, 34(4), 178–192.

Thies, I. M., Menon, N., Magapu, S., Subramony, M., & O'neill, J. (2017). How do you want your chatbot? An exploratory Wizard-of-Oz study with young, urban Indians. IFIP Conference on Human-Computer Interaction, 441–459.

Trilaksono, B. R., Riza, H., Jarin, A., Darmayanti, N. D. S., & Liawatimena, S. (n.d.). Prosiding Use Cases Artificial Intelligence Indonesia. In Embracing Collaboration for Research and Industrial Innovation in Artificial Intelligence 2023 Penerbit BRIN.

Tsai, M.-H., Chen, J. Y., & Kang, S.-C. (2019). Ask Diana: A keyword-based chatbot system for water-related disaster management. Water, 11(2), 234.

Tülübaş, T., Demirkol, M., Ozdemir, T. Y., Polat, H., Karakose, T., & Yirci, R. (2023). An interview with ChatGPT on emergency remote teaching: A comparative analysis based on human–AI collaboration Educational Process. International Journal.

Verma, A., Chandra, P., & Joshi, A. (2021). AI Chatbots, its Feasibility and Reliability in Modern World.

Vinuesa, R., Azizpour, H., Leite, I., Balaam, M., Dignum, V., Domisch, S., Felländer, A., Langhans, S. D., Tegmark, M., & Nerini, F. F. (2020). The role of artificial intelligence in achieving the Sustainable Development Goals. Nature Communications, 11(1), 1–10.

Webb, J. J. (n.d.). Proof of Concept: Using ChatGPT to Teach Emergency Physicians How to Break Bad News Cureus 15 5 2023 Cureus.

Wezel, M. M., Croes, E. A., & Antheunis, M. L. (2020). I'm Here for You": Can Social Chatbots Truly Support Their Users? A Literature Review. In International Workshop on Chatbot Research and Design (pp. 96–113). Springer.

White, B. K., Martin, A., & White, J. A. (2022). User Experience of COVID-19 Chatbots. Scoping Review Journal of Medical Internet Research, 24, 12 35903.

Yamin, M. M., Ullah, M., Ullah, H., & Katt, B. (2021). Weaponized AI for cyber attacks. Journal of Information Security and Applications, 57, 102722.

Yang, B., Kim, M., Lee, C., Hwang, S., & Choi, J. (2022). Developing an Automated Analytical Process for Disaster Response and Recovery in Communities Prone to Isolation. International Journal of Environmental Research and Public Health, 19, 21 13995.

Yoo, S., Drumwright, M. E., & Piscarac, D. (n.d.). Augmented reality and nonprofit marketing during the COVID-19 pandemic: AR factors that influence community participation and willingness to contribute to causes Nonprofit Management and Leadership 2023. Wiley Online Library.

The research leading to these results has received funding from Horizon 2020, the European Union's a Programme for Research and Innovation (H2020/2014-2020) under grant agreement n° 882850.

Zachlod, C., Samuel, O., Ochsner, A., & Werthmüller, S. (2022). Analytics of social media data–State of characteristics and application. Journal of Business Research, 144, 1064–1076.

Zhu, Y., Janssen, M., Wang, R., & Liu, Y. (2022). It is me, chatbot: Working to address the COVID-19 outbreak-related mental health issues in China. User Experience, Satisfaction, and Influencing Factors International Journal of Human–Computer Interaction, 38 12, 1182–1194.

Zifu, C. (2016). Strong AI and super AI: technology rationality and its criticism. Science and Management, 05.

(2022). In Decision Support System Design For Low-Voice Emergency Medical Calls At Smart City Based On Chatbot Management In Social Networks Webology (ISSN: 1735-188X.

The research leading to these results has received funding from Horizon 2020, the European Union's a Programme for Research and Innovation (H2020/2014-2020) under grant agreement n° 882850.

15 APPENDICES

15.1 APPENDIX A: THE FIRST EDELPHI QUESTIONNAIRE

Professionals eDelphi – AI-enabled Chatbots for Emergencies and Disasters – Blueprint Evaluation

Start of Block: Intro

Q1 ENGAGE is an EU-funded project for developing new strategies for building societal resilience. As part of the project, we develop a blueprint of an AI-enabled chatbot for emergencies and disasters. The main aims of the AI-enabled chatbot are to distribute information and answer public queries 24/7 during adversities, respond to the surge of calls to emergency call centres during disasters without depending solely on human receptionists, and neutralise false information.

As a professional in your field, we would highly appreciate your opinion about the blueprint we are creating. The blueprint aims to set guidelines and recommendations regarding future designs of such AI-enabled chatbots.

After gathering all responses, we will analyse the results and share with you how different professionals (anonymously) replied to the questions. Then, based on the other answers, we will ask you to respond again to a shortened version of the questionnaire once or twice more.

Q2 In the next screen, we will present you with an initial blueprint (including) design directions for creating an AI-enabled chatbot for emergencies and disasters.

Q78 We want to achieve a consensus among professionals concerning the efficacy of the blueprint. Thus, with your permission, following the initial contribution, we will collate all the responses and send the assimilated response to all participants for up to two additional cycles of response.

*

The research leading to these results has received funding from Horizon 2020, the European Union's a Programme for Research and Innovation (H2020/2014-2020) under grant agreement n° 882850.

Q45 As mentioned above, after gathering all responses, we will analyse the results and share with you how different experts replied to the questions. Please kindly write down your email address so we can email you in a few weeks the results.

____

End of Block: Intro

Start of Block: The Whole Blueprint

Q3
Please take a few moments to review the blueprint, its various components, and how they are connected (the process). Then, in the following two sections, we will ask you several questions about the blueprint. Please hit next when you are done.

End of Block: The Whole Blueprint

Start of Block: General - Before

Q4 The blueprint includes 12 components: input types, channels & platforms, UX and logic, security & privacy, logging and documentation, existing platforms and models for training the chatbot, quality assurance, monitoring and reporting, statistics, alerts, datasets and manual human monitoring.

In the next section, please review the 12 components and answer a few questions about each. **However, at this moment, is there a component you think may be missing in the blueprint or anything else you would like to comment on?**

◯ Yes (1)

◯ No (2)

End of Block: General - Before

Start of Block: General - Before (Cont)

Display This Question:

If The blueprint includes 12 components: input types, channels & platforms, UX and logic, security &... = Yes

Q5 Following the previous question, what do you want to comment on or add? Please elaborate as much as you can.

____

End of Block: General - Before (Cont)

The research leading to these results has received funding from Horizon 2020, the European Union's a Programme for Research and Innovation (H2020/2014-2020) under grant agreement n° 882850.

Start of Block: 1. Input Types

Q6 **1 (Out of 12): Input Types**

Reminder: This is the full blueprint (the section of input types is circled in red).

Q7 The component of channels of platforms refers to where the chatbot's interface for communicating with users will be implemented. It refers to the various possibilities, from emergency service call centres to social media and emergency apps.

To what extent do you think that this component:

	1. Very Low Extent (1)	2. Low Extent (2)	3. Neutral (3)	4. High Extent (4)	5. Very High Extent (5)
Can have a technological contribution to the goals of the AI-enabled chatbot. (1)	○	○	○	○	○
Includes all relevant considerations that you, as an expert in your field, would expect from an AI-enabled chatbot blueprint. (2)	○	○	○	○	○
Is located in the correct place in the process of the blueprint. (3)	○	○	○	○	○

The research leading to these results has received funding from Horizon 2020, the European Union's a Programme for Research and Innovation (H2020/2014-2020) under grant agreement n° 882850.

Display This Question:

If The component of channels of platforms refers to where the chatbot's interface for communicating... = Includes all relevant considerations that you, as an expert in your field, would expect from an AI-enabled chatbot blueprint. [1. Very Low Extent]

Or The component of channels of platforms refers to where the chatbot's interface for communicating... = Includes all relevant considerations that you, as an expert in your field, would expect from an AI-enabled chatbot blueprint. [2. Low Extent]

Optional: As an expert in your field, what other considerations would you expect this component to include, as part of an AI-enabled chatbot blueprint?

____

Display This Question:

If The component of channels of platforms refers to where the chatbot's interface for communicating... = Is located in the correct place in the process of the blueprint. [1. Very Low Extent]

Or The component of channels of platforms refers to where the chatbot's interface for communicating... = Is located in the correct place in the process of the blueprint. [2. Low Extent]

Optional: Where would you locate this component, otherwise?

____

End of Block: 1. Input Types

Start of Block: 2. Channels & Platforms

Q8 **2 (Out of 12): Channels & Platforms**

Reminder: This is the full blueprint (the section of channels & platforms is circled in red).

Q9 The component of channels of platforms refers to where the chatbot's interface for communicating with users will be implemented. It refers to the various possibilities, from emergency service call centres to social media and emergency apps.

The research leading to these results has received funding from Horizon 2020, the European Union's a Programme for Research and Innovation (H2020/2014-2020) under grant agreement n° 882850.

To what extent do you think that this component:

	1. Very Low Extent (1)	2. Low Extent (2)	3. Neutral (3)	4. High Extent (4)	5. Very High Extent (5)
Can have a technological contribution to the goals of the AI-enabled chatbot. (1)	○	○	○	○	○
Includes all relevant considerations that you would expect from an AI-enabled chatbot blueprint. (2)	○	○	○	○	○
Is located in the correct place in the process of the blueprint. (3)	○	○	○	○	○

Display This Question:

If The component of channels of platforms refers to where the chatbot's interface for communicating... = Includes all relevant considerations that you would expect from an AI-enabled chatbot blueprint. [1. Very Low Extent]

Or The component of channels of platforms refers to where the chatbot's interface for communicating... = Includes all relevant considerations that you would expect from an AI-enabled chatbot blueprint. [2. Low Extent]

Q52 Optional: As an expert in your field, what other considerations would you expect this component to include, as part of an AI-enabled chatbot blueprint?

____

Display This Question:

If The component of channels of platforms refers to where the chatbot's interface for communicating... = Is located in the correct place in the process of the blueprint. [1. Very Low Extent]

Or The component of channels of platforms refers to where the chatbot's interface for communicating... = Is located in the correct place in the process of the blueprint. [2. Low Extent]

The research leading to these results has received funding from Horizon 2020, the European Union's a Programme for Research and Innovation (H2020/2014-2020) under grant agreement n° 882850.

Q65 Optional: Where would you locate this component, otherwise?

____

End of Block: 2. Channels & Platforms

Start of Block: 3. UX & Logic

Q10 **3 (Out of 12): UX & Logic**

Reminder: This is the full blueprint (the section of UX and logic is circled in red).

Q11 The component of UX and logic refers to the services needed to facilitate the connection between the various communication channels and the chatbot – and vice-versa.

To what extent do you think that this component:

	1. Very Low Extent (1)	2. Low Extent (2)	3. Neutral (3)	4. High Extent (4)	5. Very High Extent (5)
Can have a technological contribution to the goals of the AI-enabled chatbot. (1)	○	○	○	○	○
Includes all relevant considerations that you, as an expert in your field, would expect from an AI-enabled chatbot blueprint. (2)	○	○	○	○	○
Is located in the correct place in the process of the blueprint. (3)	○	○	○	○	○

The research leading to these results has received funding from Horizon 2020, the European Union's a Programme for Research and Innovation (H2020/2014-2020) under grant agreement n° 882850.

Display This Question:

If The component of UX and logic refers to the services needed to facilitate the connection between... = Includes all relevant considerations that you, as an expert in your field, would expect from an AI-enabled chatbot blueprint. [1. Very Low Extent]

Or The component of UX and logic refers to the services needed to facilitate the connection between... = Includes all relevant considerations that you, as an expert in your field, would expect from an AI-enabled chatbot blueprint. [2. Low Extent]

Q53 Optional: As an expert in your field, what other considerations would you expect this component to include, as part of an AI-enabled chatbot blueprint?

____

Display This Question:

If The component of UX and logic refers to the services needed to facilitate the connection between... = Is located in the correct place in the process of the blueprint. [1. Very Low Extent]

Or The component of UX and logic refers to the services needed to facilitate the connection between... = Is located in the correct place in the process of the blueprint. [2. Low Extent]

Q66 Optional: Where would you locate this component, otherwise?

____

End of Block: 3. UX & Logic

Start of Block: 4. Security & Privacy

Q12 **4 (Out of 12): Security & Privacy**

Reminder: This is the full blueprint (the section of security & privacy is circled in red).

Q13 Security and privacy are needed to make the connection to the chatbot secure and private and to prevent the abuse of information. This component refers to the services needed to provide security and authentication measures for data storage, GDPR and more.

The research leading to these results has received funding from Horizon 2020, the European Union's a Programme for Research and Innovation (H2020/2014-2020) under grant agreement n° 882850.

To what extent do you think that this component:

	1. Very Low Extent (1)	2. Low Extent (2)	3. Neutral (3)	4. High Extent (4)	5. Very High Extent (5)
Can have a technological contribution to the goals of the AI-enabled chatbot. (1)	○	○	○	○	○
Includes all relevant considerations that you, as an expert in your field, would expect from an AI-enabled chatbot blueprint. (2)	○	○	○	○	○
Is located in the correct place in the process of the blueprint. (3)	○	○	○	○	○

Display This Question:

If Security and privacy are needed to make the connection to the chatbot secure and private and to p... = Includes all relevant considerations that you, as an expert in your field, would expect from an AI-enabled chatbot blueprint. [1. Very Low Extent]

Or Security and privacy are needed to make the connection to the chatbot secure and private and to p... = Includes all relevant considerations that you, as an expert in your field, would expect from an AI-enabled chatbot blueprint. [2. Low Extent]

Q54 Optional: As an expert in your field, what other considerations would you expect this component to include, as part of an AI-enabled chatbot blueprint?

____

The research leading to these results has received funding from Horizon 2020, the European Union's a Programme for Research and Innovation (H2020/2014-2020) under grant agreement n° 882850.

Display This Question:

If Security and privacy are needed to make the connection to the chatbot secure and private and to p... = Is located in the correct place in the process of the blueprint. [1. Very Low Extent]

Or Security and privacy are needed to make the connection to the chatbot secure and private and to p... = Is located in the correct place in the process of the blueprint. [2. Low Extent]

Q67 Optional: Where would you locate this component, otherwise?

____

End of Block: 4. Security & Privacy

Start of Block: 5. Logging & Documentation

Q14 **5 (Out of 12): Logging & Documentation**

Reminder: This is the full blueprint (the section of logging and documentation is circled in red).

Q15 This component refers to the need to log the data that the chatbot can collect (e.g., conversations, monitoring content on social media, documents provided) to provide diagnostic and analysis.

The research leading to these results has received funding from Horizon 2020, the European Union's a Programme for Research and Innovation (H2020/2014-2020) under grant agreement n° 882850.

To what extent do you think that this component:

	1. Very Low Extent (1)	2. Low Extent (2)	3. Neutral (3)	4. High Extent (4)	5. Very High Extent (5)
Can have a technological contribution to the goals of the AI-enabled chatbot. (1)	○	○	○	○	○
Includes all relevant considerations that you, as an expert in your field, would expect from an AI-enabled chatbot blueprint. (2)	○	○	○	○	○
Is located in the correct place in the process of the blueprint. (3)	○	○	○	○	○

Display This Question:

If This component refers to the need to log the data that the chatbot can collect (e.g., conversatio... = Includes all relevant considerations that you, as an expert in your field, would expect from an AI-enabled chatbot blueprint. [1. Very Low Extent]

Or This component refers to the need to log the data that the chatbot can collect (e.g., conversatio... = Includes all relevant considerations that you, as an expert in your field, would expect from an AI-enabled chatbot blueprint. [2. Low Extent]

Q55 Optional: As an expert in your field, what other considerations would you expect this component to include, as part of an AI-enabled chatbot blueprint?

____

The research leading to these results has received funding from Horizon 2020, the European Union's a Programme for Research and Innovation (H2020/2014-2020) under grant agreement n° 882850.

Display This Question:

If This component refers to the need to log the data that the chatbot can collect (e.g., conversatio... = Is located in the correct place in the process of the blueprint. [1. Very Low Extent]

Or This component refers to the need to log the data that the chatbot can collect (e.g., conversatio... = Is located in the correct place in the process of the blueprint. [2. Low Extent]

Q68 Optional: Where would you locate this component, otherwise?

____

End of Block: 5. Logging & Documentation

Start of Block: 6. Existing Platforms

Q16 **6 (Out of 12): Existing Platforms and Models for Training the Chatbot**

Reminder: This is the full blueprint (the section of existing platforms and models for training is circled in red).

Q17 The component of existing platforms and models for training the chatbot refers to three processes: cognition and intelligence, technological capabilities and preparing, processing & training.

AI-enabled chatbots rely on existing language models (e.g., GPT-3, InstructGPT, Dialogflow), for understanding and generating natural language (e.g., davinci, curie) or images (e.g., Dall-E, Google Vision). This component sets the guidelines on how the chatbot can interact with such platforms' APIs in order to be able to use these models. It allows the connection of the existing knowledge (e.g., datasets) of the chatbot with external capabilities (e.g., the language models).

Alternatively, chatbots can be trained independently, without the connection to external platforms, in a self-developed process, based on the three components that such platforms, as abovementioned, are using:

<u>Cognition and Intelligence</u>
The component of cognition and intelligence refers to two sub-categories. One is the relevant algorithms that the chatbots employ, such as NLP, STT, TTS, speech recognition, image and video analysis and more. The decision about the relevant algorithms results from other chatbot components (e.g., input types). In addition, cognition and intelligence also refer to questions of classifications, such as what intents should be covered in the chatbot. Who are the relevant entities? What is the structure of dialogues that the chatbot is expected to conduct? And more. Here, there are relevant issues to consider regarding the training, development and how the AI learns and improves. Questions such as is the AI self-developed or using public cloud AI should be

The research leading to these results has received funding from Horizon 2020, the European Union's a Programme for Research and Innovation (H2020/2014-2020) under grant agreement n° 882850.

answered.

Technological capabilities
The technological capabilities support the chatbot's functionality, and they depend on its goals. They can include capabilities such as indexing data, searchability, generating automatic or manual FAQs, creating a calendar of events and more.

Preparing, processing and training
Since raw data is mainly unstructured, services are needed to allow the AI-enabled chatbot to train and develop. Training and development of the chatbot are relevant for the backend improvement of the system. They include considerations regarding the chatbot's relevant data and analysis and extraction of information. This component highlights the importance of integrating these services in the design process of the chatbot.

When **referring** to existing platforms and models for training, as a whole, to what extent do you think that this component:

	1. Very Low Extent (1)	2. Low Extent (2)	3. Neutral (3)	4. High Extent (4)	5. Very High Extent (5)
Can have a technological contribution to the goals of the AI-enabled chatbot. (1)	○	○	○	○	○
Includes all relevant considerations that you, as an expert in your field, would expect from an AI-enabled chatbot blueprint. (2)	○	○	○	○	○
Is located in the correct place in the process of the blueprint. (3)	○	○	○	○	○

The research leading to these results has received funding from Horizon 2020, the European Union's a Programme for Research and Innovation (H2020/2014-2020) under grant agreement n° 882850.

Display This Question:

If The component of existing platforms and models for training the chatbot refers to three processes... = Includes all relevant considerations that you, as an expert in your field, would expect from an AI-enabled chatbot blueprint. [1. Very Low Extent]

Or The component of existing platforms and models for training the chatbot refers to three processes... = Includes all relevant considerations that you, as an expert in your field, would expect from an AI-enabled chatbot blueprint. [2. Low Extent]

Q56 Optional: As an expert in your field, what other considerations would you expect this component to include, as part of an AI-enabled chatbot blueprint?

____

Display This Question:

If The component of existing platforms and models for training the chatbot refers to three processes... = Is located in the correct place in the process of the blueprint. [1. Very Low Extent]

Or The component of existing platforms and models for training the chatbot refers to three processes... = Is located in the correct place in the process of the blueprint. [2. Low Extent]

Q69 Optional: Where would you locate this component, otherwise?

____

End of Block: 6. Existing Platforms

Start of Block: 7. QA

Q22 **7 (Out of 12): Quality Assurance (QA)**

Reminder: This is the full blueprint (the section of quality assurance (QA) is circled in red).

Q23 The component of QA refers to the need for regular tests of the chatbot and how they are implemented. It also includes the QA of the AI itself and the methods and tools available to ensure the AI is ethical and safe to use. Finally, it allows the identification of technical failures, failed conversations and other technical issues that require attention, changes and updates in the system.

The research leading to these results has received funding from Horizon 2020, the European Union's a Programme for Research and Innovation (H2020/2014-2020) under grant agreement n° 882850.

To what extent do you think that this component:

	1. Very Low Extent (1)	2. Low Extent (2)	3. Neutral (3)	4. High Extent (4)	5. Very High Extent (5)
Can have a technological contribution to the goals of the AI-enabled chatbot. (1)	○	○	○	○	○
Includes all relevant considerations that you, as an expert in your field, would expect from an AI-enabled chatbot blueprint. (2)	○	○	○	○	○
Is located in the correct place in the process of the blueprint. (3)	○	○	○	○	○

Display This Question:

If The component of QA refers to the need for regular tests of the chatbot and how they are implemen... = Includes all relevant considerations that you, as an expert in your field, would expect from an AI-enabled chatbot blueprint. [1. Very Low Extent]

Or The component of QA refers to the need for regular tests of the chatbot and how they are implemen... = Includes all relevant considerations that you, as an expert in your field, would expect from an AI-enabled chatbot blueprint. [2. Low Extent]

Q59 Optional: As an expert in your field, what other considerations would you expect this component to include, as part of an AI-enabled chatbot blueprint?

____

The research leading to these results has received funding from Horizon 2020, the European Union's a Programme for Research and Innovation (H2020/2014-2020) under grant agreement n° 882850.

Display This Question:

If The component of QA refers to the need for regular tests of the chatbot and how they are implemen... = Is located in the correct place in the process of the blueprint. [1. Very Low Extent]

Or The component of QA refers to the need for regular tests of the chatbot and how they are implemen... = Is located in the correct place in the process of the blueprint. [2. Low Extent]

Q72 Optional: Where would you locate this component, otherwise?

____

End of Block: 7. QA

Start of Block: 8. Monitoring & Reporting

Q24 **8 (Out of 12): Monitoring & Reporting**

Reminder: This is the full blueprint (the section of monitoring & reporting is circled in red).

Q25 The monitoring and reporting component refers to services that provide live and a-synchronous analyses of big datasets, providing real-time feedback, notifications and alerts about statistics, technical issues, occurrences, and anomalies. These can be analysed for further consideration or generate notifications for either the organisation's end-users (the public) or the professional emergency workers.

The research leading to these results has received funding from Horizon 2020, the European Union's a Programme for Research and Innovation (H2020/2014-2020) under grant agreement n° 882850.

To what extent do you think that this component:

	1. Very Low Extent (1)	2. Low Extent (2)	3. Neutral (3)	4. High Extent (4)	5. Very High Extent (5)
Can have a technological contribution to the goals of the AI-enabled chatbot. (1)	○	○	○	○	○
Includes all relevant considerations that you, as an expert in your field, would expect from an AI-enabled chatbot blueprint. (2)	○	○	○	○	○
Is located in the correct place in the process of the blueprint. (3)	○	○	○	○	○

Display This Question:

If The monitoring and reporting component refers to services that provide live and a-synchronous ana... = Includes all relevant considerations that you, as an expert in your field, would expect from an AI-enabled chatbot blueprint. [1. Very Low Extent]

Or The monitoring and reporting component refers to services that provide live and a-synchronous ana... = Includes all relevant considerations that you, as an expert in your field, would expect from an AI-enabled chatbot blueprint. [2. Low Extent]

Q60 Optional: As an expert in your field, what other considerations would you expect this component to include, as part of an AI-enabled chatbot blueprint?

____

The research leading to these results has received funding from Horizon 2020, the European Union's a Programme for Research and Innovation (H2020/2014-2020) under grant agreement n° 882850.

Display This Question:

If The monitoring and reporting component refers to services that provide live and a-synchronous ana... = Is located in the correct place in the process of the blueprint. [1. Very Low Extent]

Or The monitoring and reporting component refers to services that provide live and a-synchronous ana... = Is located in the correct place in the process of the blueprint. [2. Low Extent]

Q73 Optional: Where would you locate this component, otherwise?

____

End of Block: 8. Monitoring & Reporting

Start of Block: 9. Datasets

Q26 **9 (Out of 12): Datasets**

Reminder: This is the full blueprint (the section of datasets is circled in red).

Q27 The component of datasets refers to the need for data to train the chatbot and provide the desired information.

The research leading to these results has received funding from Horizon 2020, the European Union's a Programme for Research and Innovation (H2020/2014-2020) under grant agreement n° 882850.

To what extent do you think that this component:

	1. Very Low Extent (1)	2. Low Extent (2)	3. Neutral (3)	4. High Extent (4)	5. Very High Extent (5)
Can have a technological contribution to the goals of the AI-enabled chatbot. (1)	○	○	○	○	○
Includes all relevant considerations that you, as an expert in your field, would expect from an AI-enabled chatbot blueprint. (2)	○	○	○	○	○
Is located in the correct place in the process of the blueprint. (3)	○	○	○	○	○

Display This Question:

If The component of datasets refers to the need for data to train the chatbot and provide the desire... = Includes all relevant considerations that you, as an expert in your field, would expect from an AI-enabled chatbot blueprint. [1. Very Low Extent]

Or The component of datasets refers to the need for data to train the chatbot and provide the desire... = Includes all relevant considerations that you, as an expert in your field, would expect from an AI-enabled chatbot blueprint. [2. Low Extent]

Q61 Optional: As an expert in your field, what other considerations would you expect this component to include, as part of an AI-enabled chatbot blueprint?

____

The research leading to these results has received funding from Horizon 2020, the European Union's a Programme for Research and Innovation (H2020/2014-2020) under grant agreement n° 882850.

Display This Question:

If The component of datasets refers to the need for data to train the chatbot and provide the desire... = Is located in the correct place in the process of the blueprint. [1. Very Low Extent]

Or The component of datasets refers to the need for data to train the chatbot and provide the desire... = Is located in the correct place in the process of the blueprint. [2. Low Extent]

Q74 Optional: Where would you locate this component, otherwise?

____

End of Block: 9. Datasets

Start of Block: 10. Statistics

Q28 **10 (Out of 12): Statistics**

Reminder: This is the full blueprint (the section of statistics is circled in red).

Q29 The statistics component refers to the statistical analysis of the data received and produced by the chatbot for creating routine reports.

The research leading to these results has received funding from Horizon 2020, the European Union's a Programme for Research and Innovation (H2020/2014-2020) under grant agreement n° 882850.

To what extent do you think that this component:

	1. Very Low Extent (1)	2. Low Extent (2)	3. Neutral (3)	4. High Extent (4)	5. Very High Extent (5)
Can have a technological contribution to the goals of the AI-enabled chatbot. (1)	o	o	o	o	o
Includes all relevant considerations that you, as an expert in your field, would expect from an AI-enabled chatbot blueprint. (2)	o	o	o	o	o
Is located in the correct place in the process of the blueprint. (3)	o	o	o	o	o

Display This Question:

If The statistics component refers to the statistical analysis of the data received and produced by... = Includes all relevant considerations that you, as an expert in your field, would expect from an AI-enabled chatbot blueprint. [1. Very Low Extent]

Or The statistics component refers to the statistical analysis of the data received and produced by... = Includes all relevant considerations that you, as an expert in your field, would expect from an AI-enabled chatbot blueprint. [2. Low Extent]

Q62 Optional: As an expert in your field, what other considerations would you expect this component to include, as part of an AI-enabled chatbot blueprint?

____

The research leading to these results has received funding from Horizon 2020, the European Union's a Programme for Research and Innovation (H2020/2014-2020) under grant agreement n° 882850.

Display This Question:

If The statistics component refers to the statistical analysis of the data received and produced by... = Is located in the correct place in the process of the blueprint. [1. Very Low Extent]

Or The statistics component refers to the statistical analysis of the data received and produced by... = Is located in the correct place in the process of the blueprint. [2. Low Extent]

Q75 Optional: Where would you locate this component, otherwise?

____

End of Block: 10. Statistics

Start of Block: 11. Alerts

Q30 **11 (Out of 12): Alerts**

Reminder: This is the full blueprint (the section of alerts is circled in red).

Q31 The component of alerts refers to the ability of the chatbot to generate urgent alerts, whether to the public or the emergency workers, based on its monitoring process.

The research leading to these results has received funding from Horizon 2020, the European Union's a Programme for Research and Innovation (H2020/2014-2020) under grant agreement n° 882850.

To what extent do you think that this component:

	1. Very Low Extent (1)	2. Low Extent (2)	3. Neutral (3)	4. High Extent (4)	5. Very High Extent (5)
Can have a technological contribution to the goals of the AI-enabled chatbot. (1)	○	○	○	○	○
Includes all relevant considerations that you, as an expert in your field, would expect from an AI-enabled chatbot blueprint. (2)	○	○	○	○	○
Is located in the correct place in the process of the blueprint. (3)	○	○	○	○	○

Display This Question:

If The component of alerts refers to the ability of the chatbot to generate urgent alerts, whether t... = Includes all relevant considerations that you, as an expert in your field, would expect from an AI-enabled chatbot blueprint. [1. Very Low Extent]

Or The component of alerts refers to the ability of the chatbot to generate urgent alerts, whether t... = Includes all relevant considerations that you, as an expert in your field, would expect from an AI-enabled chatbot blueprint. [2. Low Extent]

Q63 Optional: As an expert in your field, what other considerations would you expect this component to include, as part of an AI-enabled chatbot blueprint?

____

The research leading to these results has received funding from Horizon 2020, the European Union's a Programme for Research and Innovation (H2020/2014-2020) under grant agreement n° 882850.

Display This Question:

If The component of alerts refers to the ability of the chatbot to generate urgent alerts, whether t... = Is located in the correct place in the process of the blueprint. [1. Very Low Extent]

Or The component of alerts refers to the ability of the chatbot to generate urgent alerts, whether t... = Is located in the correct place in the process of the blueprint. [2. Low Extent]

Q76 Optional: Where would you locate this component, otherwise?

____

End of Block: 11. Alerts

Start of Block: 12. Manual Human Monitoring

Q32 **12 (Out of 12): Manual Human Monitoring & Moderation**

Reminder: This is the full blueprint (the section of manual human monitoring is circled in red).

Q33 Manual human monitoring and moderation refers to the need to monitor the activity of the chatbot and its conversations, whether in the initial period or all the time, by human operators beyond automatic monitoring.

The research leading to these results has received funding from Horizon 2020, the European Union's a Programme for Research and Innovation (H2020/2014-2020) under grant agreement n° 882850.

To what extent do you think that this component:

	1, Very Low Extent (1)	2, Low Extent (2)	3. Neutral (3)	4. High Extent (4)	5. Very High Extent (5)
Can have a technological contribution to the goals of the AI-enabled chatbot. (1)	○	○	○	○	○
Includes all relevant considerations that you, as an expert in your field, would expect from an AI-enabled chatbot blueprint. (2)	○	○	○	○	○
Is located in the correct place in the process of the blueprint. (3)	○	○	○	○	○

Display This Question:

If Manual human monitoring and moderation refers to the need to monitor the activity of the chatbot... = Includes all relevant considerations that you, as an expert in your field, would expect from an AI-enabled chatbot blueprint. [1, Very Low Extent]

Or Manual human monitoring and moderation refers to the need to monitor the activity of the chatbot... = Includes all relevant considerations that you, as an expert in your field, would expect from an AI-enabled chatbot blueprint. [2, Low Extent]

Q64 Optional: As an expert in your field, what other considerations would you expect this component to include, as part of an AI-enabled chatbot blueprint?

____

The research leading to these results has received funding from Horizon 2020, the European Union's a Programme for Research and Innovation (H2020/2014-2020) under grant agreement n° 882850.

Display This Question:

If Manual human monitoring and moderation refers to the need to monitor the activity of the chatbot... = Is located in the correct place in the process of the blueprint. [1, Very Low Extent]

Or Manual human monitoring and moderation refers to the need to monitor the activity of the chatbot... = Is located in the correct place in the process of the blueprint. [2, Low Extent]

Q77 Optional: Where would you locate this component, otherwise?

____

End of Block: 12. Manual Human Monitoring

Start of Block: Blueprint as a whole

Q35 After reviewing the 12 components separately, please answer the following questions, referring once again to the blueprint as a whole.

Q34 To what extent do you agree or disagree that the process and the interaction between the different components in the blueprint (i.e., the arrows) are appropriate?

- ○ 1. Very Low Extent (1)
- ○ 2. Low Extent (2)
- ○ 3. Neutral (3)
- ○ 4. High Extent (4)
- ○ 5. Very High Extent (5)

Q36 Are there any components and/or interactions and/or something else in the blueprint? If yes, please elaborate.

____

Q79 In November 2022, OpenAI launched ChatGPT (Generative Pre-trained Transformer), based on large language models, with both supervised and reinforcement learning techniques. **How relevant or irrelevant do you think that the last blueprint, that we presented you, is in light of**

The research leading to these results has received funding from Horizon 2020, the European Union's a Programme for Research and Innovation (H2020/2014-2020) under grant agreement n° 882850.

the release of ChatGPT?

◯ 1. Highly Irrelevant (1)

◯ 2. Irrelevant (2)

◯ 3. Somewhat Irrelevant (3)

◯ 4. Neither Irrelevant not Relevant (4)

◯ 5. Somewhat Relevant (5)

◯ 6. Relevant (6)

◯ 7. Highly Relevant (7)

Display This Question:

If In November 2022, OpenAI launched ChatGPT (Generative Pre-trained Transformer), based on large la... = 5. Somewhat Relevant

Or In November 2022, OpenAI launched ChatGPT (Generative Pre-trained Transformer), based on large la... = 6. Relevant

Or In November 2022, OpenAI launched ChatGPT (Generative Pre-trained Transformer), based on large la... = 7. Highly Relevant

Q80 Optional: Can you shortly explain why you think it is still relevant?

____

Display This Question:

If In November 2022, OpenAI launched ChatGPT (Generative Pre-trained Transformer), based on large la... = 1. Highly Irrelevant

Or In November 2022, OpenAI launched ChatGPT (Generative Pre-trained Transformer), based on large la... = 2. Irrelevant

Or In November 2022, OpenAI launched ChatGPT (Generative Pre-trained Transformer), based on large la... = 3. Somewhat Irrelevant

Q81 Optional: Can you shortly explain why you think it is irrelevant?

____

The research leading to these results has received funding from Horizon 2020, the European Union's a Programme for Research and Innovation (H2020/2014-2020) under grant agreement n° 882850.

End of Block: Blueprint as a whole

Start of Block: Statistics

Q37 Finally, we would appreciate if you could answer about a few questions for statistical purposes.

Q44 What country are you from?

____

Q41 What is your role in your organisation?

____

*

Q42 How many years of experience working with either AI in general or chatbots, in particular, do you have?

____

Q43 What is the level of proficiency in designing or working with blueprints/architectures of AI-enabled chatbots or other AI designs that you have?

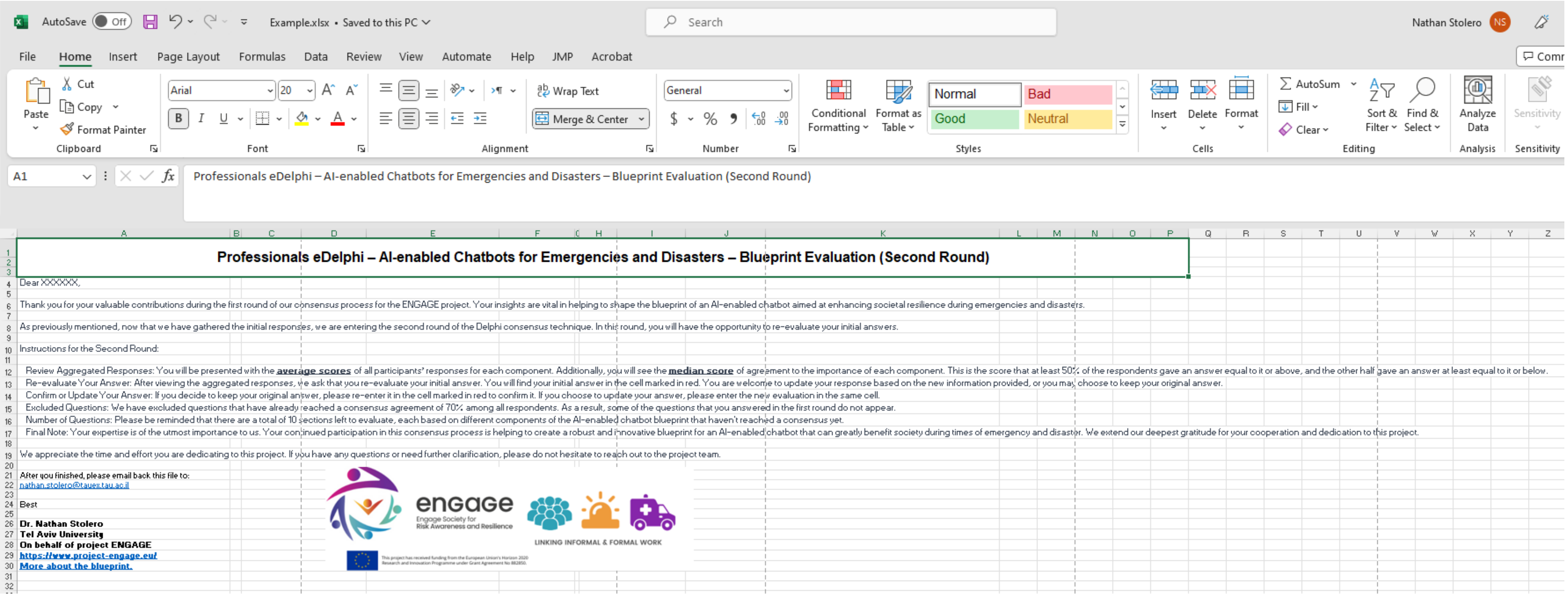
- 1. Not at all (1)
- 2. Very little (2)
- 3. Average (3)
- 4. High (4)
- 5. Very high (5)

End of Block: Statistics

The research leading to these results has received funding from Horizon 2020, the European Union's a Programme for Research and Innovation (H2020/2014-2020) under grant agreement n° 882850.

The research leading to these results has received funding from Horizon 2020, the European Union's a Programme for Research and Innovation (H2020/2014-2020) under grant agreement n° 882850.

15.2 Appendix B: The second eDelphi questionnaire



Professionals eDelphi – AI-enabled Chatbots for Emergencies and Disasters – Blueprint Evaluation (Second Round)

Dear XXXXXX,

Thank you for your valuable contributions during the first round of our consensus process for the ENGAGE project. Your insights are vital in helping to shape the blueprint of an AI-enabled chatbot aimed at enhancing societal resilience during emergencies and disasters.

As previously mentioned, now that we have gathered the initial responses, we are entering the second round of the Delphi consensus technique. In this round, you will have the opportunity to re-evaluate your initial answers.

Instructions for the Second Round:

Review Aggregated Responses: You will be presented with the **average scores** of all participants' responses for each component. Additionally, you will see the **median score** of agreement to the importance of each component. This is the score that at least 50% of the respondents gave an answer equal to it or above, and the other half gave an answer at least equal to it or below.
Re-evaluate Your Answer: After viewing the aggregated responses, we ask that you re-evaluate your initial answer. You will find your initial answer in the cell marked in red. You are welcome to update your response based on the new information provided, or you may choose to keep your original answer.
Confirm or Update Your Answer: If you decide to keep your original answer, please re-enter it in the cell marked in red to confirm it. If you choose to update your answer, please enter the new evaluation in the same cell.
Excluded Questions: We have excluded questions that have already reached a consensus agreement of 70% among all respondents. As a result, some of the questions that you answered in the first round do not appear.
Number of Questions: Please be reminded that there are a total of 10 sections left to evaluate, each based on different components of the AI-enabled chatbot blueprint that haven't reached a consensus yet.
Final Note: Your expertise is of the utmost importance to us. Your continued participation in this consensus process is helping to create a robust and innovative blueprint for an AI-enabled chatbot that can greatly benefit society during times of emergency and disaster. We extend our deepest gratitude for your cooperation and dedication to this project.

We appreciate the time and effort you are dedicating to this project. If you have any questions or need further clarification, please do not hesitate to reach out to the project team.

After you finished, please email back this file to:
nathan.stolero@tauex.tau.ac.il

Best

Dr. Nathan Stolero
Tel Aviv University
On behalf of project ENGAGE
https://www.project-engage.eu/
More about the blueprint.

The research leading to these results has received funding from Horizon 2020, the European Union's a Programme for Research and Innovation (H2020/2014-2020) under grant agreement n° 882850.

Here is an example of how to approach the questions:

This is the answer you originaly submitted

This is the average of all other responses

Here you can see the median answer

Here you need to input your answer. If you did not change your opinion, copy the number in the cell "Your Answer". If you have a new evaluation write in the new answer as follows.

This is just an example. Do not answer.

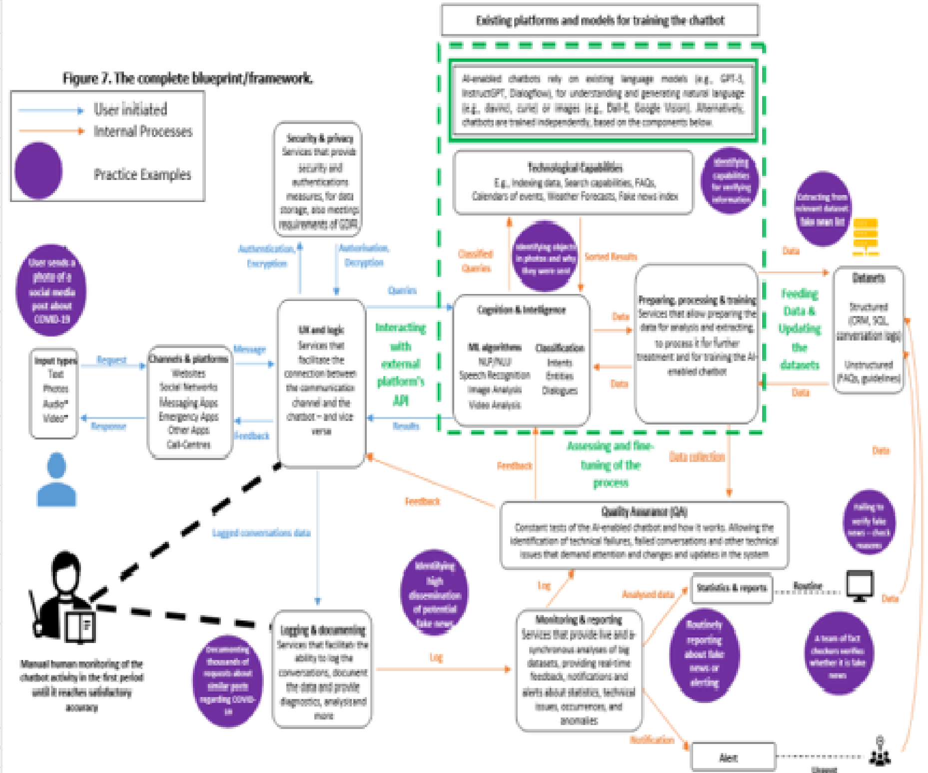
component: (1 very low extent, 2 low extent, 3 neut	Your Answer	The average of all responses	Median	What is your answer after viewing other respondents' evaluations?
Can have a technological contribution to the goals of the AI-enabled chatbot.	5	4.5	5	5
Includes all relevant considerations that you, as an expert in your field, would expect from an AI-enabled chatbot blueprint.	3	3.6	4	4
Is located in the correct place in the process of the	4	3.7	3	3

An example keeping the original answer

An example of a new evaluation - higher

An example of a new evaluation - lower

Reminder: This is the original blueprint

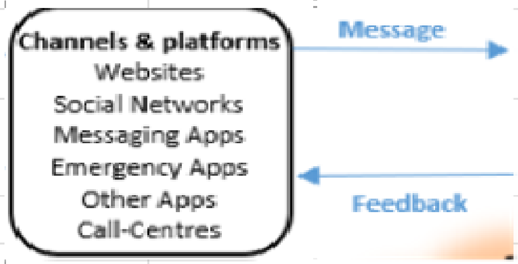


a Programme for Research and Innovation (H2020/2014-2020) under grant agreement n° 882850.

Q1/10

The component of <u>channels of platforms</u> refers to where the chatbot's interface for communicating with users will be implemented. It refers to the various possibilities, from emergency service call centres to social media and emergency apps.

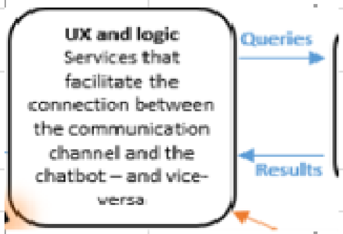
To what extent do you think that this component: (1 very low extent, 2 low extent, 3 neutral, 4 high extent, 5 very high extent)	Your Answer	The average of all responses	Median	What is your answer after viewing other respondents' evaluations?
Can have a technological contribution to the goals of the AI-enabled chatbot.	2	3.63	4	
Includes all relevant considerations that you, as an expert in your field, would expect from an AI-enabled chatbot blueprint.	3	3.75	4	



Q2/10

The component of <u>UX and logic</u> refers to the services needed to facilitate the connection between the various communication channels and the chatbot and vice-versa.

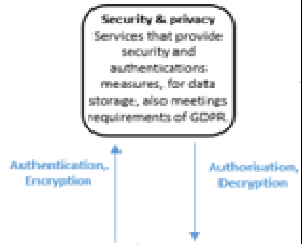
To what extent do you think that this component: (1 very low extent, 2 low extent, 3 neutral, 4 high extent, 5 very high extent)	Your Answer	The average of all responses	Median	What is your answer after viewing other respondents' evaluations?
Can have a technological contribution to the goals of the AI-enabled chatbot.	2	3.5	4	
Includes all relevant considerations that you, as an expert in your field, would expect from an AI-enabled chatbot blueprint.	2	3.38	3	



Q3/10

and to prevent the abuse of information. This component refers to the services needed to provide security and authentication measures for data storage, GDPR and more. <u>This component is applied to all parts of the chatbot.</u>

To what extent do you think that this component: (1 very low extent, 2 low extent, 3 neutral, 4 high extent, 5 very high extent)	Your Answer	The average of all responses	Median	What is your answer after viewing other respondents' evaluations?
Can have a technological contribution to the goals of the AI-enabled chatbot.	3	3.69	4	
Includes all relevant considerations that you, as an expert in your field, would expect from an AI-enabled chatbot blueprint.	3	3.5	4	
Is located in the correct place in the process of the	4	3.38	4	



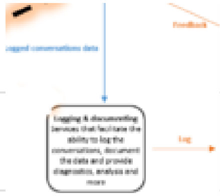
Changes

We added a clarification tha
component is applied to
providing multiple layers c
just the chatbot itself but a
interacts with. Therefore, a
of unintentionally sharing i
potential "data leaks"

a Programme for Research and Innovation (H2020/2014-2020) under grant agreement n° 882850.

his component refers to the need to log the data that the chatbot can collect (e.g., sations, monitoring content on social media, documents provided) to provide diagnostic and analyses, based on privacy by design.

xtent do you think that this component: u extent, 2 low extent, 3 neutral, 4 high extent, 5 very high extent)	Your Answer	The average of all responses	Median	What is your answer after viewing other respondents' evaluations?
ave a technological contribution to the goals of the AI-enabled chatbot.	3	3.81	4	
ll relevant considerations that you, as an : in your field, would expect from an AI-enabled chatbot blueprint.	4	3.81	4	
l in the correct place in the process of the	3	3.88	4	

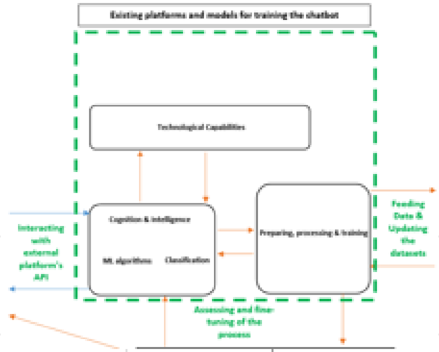


Changes applied:

We added a clarification that the chatbot's blueprint, especially in the logging part, is following privacy by design principles.

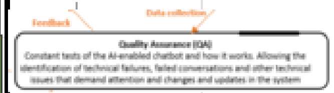
ses: cognition and intelligence, technological capabilities and preparing, processing & ing. AI-enabled chatbots rely on existing language models (e.g., GPT-3, InstructGPT, low), for understanding and generating natural language (e.g., davinci, curie) or images ., Dall-E, Google Vision). This component sets the guidelines on how the chatbot can eract with such platforms' APIs in order to be able to use these models. It allows the ction of the existing knowledge (e.g., datasets) of the chatbot with external capabilities he language models). Alternatively, chatbots can be trained independently, without the connection to external platforms, in a self-developed process.

erring to existing platforms and models ning, as a whole, to what extent do you t this component: (1 very low extent, 2 nt, 3 neutral, 4 high extent, 5 very high extent)	Your Answer	The average of all responses	Median	What is your answer after viewing other respondents' evaluations?
ll relevant considerations that you, as an : in your field, would expect from an AI-enabled chatbot blueprint.	4	3.81	4	



rithms, and how they are implemented. It also includes the QA of the AI itself and the ods and tools available to ensure the AI is ethical and safe to use. Finally, it allows the cation of technical failures, failed conversations and other technical issues that require attention, changes and updates in the system.

xtent do you think that this component: u extent, 2 low extent, 3 neutral, 4 high extent, 5 very high extent)	Your Answer	The average of all responses	Median	What is your answer after viewing other respondents' evaluations?
ll relevant considerations that you, as an : in your field, would expect from an AI-enabled chatbot blueprint.	5	4	4	
l in the correct place in the process of the	5	4	4	



Changes applied:

We added a clarification that QA is achieved both by humans (e.g., fact-checkers, QA workers) and algorithms).

Q7/10

The component of datasets refers to the need for data to train the chatbot and provide the desired information and also to feed manual checks (e.g., fact-checkers). Data is separated into seen data and unseen data (e.g., to assess human evaluations).

To what extent do you think that this component: (1 very low extent, 2 low extent, 3 neutral, 4 high extent, 5 very high extent)	Your Answer	The average of all responses	Median	What is your answer after viewing other respondents' evaluations?
Includes all relevant considerations that you, as an expert in your field, would expect from an AI-enabled chatbot blueprint.	3	3.81	4	

Changes applied:

Datasets are also used for manual checks and are separated into seen data, revealed to those who evaluate the chatbot and unseen data, which serves to assess the evaluation process.

Q8/10

by the chatbot for creating routine reports. It also refers to statistics produced by external methods and sources.

To what extent do you think that this component: (1 very low extent, 2 low extent, 3 neutral, 4 high extent, 5 very high extent)	Your Answer	The average of all responses	Median	What is your answer after viewing other respondents' evaluations?
Can have a technological contribution to the goals of the AI-enabled chatbot.	3	3.5	4	
Includes all relevant considerations that you, as an expert in your field, would expect from an AI-enabled chatbot blueprint.	3	3.5	4	
Is located in the correct place in the process of the	5	3.75	4	

Changes applied:

The component of statistics is applied not just to the internal data of the chatbot but also to external methods and sources.

Q9/10

The component of alerts refers to the ability of the chatbot to generate urgent alerts, whether to the public or the emergency workers, based on its monitoring process. It also refers to alerts produced by external methods and sources.

To what extent do you think that this component: (1 very low extent, 2 low extent, 3 neutral, 4 high extent, 5 very high extent)	Your Answer	The average of all responses	Median	What is your answer after viewing other respondents' evaluations?
Can have a technological contribution to the goals of the AI-enabled chatbot.	3	3.56	4	
Includes all relevant considerations that you, as an expert in your field, would expect from an AI-enabled chatbot blueprint.	3	3.69	4	

Changes applied:

The component of alerts is applied not just to the internal data of the chatbot but also to external methods and sources.

The research leading to these results has received funding from Horizon 2020, the European Union's a Programme for Research and Innovation (H2020/2014-2020) under grant agreement n° 882850.

Q10/10

<u>Manual human monitoring and moderation</u> refers to the need to monitor the activity of the chatbot and its conversations, <u>in addition, and not instead algorithm monitoring</u>, whether in the initial period or all the time, by human operators beyond automatic monitoring.

To what extent do you think that this component: (1 very low extent, 2 low extent, 3 neutral, 4 high extent, 5 very high extent)	Your Answer	The average of all responses	Median	What is your answer after viewing other respondents' evaluations?
Can have a technological contribution to the goals of the AI-enabled chatbot.	4	3.63	4	
Includes all relevant considerations that you, as an expert in your field, would expect from an AI-enabled chatbot blueprint.	4	3.63	4	

Changes applied:

Manual human monitoring occures in addition to algorithm monitoring and not instead of it.

The research leading to these results has received funding from Horizon 2020, the European Union's a Programme for Research and Innovation (H2020/2014-2020) under grant agreement n° 882850.