LAUREA
AMMATTIKORKEAKOULU
Yhdessä enemmän

BETWEEN PEACE AND TECHNOLOGY

A case study on opportunities and responsible design of artificial intelligence in peace technology

Niina Mäki

2020 Laurea

Laurea-ammattikorkeakoulu

Between peace and technology

A case study on opportunities and responsible design of artificial intelligence in peace technology

Niina Mäki
Palvelumuotoilu
Opinnäytetyö
Toukokuu 2020

Laurea-ammattikorkeakoulu **Tiivistelmä**
Palvelumuotoilu
Restonomi (YAMK)

Niina Mäki

Rauhan ja teknologian välissä – Tapaustutkimus tekoälyn mahdollisuuksista ja vastuullisesta muotoilusta rauhanteknologiassa

Vuosi 2020 Sivumäärä 118

Tekoälyä pidetään usein aikamme lupaavimpana teknologiana. Sen eri sovellukset ovat kiinteä osa arkeamme. Tekoälyn liiketoiminnallisista hyödyistä ja vaikutuksista puhutaan paljon. Se voi kuitenkin myös pahentaa olemassa olevia yhteiskunnallisia ongelmia. Sitä, kuinka tekoälyä voisi hyödyntää aikamme suurten haasteiden ratkomisessa, on pohdittu huomattavasti vähemmän. Konfliktien lisääntyminen ja niiden monimutkaistuminen ovat varmasti nykypäivän viheliäisimpiä ongelmia. Samalla keinot rauhan rakentamiseksi ovat jäämässä jälkeen.

Tämän opinnäytetyön tarkoituksena on tutkia tekoälyä hyödyntävän rauhanteknologian tulevaisuuden mahdollisuuksia ja sen vastuullista muotoilua. Kehittämistyö tunnistaa alueita, joissa rauhanrakentamisen tarpeet ja tekoälyn osaaminen kohtaavat, sekä kehittää konkreettisia työkaluja vastuulliselle muotoilulle. Opinnäytetyö vastaa tutkimuskysymyksiin: mitkä ovat tulevaisuuden kehitysalueet, joissa tekoäly voi tukea rauhanrakentamista, ja kuinka rauhanteknologiaa voidaan muotoilla vastuullisesti? Vaikka työ on tapaustutkimus, voi sen tuloksia hyödyntää myös muilla sektoreilla, joilla tekoälyä käytetään kompleksisten yhteiskunnallisten haasteiden ratkomiseen.

Työn teoreettinen viitekehys yhdistää eri aloja. Se muodostuu rauhan- ja konfliktintutkimuksesta, tekoälystä ja sen eettisistä kysymyksistä sekä vastuullisen muotoilun ja kollektiivisen älykkyyden tutkimuksesta. Rauhanteknologian tutkimusta käytetään tulosten luotettavuuden varmistamiseksi. Tutkimusprosessi soveltaa tapaustutkimuksen mallia ja hyödyntää palvelumuotoilun menetelmiä. Se koostuu kirjallisuuskatsauksesta, viidestä työpajasta, 15 temaattisesta asiantuntijahaastattelusta, kolmesta muotoiluluotaimesta sekä teknologiademonstraatiosta. Tutkimuksen materiaalia kerättiin osana Futuricen Rauhankone-vastuullisuusprojektia.

Tulokset osoittavat, että tekoäly voi olla strategisesti tärkeä tuki rauhanrakentamiselle. Sitä voi hyödyntää konfliktien eri vaiheissa ja eri tasoilla. Tutkimus tunnistaa kolme kehittämisaluetta tulevaisuuden sovelluksille: tekoälyavusteinen konfliktianalyysi, ennakointijärjestelmät sekä tuki ihmisten väliselle vuorovaikutukselle. Toisena lopputuloksena kehitetyt kolme muotoilukanvaasia toimivat minimimuistilistana rauhanteknologian tekoälysovellusten vastuulliselle muotoilulle.

Tekoälyä hyödyntävä rauhanteknologia on ihmisen ja koneen yhteistyötä, jonka tulee perustua molempien vahvuuksiin ja selkeään työnjakoon. Muotoilun ihmiskeskeistä lähestymistapaa ja osallistavia metodeja on täydennettävä systeemisen tason työkaluilla. Työ ehdottaa integroitua mallia vastuulliselle muotoilulle. Kehitetyt työkalut auttavat tuomaan yhteen eri alojen osaamista ja osaajia, sijoittamaan tulevaisuuden sovellukset laajempaan ympäristöönsä, ennakoimaan niiden odottamattomiakin vaikutuksia sekä arvioimaan vastuullisuutta mahdollisimman aikaisessa vaiheessa muotoiluprosessia. Toimiessaan rauhan ja teknologian välissä muotoilun on huomioitava jokaisen kontekstin erityisyys ja sen koko ekosysteemi, asetettava ihmiset keskiöön sekä varmistettava sovellusten vastuullinen kehitys ja turvallinen käyttö.

Asiasanat: rauhan rakentaminen, rauhan teknologia, tekoäly, palvelumuotoilu, vastuullinen muotoilu

Laurea University of Applied Sciences **Abstract**
Service Design
Master of Hospitality Management

Niina Mäki

Between Peace and Technology – A Case Study on Opportunities and Responsible Design of Artificial Intelligence in Peace Technology

Year 2020 Pages 118

A growing number of artificial intelligence (AI) applications are currently shaping the human experience in more ways than most realize. There is a lot of discussion about the impacts and business potential of AI. While the development of AI can bring many benefits, it can also further exaggerate the existing social challenges. Despite this, much less attention is paid on how AI could be utilized to tackle those challenges. The increase in number and complexity of violent conflicts is undoubtedly one of the most wicked problems of today. We are experiencing the biggest number of conflicts in decades, yet the existing solutions for building peace are quickly becoming obsolete.

The purpose of this thesis is to understand how AI could be used to support peace. It answers two questions: what are the future opportunity areas for AI in peacebuilding? How can AI peace technology be designed responsibly? The development work identifies where the capabilities of AI and the needs of peacebuilding overlap and develops a set of tools and considerations for responsible design of AI peace technology. While a case study, the results can be applicable also to other sectors where AI is used for solving complex social challenges.

The research and the theoretical framework merge several fields of expertise. Peace and conflict studies support the understanding of the needs and context of peacebuilding. Insights on the capabilities and limitations of AI are combined with ideas on how to carry out responsible design and innovation in a collective intelligence setting and with an intended social impact. Peace technology research is used to draw inspiration and validate the findings.

The case study research process consists of literature review, five workshops, 15 expert interviews, three design probes and a technology demonstration. Service design methods and iteration are used to develop a set of three canvases for responsible design of AI peace technology. The main part of the data collection for the development work was done as part of Futurice's Peace Machine social responsibility project.

This thesis demonstrates that there is potential for AI applications to strategically support the objectives of peacebuilding at different stages of conflict and on various levels. The three future opportunity areas identified are AI assisted conflict analysis, early warning systems and support to human communication. The set of three canvases developed are a checklist of issues that should at least be considered in the responsible design of AI peace technology.

AI peace technology is about collective intelligence: combining the strengths of human and machine. This thesis suggests that the human-centered design approach and use of participatory methods should be complemented by systems-minded tools. An integrated framework of design is proposed. An assembly of experts should work together to both define the problem and to create the solutions. The canvases developed help to consider different levels of impact in advance, to place the solutions into their wider ecosystem and to summarize the issues to consider at an early stage of design. Responsible design done between peace and technology needs to be context specific, place people at its heart, consider the whole ecosystem and cover both the development and the use of future applications.

Keywords: peacebuilding, peace technology, artificial intelligence, service design, responsible design

Acronyms

AI	Artificial Intelligence
EU	European Union
EC	European Commission
HCD	Human-Centered Design
ML	Machine Learning
NGO	Non-Governmental Organization
NLP	Natural Language Processing
SDGs	Sustainable Development Goals
UN	United Nations
WB	World Bank
HQ	Head Quarters
USAID	United States Agency for International Development
VR	Virtual Reality

Table of Contents

1 Introduction

The latest technology has always been used to wage war. Today different applications of artificial intelligence (AI) and of machine learning (ML) in particular, are developing rapidly. They are currently the most influential technologies shaping our time. The increase of AI applications is almost matched by the surge of various types of conflicts. According to Uppsala Conflict Data Program (2019) there were 52 active state-based armed conflicts in the world in 2018. This is the highest number since 1946 (Pettersson et al. 2019).

AI and other emerging technologies, such as blockchain, virtual reality (VR) and robotics, already have a profound impact on all sectors of our society. From the way we consume information to the way data on our behavior is utilized to tailor and target services. AI is widely predicted to bring changes to all sectors of life. Boosting effectiveness of businesses and altering the future of work are among the ones that could disrupt our lives in an even more dramatic way than the transformations prompted by industrialization. (Kelly 2016; Brynjolfsson 2014.)

While the business side of AI applications have received most of the attention in public debate, the potential of AI is also increasingly used for societal impact. AI is utilized to augment human's ability to diagnose and treat illnesses such as cancer (Brynjolfsson 2014, 92-93), to detect posttraumatic stress disorders (Marmar et al. 2019) and to even predict the need for child protection services (Vuolteenaho 2018). While there is much discussion on what the impacts of AI will be on our lives and on our societies, much less attention has been paid on how AI could proactively be used to change them for the better. How could the potential of AI be used to create positive social impacts? Could it help us to solve some of the biggest societal challenges of our time?

From climate crises to growing social polarization – there is a myriad of global challenges that impact people's lives everywhere. Among the most pressing ones is the increasing number and complexity of violent conflicts in the world. New ways are needed to end violence and tackle the reasons behind it. United Nations and the World Bank (2018) estimated that in 2016 more countries were experiencing violent conflicts than at any time during the last three decades. Preventing, managing and resolving conflicts will require using the latest technology to wage peace instead of war.

The emerging field of practice around **peace technology**, or 'peace tech' for short, (Puig Larrauri at al. 2018) is central inspiration for this thesis. How could the technological advancements in the field of AI help to build peace? Furthermore, why isn't the latest and most talked about technology used more for that purpose? What are the issues to consider when

designing AI applications for peacebuilding? How to design AI peace technology? What are the future opportunity areas for that? How about the necessary guidelines and tools when designing for life or death situations?

This thesis begins not from the technological perspective, but from gathering insights on the challenges and everyday realities faced by people involved in peacebuilding efforts. Would it be possible to identify areas where AI could meet their needs and support peace? What tools and methods should be used for what inherently needs to be a process of responsible design? Is it possible to find a balance between the opportunities and risks of using AI in peace technology? What is the role of design when building applications for the most fragile and risky of contexts?

1.1 Background for the thesis

The dramatic increase of conflicts is undoubtedly among the most pressing global challenges today. According to the estimation by the World Bank and UN (2018), there has been a steep increase in the number and complexity of violent conflicts. In 2016 countries experienced most violent conflicts in the last 30 years. The number of battle related deaths is highest in 20 years, the number of interstate conflicts has doubled, and civilian deaths related to conflicts are on the rise. (United Nations & World Bank 2018.)

Prolonged conflicts such as in Syria and in Afghanistan have impacts that are felt even in countries like Finland. Instead of decreasing, the amount of displaced populations and of military spending have surged since the beginning of the century. The economic cost of conflict response in 2012 alone was 9,46 trillion US dollars, which equals 2,4 times the total GDP of Africa. (United Nations & World Bank 2018.) The human cost of conflicts is unmeasurable.

A record number of people are forcibly displaced due to conflicts. An estimated 70,8 million people have been forced to leave their homes (UNHCR 2019). The number of internally displaced persons has increased more than fivefold between 2005 and 2016, while the number of refugees has almost doubled during the same period (United Nations & World Bank 2018). Globally there is both political and business interest to combat different forms of fragility, prevent conflicts and resolve the ones that exist.

The increasing number and complexity of violent conflicts can be described as one the most "**wicked problems**" of our time (Rittel & Webber 1973). There is no simple, nor tested solution. The definition of a wicked problem according to Peters (2017, 388) could also be used to define the messy bundle of causalities and interlinkages that characterize many conflicts:

1.) Wicked problems are difficult to define. There is no definite formulation.
2.) They have no stopping rule.
3.) Solutions are not true or false, but good or bad.
4.) There is no immediate or ultimate test for solutions.
5.) All attempts to find solutions have effects that may not be reversible or forgettable.
6.) These problems have no clear solution.
7.) Every problem is essentially unique.
8.) Every problem may be a symptom of another problem.
9.) There are multiple explanations for the problem.
10.)The planner (policymaker) has no right to be wrong.

The concept of "**super wicked problems**" is used to describe issues with additional challenges. These include having no central authority to manage the problem, having the same actors to both cause and to solve the problem and being in a situation where current focused solutions are becoming obsolete while running out of time to solve the problem. (Peters 2017.) Modern-day conflicts tend to hit many of the characteristics of a wicked and furthermore also of a super wicked problem. For example, many conflicts are fueled by climate change, aggravated by geo-political interests and managed by a fragmented international system including actors that are central to both causing as well as solving them.

Thinking of conflicts as super wicked problems means remedies should meet similar requirements: they need immediate, comprehensive action and system-wide solutions. Understanding complexity is central to analyzing the problem as well as to developing solutions. (Peters 2017, 392.) Dynamic, systems-orientated approach is a prerequisite for tackling a super wicked problem. Here design thinking can be useful as it focuses on coming up with creative ways to solve problems based on a deep understanding of the needs and constraints of end-users and of each unique context and situation. (Mootee 2013, 35.)

Rapidly developing AI applications could help us to better understand complex, super wicked problems such as conflicts. The expansion of AI has been propelled by the combination of available digital data and machines' capacity to analyze it. From mere 25 percent in 2000 to 94 percent in 2007, the majority of all data is now stored in digital form. (Ross 2016, 154.) Analyzing large amounts of data to spot patterns and forecast future trends can be used to supports human's ability to understand complicated dynamics and make smarter, more efficient decisions. The whole field of data visualization has developed around making patterns and trends visible and easier to understand. (Ross 2016, 154-155.)

Could the same data processing ability used for things like testing Barack Obama's presidential campaign messages also be applied to peacebuilding? (Ross 2016, 156.) The AI processing

of data powers recommendation engines in Google, Netflix and Amazon. It is wiping out language barriers as machine translation applications take huge leaps forward. The potential for social impact by simply altering the use of already existing commercial AI applications is impressive. For example, precision agriculture, which is using big data to inform agricultural activities, could help to feed the world's 800 million undernourished. It could decrease the potential for conflict in many fragile, hunger prone areas around the globe. (Ross 2016, 161-162.)

Issues of equality and accessibility are becoming a significant part of technology discussions. Ensuring that digitalization, AI and other new technologies do not magnify the existing social, gender, racial and other divides is vital. The European Commission (EC) White Paper published in February 2020 puts forward priorities to safeguard the trustworthiness and human-centric development of AI through highlighting the importance of transparency, traceability and human oversight that apply particularly to high-risk areas. (European Commission 2020)

Including the concerns of those affected by new technologies should happen even if the technology is developed on the other side of the globe. It should also happen preemptively, before the effects and consequences are evident. (UN 2018a, 8.) What will be the values, principles and responsibilities that guide the design and use of AI applications? How to ensure that with all its promise AI is anchored to human rights and to other normative frameworks and best practices developed? Can human-centered design (HCD) (Gould & Lewis 1985) offer useful insights or tools for this? What should be done to strike a balance between harvesting the benefits of AI in supporting peace and mitigating its risks? What is the role of design and implications for the designer?

In his book "**Peace Machine**" Timo Honkela (2017) suggests that AI and machine learning applications should be used to facilitate human communication, increase understanding between people and through that build more peaceful societies. The idea of applying AI to promote peace has been received with interest and enthusiasm. It merges two very topical and much debated issues – the huge interest in AI and the dramatic need for peace. The compelling aspiration of combining technological potential to the needs of peacebuilding is also mostly overlooked by others. Testing the idea of what "a peace machine" could look like in practice, is key inspiration for this thesis.

For an international development professional global challenges like poverty and inequality are familiar ones. Having lived and worked in some of the poorest countries in the world means also having a distinctive personal perspective on the opportunities and challenges of building technical solutions for those contexts. The thesis writer's personal position and po-

tential for bias should also be noted. Simultaneously 15 years of experience in the “neighboring” field, as peace and development are closely interlinked, has also greatly helped the research process. Ability to reach out to peace experts, including people working in conflict zones, has been an advantage. At the same time, the learning curve on both peacebuilding and on AI has been a steep one.

The thesis started with a bundle of initial questions on how to combine the two distinctive disciplines of peacebuilding and AI. Could technological advancements in AI unlock new ways to solve conflicts? Could emerging technologies be utilized for both ending violent conflicts as well as tackling some of the issues causing them? Is yes, what are the opportunities for future applications? What is the role of design and designer? How to make sure that the future applications in this sector are designed in a responsible manner?

1.2 Goals and research questions

The purpose of this thesis is to understand the opportunities and limitations that AI could have in supporting peace. It seeks to identify areas where technological development of AI and needs of peacebuilding meet and overlap.

The first objective of this thesis is to **identify future opportunity areas where AI could be applied to strategically support peacebuilding.** For this purpose, a research process consisting of literature review, workshops, expert interviews, design probes and a technology demo is carried out.

Secondly, based on the understanding and insights obtained during the research process, this case study will **develop tools and considerations for responsible design of AI peace technology**. In order to complete the development task, an iterative process using design mindset and methods is used.

To achieve the two objectives set out for this study, the thesis will answer the following two questions:

1.) **What are the future opportunity areas for AI in peacebuilding?**
2.) **How can AI peace technology be designed responsibly?**

In order to answer the first of the study questions, insights on the realities and needs of peacebuilding is acquired. Simultaneously, an understanding of the issues that AI is good at solving is needed. Gathering insight from both sectors and merging them in analysis seeks to answer the first research question. For that purpose, the following areas are investigated in a

case study setting: what are the most pressing needs in the field of peacebuilding? What are the current areas that AI is good at? What are the ethical and safety issues to consider from both fields? The result of this development task is identification of a set of future opportunity areas where the needs of peacebuilding and the capabilities of AI overlap.

Based on the insights gathered during the research process, formulating a set of design tools and considerations seeks to answer the second research question. In this task, understanding the context, needs, motivations and limitations of potential customers, end-users and beneficiaries as well as those involved in the design process is central. In order to answer the second study question, the following areas are explored in a case study setting: What are the issues that should at least be considered in the design of AI peace technology applications? What kind of methods and tools could be useful when designing AI for highly complex and fragile environments?

On a general level, the development task of this case study tests the usefulness of design-thinking and tools. When developing peace technology solutions utilizing AI, what are the necessary tools and considerations for responsible design? How to design when including the end-user or intended beneficiary to the process might be difficult? How can the design process be informed by the existing good practices and frameworks from both AI and peacebuilding? How to predict potential impacts at multiple levels?

The design tools developed are used to summarize main characteristics of the future opportunity areas identified. While this is a case study, the potential for the same considerations and tools to be relevant also for design of AI applications with intended social impact in other sectors is discussed.

1.3 Key concepts

Both fields of AI and peacebuilding are impregnated with terminology much of which is evolving or even ambiguous on their definitions. This thesis uses responsible design as an entry point to look at peace technology, a developing field combining the two – emerging technologies and peacebuilding. Operating in the intersection of distinctive yet interlinked fields of expertise means tackling elusiveness and ambiguity when defining the key concepts for this study.

The conceptual choices always tend to be a compromise and this thesis is no exception. Terms are chosen based on ease of communication and their ability to break the silos between the above-mentioned disciplines. It was important to use concepts that are inclusive and understandable also to people that are not experts on neither AI nor peacebuilding.

Making sure the thesis is asking the right questions before suggesting technology solutions was central. Understanding how to describe the needs and realities of people working in peacebuilding or living in conflict areas in a way that would translate to technology development was important. What are the most important challenges for them? Identifying the relevant concepts and insights on AI was equally pertinent. What is AI currently good at doing? A common, communicable understanding of the main concepts was a prerequisite for carrying out the analysis: identifying areas that answer to real needs and that are suitable for AI to solve.

Peacebuilding

This thesis uses the term **peacebuilding** to refer to the holistic and complementary mix of strategic interventions, in all stages and at all levels, to prevent, stop, manage or mitigate violent conflicts. The term peacebuilding thus covers both the efforts to end violence as well as the actions tackling issues that create the need for violence.

The definition used in this development work echoes the United Nations Secretary-General's Policy Committee (2007, 5) definition that states: *"Peacebuilding involves a range of measures targeted to reduce the risk of lapsing or relapsing into conflict by strengthening national capacities at all levels for conflict management, and to lay the foundations for sustainable peace and development."*

The academic discipline of peace and conflict studies concentrates on **conflict resolution** and interventions at different stages of conflict. However, it can neglect to include activities like state building or socio-economic development that are key to sustainable peace. (Liesinen 2018, 23-28; Ramsbotham et al. 2018, 5-15) In the thesis peacebuilding is understood to cover activities that address both the symptoms and the causes of conflict, including structural issues and dynamic relationships between the conflict parties. It is seen as *"broader policy framework that strengthens the synergy among the related efforts of conflict prevention, peacemaking, peacekeeping, recovery and development"* (Alliance for Peacebuilding 2013).

The holistic approach of this thesis to peacebuilding means drawing on different disciplines that have to do with peace and conflict as well as with sustainable societal development that all form a part of the continuum from conflict prevention to sustainable peace.

Artificial intelligence

Artificial intelligence (AI) is a broad term often used to refer to a wide variety of different technological subfields ranging from general issues like data processing, language or learning

to very specific tasks such as self-driving cars or composing poetry (Russell & Norvig 2016, 1). The Encyclopedia Britannica defines AI as *"the ability of a digital computer or computer-controlled robot to perform tasks commonly associated with intelligent beings. The term is frequently applied to the project of developing systems endowed with the intellectual processes characteristic of humans, such as the ability to reason, discover meaning, generalize, or learn from past experience* (Copeland 2018)."

In the fast-evolving field of AI there is no singular, clear-cut definition for artificial intelligence. It rather seems to escape or be somewhat inconsequential even for AI researchers. Key properties that are often used to describe AI include 1.) autonomy or "the ability to perform tasks in complex environments without constant guidance by a user" and 2.) adaptivity or "the ability to improve performance by learning from experience" (Elements of AI 2019). These are also the most important characteristics of AI for the purpose of this thesis.

Much of what is labeled as "AI" in public discussions falls under the category of **machine learning (ML),** or *ability of machine systems to adapt to new circumstances, detect and extrapolate patterns and improve its performance with more experience or data* (Russell & Norvig 2016, 2). In this thesis ML and other types of AI subfields are referred to simply as AI.

The understanding of AI in this development work follows the classic encyclopedia definition focusing on autonomy and adaptivity as key characteristics. In the text "machine analysis" or "machine" are also used interchangeably with AI, as per the language used in the data collection workshops. Distinction between ML and other AI technologies is only made if it bears significance to the results.

Peace technology

Peace technology is widely used as a term referring to the multitude of ways technology can support efforts to build peace. Elusiveness of definitions for peace technology are both accurate and symptomatic. **Peace technology**, as per the definition put forward by Stanford Peace Innovation Lab is "*is fundamentally mediating technology — it "mediates" our engagement with each other. That is, it acts as an intervening agent, augmenting our ability to engage positively with others*" (Quihuis et al. 2015).

In this study the focus is on peace technology that utilizes AI and is designed to strategically support peacebuilding outcomes. However, blindly adopting technological applications without accounting for local context can have the opposite effect of worsening conflicts. (Miklian & Hoelscher 2017) In this thesis, peace technology is understood in a holistic sense. In line with Miklian & Hoelscher (2017) it used to refer to context-specific innovations that utilize

technology (in this case AI) in order to prevent and end conflicts, to tackle reasons behind them and to alleviate human suffering during them.

Responsible design and innovation

When tackling complex societal challenges, such as conflicts, an interdisciplinary and inclusive approach to design and innovations is needed. Von Schomberg (2011, 9) defines **responsible innovation** as *"a transparent, interactive process by which societal actors and innovators become mutually responsive to each other with a view to the (ethical) acceptability, sustainability and societal desirability of the innovation process and its marketable products (in order to allow a proper embedding of scientific and technological advances in our society)."*

According to Stilgoe et al. (2013) key dimensions for responsible innovation process are 1) anticipation (foreseeing risks and dangers), 2) reflexivity (awareness of one's own role, values and responsibilities), 3) inclusiveness (diversity of opinions and public legitimization) and 4) responsiveness (making sure principles, norms and standards have an impact on the way the innovation process is done).

De Hoop et al. (2016) see the strength of responsible innovation in providing methods for real-life innovation processes. They underline the need to always consider not innovating as a possible outcome of a responsible innovation process. Being able to compare and consider different options, their potential impacts as well as the alternative of not moving forward with innovating are particularly relevant to the design of AI peace technology. This thesis considers the principles of responsible innovation to be applicable also to the responsible design of peace technology.

Fostering responsible peace innovations with deeper impact requires incorporating guidance that is contextual, area-specific and conflict-sensitive. As technology and peace and conflict studies advance, opportunities for innovation are created at their intersection. Finding ways for meaningful collaboration that brings together technology experts, peace scholars and innovation researchers could hold great potential for tackling some of today's biggest global challenges. (Miklian & Hoelscher 2017.)

This thesis seeks to both find the opportunities at the junction of AI and peacebuilding as well as to develop tools for meaningful collaboration needed for responsible design. Responsible design is considered as a must rather than optional in the development of AI peace technology innovations.

1.4 Project Peace Machine

The research phase of this thesis was carried out as part of a social responsibility project by Chilicorn Fund program by Futurice. Their “Peace Machine” project made it possible to gather a rich and diverse set of data. The project was inspired by the Peace Machine concept and book of Timo Honkela (2017) describing how AI could potentially be used to ease human communication and, through that, build world peace. However, he does not elaborate on the practicalities of what different peace machine applications would look like.

The Futurice project’s goal was to push Honkela’s inspirational concept further and learn from that process. Peace Machine project took up the challenge to investigate how to get one step closer to what peace machines could look like in real life. The main goal was to see what is possible to do with AI and based on that define what future applications for supporting peace could look like. As the project evolved, the role of responsible design and AI ethics became increasingly significant.

Simultaneously the Futurice project sought to produce open, accessible and easy to understand knowledge on AI as a tool for better understanding of human communication. What is AI’s current ability to analyze human communication? How could that be used to support peace? What is needed to combine the strengths of humans and machines to promote more peaceful societies? What are the future opportunities for real-life peace machines? How to tackle the risks involved?

The Peace Machine project consisted of a series of four workshops and a technology demo utilizing natural language processing (NLP) in a mass-event called “Mitä tapahtuu huomenna kohtaamiselle?” (What happens to encounters tomorrow?). They all were used to gather material for this thesis. As part of the project team the thesis writer was part of planning and carrying out all the workshops based on previously defined headline level titles. The workshop design, including selection of methods, participants and speakers, were defined to serve both the goals of the project and the data collection for the thesis. Four podcasts were recorded after each workshop summarizing their key findings. This supported the iterative nature of the case study’s research phase.

In addition to already planned activities of the Peace Machine project – the four workshops, technology demo and podcasts – an additional workshop was organized to test peace technology canvases created during the previous workshops. Additionally, the data collection for this thesis included 13 thematic expert interviews and three design probes. As the timing was overlapping, their insights were used to inform and complement those arising from the workshops.

1.5 Structure of the thesis

This thesis is divided into six chapters. The first chapter introduces the topic and its background. It also presents the key concepts and defines the goals and research questions for the development work. It explains why the subject was chosen for research and describes the case study context of the thesis.

The second chapter discusses the theoretical framework of the thesis. This is an interdisciplinary development work that merges the fields of artificial intelligence, peace and conflict studies as well as the emerging fields of peace technology and responsible design and innovation. The chapter provides the theoretical background of this case study that is described in more detail in the next chapter. The third chapter explains the case study as a method and the reasons for choosing it, along with the focus and limitations of the thesis.

Chapter four contains the case study process of this development work. It explains comprehensively how the traditional case study is tailored, gives an overview of the process and describes how the phases of design, collect, analyze and share were carried out in this thesis. Even though the case study process is dialogical, the different phases are presented in a linear fashion in order to make them easier to follow for the reader.

The fifth chapter introduces the results of the development work. The three opportunity areas identified are explained and tools developed based on the insights gathered during the development work are introduced. Lastly the learnings and conclusions are presented in the sixth chapter that answers the research questions and reflects on the significance of the findings. Evaluation of this development work as well as prospects for future research are also discussed at the end.

2 Theoretical framework

The theoretical background for this thesis draws from four different fields of expertise, namely from peace and conflict studies, from research on peace technology, on artificial intelligence and on responsible design and innovation. Insights gathered from a desk study as well as from the experts that participated in this development work are combined in the analysis to cover a wide range of insights from different disciplines.

Firstly, the expertise arising from the academic discipline of peace and conflict studies is vital for setting the scene. It is a way to understand the context as well as to evaluate the

needs for the technology solutions that are strategically important for peacebuilding. Secondly, the emerging body of work around peace technology and peace innovations is discussed in more detail for inspiration and for checking the validity of the insights of this case study development work. Thirdly, understanding on the capabilities and limitations of AI as technology is merged with the peacebuilding needs in order to identify the future opportunity areas for AI in peace technology. Lastly, lessons learned from the analysis and insights from desk study on responsible design and innovation in situations of collective intelligence, such as the one under study, are used to inform the development of the tools for responsible design of AI peace technology.

The interest in using AI as a technology to help us tackle complex societal challenges has emerged simultaneously with awareness of the risks involved. When AI is used in fragile environments and situations, a special consideration of risks needs to be a default integrated to the design from ideation to implementation. The evolving discussions on ethics, biases and risks of AI are relevant not just for peace technology, but everywhere where AI is utilized. They also inform this development work.

2.1 Needs and context of peacebuilding

The academic discipline of peace and conflict studies has expanded significantly during the last 25 years. There is an increasing amount of understanding on why conflicts begin and how they can be prevented, managed and resolved. (Ramsbotham et al. 2018, 5; Miklian & Hoelscher 2016.) Ramsbotham et al. (2018, 10) define **conflict** as "*the widest set of circumstances in which conflict parties perceive that they have mutually incompatible goals*".

In their broadest sense conflicts can be considered as a natural part of human coexistence. From very early on, a distinction between destructive and constructive conflicts has been made in peace and conflict studies. While the former is something to avoid, the latter is a necessary condition for human creativity. (Ramsbotham et al. 2018, 9.) Today's peace and conflict studies is a multilevel, multidisciplinary and multicultural field that includes both analytic and normative as well as theoretic and practical approaches (Ramsbotham et al. 2018, 10; Eronen 2018, 123-130).

The emphasis of this development work is on understanding the conflict context and dynamics related to them in order to find potential areas for preventing, managing and resolving violent conflicts. **Peacebuilding** is used as an umbrella term to refer to the different strategic interventions that happen at all stages and at all levels to end violence and tackle issues that create the need for violence. Beyond understanding the context where conflicts happen, also

insights on the ecosystem related to efforts of building peace is necessary. As conflicts develop, the methods and actors of resolving them change. This is the ecosystem in which potential AI peace technology solutions should work in.

A strong humanitarian and business case can be made for **conflict prevention**. Potential net savings of preventing violent conflicts are estimated to range between $5 billion per year to almost $70 billion per year (Mueller 2017). In 2020 there are over 167 million people in over 50 countries in need of humanitarian assistance due to both conflicts and other factors such as natural disasters. The imminent need for funding is estimated to be $28,8 billion. (UNOCHA 2020) The lessons learned from conflicts like Rwanda and Yugoslavia in the 90s, there is now a wide consensus that the best way to deal with conflicts is to prevent them.

The goal for conflict prevention is not to avoid them totally, but rather avert potential conflicts from becoming violent. (Liesinen 2018, 24; Ramsbotham et al. 2018, 144-145.) A comprehensive set of policies and intervention instruments have been developed both in the EU as well as in the UN system that underline the importance of prevention. Both direct measures to avert violent conflict as well as structural measures to cultivate peaceful change are continuously getting more attention in the peace and conflict sector. (Ramsbotham et al. 2018, 153-171.) United Nations and World Bank (2018, 37) also advocate for an integrated approach to prevention that needs to be sustained throughout the conflict phases, to include all levels, approaches and actors and to be targeted to tackle a wide range of risks that can lead to conflict.

The academic discipline of **conflict resolution** looks at all stages of conflict and different approaches to solving them. Galtung (1969) extends the definition of peace from “absence of violence” (Galtung 1969, 168) by separating personal and structural violence. The absence of former is considered negative peace and the absence of latter as positive peace, or “social justice” (Galtung 1969, 183). This closely links the study of peace to both conflict and to development theory. The notion of **negative peace** in terms of stopping violence is complemented with **positive peace** that is restoration of relationships and building of social systems (Galtung 1969, 183-186). This thesis uses the term sustainable peace interchangeably to refer to the dimensions that are also included in the term positive peace.

There are multitude of different ways to illustrate the development of a conflict. Ramsbotham et al. (2018, 16) use an hourglass model to depict conflict phases of containment, settlement and transformation and their corresponding resolution responses. In this model the combination of suitable strategic responses varies according each stage and the political space associated to it. Escalation of conflict is characterized by the narrowing of political space, while the widening of space signifies conflict de-escalation. The ability to act with a

complementary mix of suitable responses maximizes the potential to resolve conflicts and build peace. (Ramsbotham et al. 2018, 16.)

The hourglass model is aligned with the integrated approach to prevention put forward by the United Nations and World Bank (Mueller 2017, 2017, 39). Figure 1 presents the two models combined illustrating the collaboration of different actors and actions needed throughout the conflict phases.

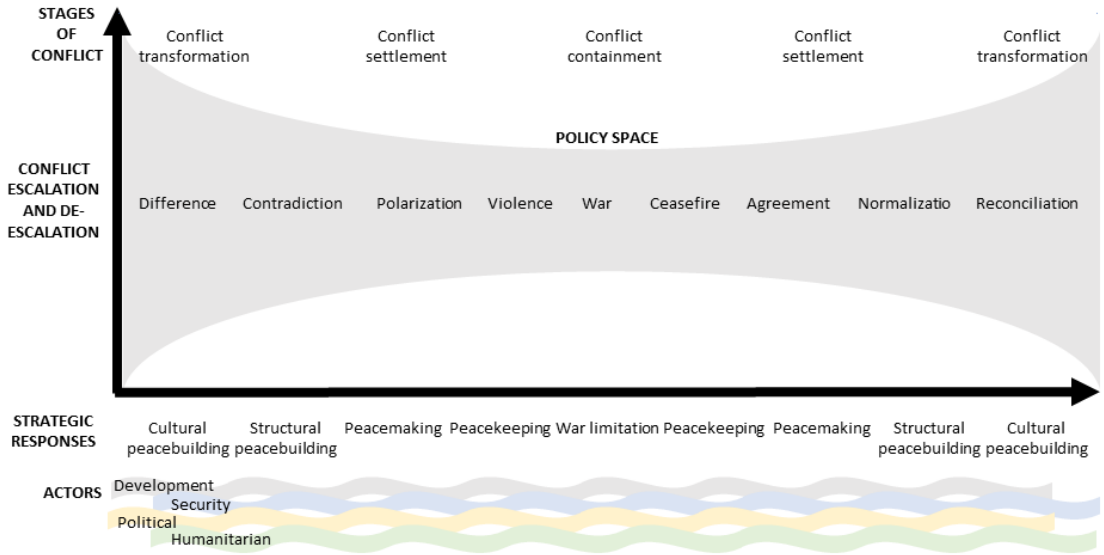


Figure 1: The hourglass model of conflict stages combined with the integrated approach to conflict prevention (Ramsbotham et al. 2018, 16; Mueller 2017, 39)

While there are different perspectives to conflict resolution, common features among them include considering **modern conflicts** as complex, dynamic and often asymmetric situations that can develop rapidly and escalate unpredictably due to emergence of new actors or internal power struggles among the different parties involved. The appearance of secondary conflicts can further complicate the situation. (Ramsbotham at al. 2018, 28-33.) Prolonged, fragmented and mutating conflicts, that include or lead to secondary conflicts and situations of severe humanitarian crises, meet many of the characteristics of “super wicked problems” as per definition by Peters (2017).

For the purpose of this development work a holistic approach on **peacebuilding** is chosen. According to Galtung (1976) the path towards modern thinking on peacebuilding is encapsulated by the notion that it is not enough to stop the violence but there needs to be efforts to address the reasons causing the need for violence. Ramsbotham et al. (2018, 266-285) see peacebuilding being at the heart of conflict resolution. The modern comprehensive approach

to peacebuilding encompasses all conflict levels and dimensions from responses by the international institutions and their normative frameworks to local level approaches such as peace education and engaging with communities.

According to the United Nations Peacekeeping Operations Principles and Guidelines (2008) peacebuilding "*involves a range of measures aimed at reducing the risk of lapsing or relapsing into conflict, by strengthening national capacities for conflict management, and laying the foundations for sustainable peace. It is a complex, long-term process aimed at creating the necessary conditions for positive and sustainable peace by addressing the deep-rooted structural causes of violent conflict in a comprehensive manner.*" (United Nations 2008, 18.)

Thus peacebuilding, as understood in this thesis, is in addition to conflict prevention, management and resolution closely linked to functioning society and the collaborative efforts of different actors including state, private sector actors, NGOs and international institutions and partners like different UN agencies and the EU. Peace must be built at large. In this thesis peacebuilding is seen as a holistic combination of processes and actions at all levels aimed to reduce the potential for lapsing or relapsing into conflict (United Nations 2010, 5).

Peacebuilding is as much about what is done as it is about how the activities are carried out (United Nations 2010,14). The need for **conflict sensitivity** – making sure that no action increases tensions or creates new ones – is a must when working in conflict or fragile situations. For this one must understand the potential impact that an intervention could have in the specific, intended context. (Doty 2016, 71-77; United Nations 2010, 14.) OECD predicts that at this rate the majority of people living in poverty will also be living in countries with high levels of violence (OECD 2015). It is thus critical to understand how peace and development processes interact with security, political, and human rights tools. (United Nations & World Bank 2018).

As AI applications could strategically support peacebuilding outcomes in any sector, all solutions must inherently be built using the lenses of conflict sensitivity as a default and point of departure. A comprehensive approach to conflicts also means that different interventions and instruments should coordinate and complement each other to support peace efforts at different levels. **In this thesis a conflict prevention, management and resolution are treated holistically as different aspects of peacebuilding.** The term peacebuilding is used to cover interventions at all conflict phases and to also include development activities that aim to build resilience and tackle the root causes of conflicts.

2.2 Capabilities and limitations of artificial intelligence

The foundations of **artificial intelligence (AI)** as a field of science can be found in disciplines like mathematics, economics, neuroscience, psychology, computer engineering and linguistics. The birth of AI, including the term artificial intelligence, is attributed to John McCarthy and a two-month workshop in Dartmouth College in summer 1956. Things now considered classics in the field - like the Turing test originating from the 1950's and the chess match of IBM's Deep Blue beating Garry Kasparov in the 1990s - have been influential in the current debate on what is considered as AI. (McCarthy et al. 1955, Russell & Norvig 2016, 5-18; Kaplan & Haenlein 2019, 17-18.) However, instead of scientific discipline, this thesis prioritizes a functional approach to understanding the phenomenon that is AI.

In 2019 the European Commission instituted Independent High-Level Expert Group on Artificial Intelligence (2019a, 6) put forward an extended the definition of Artificial intelligence as *"software (and possibly also hardware) systems designed by humans that, given a complex goal, act in the physical or digital dimension by perceiving their environment through data acquisition, interpreting the collected structured or unstructured data, reasoning on the knowledge, or processing the information, derived from this data and deciding the best action(s) to take to achieve the given goal."*

In many cases more compact definitions of AI are used. It is seen to be both about science and about technology that creates **intelligent systems** that can analyze their environment, learn and adjust to new information by using algorithms and take actions with a degree of autonomy to achieve specific goals and tasks. AI systems can be purely **software**, such as voice assistants or face and speech recognition, or embedded in **hardware** such as autonomous cars or robotics (SAS 2018; European Commission 2018; European Commission 2019, 1; Kaplan & Haenlein 2019.)

Characteristics of artificial intelligence systems

Despite the exponential growth in interest, applications and research, the definition of AI remains surprisingly fuzzy and to some even irrelevant (Kaplan & Haenlein 2019, 17). While there is no single universally accepted definition, common key characteristics of AI include the ability to perceive their environment through some type of sensors, collect, process and interpret data and based on this decide the best action. AI systems learn from the past experiences and adapt their behavior as per the analysis of how the previous actions have affected the environment. This is done in order to achieve a defined goal or to perform a task without constant guidance from human user. (Elements of AI 2019; European Commission 2019; Kaplan & Haenlein 2019.)

The content of components in an AI system varies significantly. The **sensors** used for perceiving the environment can include almost anything according to what type of data is relevant for achieving the goal given to the AI system by its designer. The **data** that is used for reasoning and making the decisions can be structured (like databases) or unstructured (like images) in format. The decision made by AI systems can have various degrees of **autonomy** from totally automatic to selection of recommendations for human users to decide on. To act on the decisions, **actuators** such as chatbot producing responses, are needed. As this can modify the environment, the sensing and sensemaking is used again **to learn from the actions and to adapt** the reasoning in order to achieve the task given to the system. (European Commission 2019, 2-3.)

Kaplan & Haenlein (2019, 18-20) classify AI systems into three groups: analytical AI, human-inspired AI and humanized AI. The first two can loosely be aligned with the approaches identified by Russell & Norvig (2016, 2): the human-inspired AI systems (the category of thinking and acting humanly) and the analytical AI (thinking and acting rationally). The analytical AI systems use representations of their environment and learn from past experiences to inform future decisions. The human inspired AI systems can include elements of emotional intelligence and can thus consider human emotions in their decision-making. Applications of the former group include self-driving cars and fraud-detection systems while applications of the latter are used in things like customer interaction and recruitment systems. (Kaplan & Haenlein 2019, 18-20.) The last group of humanized AI is largely still at the earliest stages of development.

The classification of AI systems based on cognitive, emotional and social competencies used by Kaplan and Haenlein (2019, 18) leaves out expert systems (like the Deep Blue that beat Kasparov) that are based on if-then type of rule collections programmed by humans. This top-down approach lacks the autonomous learning and adaptation capabilities central to defining what AI is and what it is not. The bottom-up approach characterized by autonomy and adaptivity is often used to describe what is nowadays considered as AI. (Russell & Norvig 2016, 2; Kaplan & Haenlein 2019, 18.)

The ability to learn from past data is common to all AI systems. It is not a surprise that much of what is labeled as "AI" in public discussion falls under the category of **machine learning (ML).** It is characterized by the ability of systems to adapt, detect patterns and improve performance through experience (Russell & Norvig 2016, 2). ML is a wide array of methods that can be used to train computers to learn by identifying patterns, structures or irregularities

from datasets. Using data to make predictions and automate decisions are among key features of ML systems. Finding patterns in training data and analyzing them into a model enables the system to make predictions also for new, unseen data. (USAID 2018, 10-12).

The processes of machine learning can be grouped into three categories. Firstly, **supervised learning**, that can use things like simple classification trees to things like more complex neural networks to do tasks such as image recognition. (Kaplan & Haenlein 2019, 19; Russell & Norvig 2016, 693-697.) Secondly, **unsupervised learning** like cluster analysis is utilized to carry out tasks like speech recognition. Here the need to trust the AI system is greater as only the input, but not the output is defined in advance. Thirdly, **reinforcement learning** is used in tasks like playing games or selecting social media content to maximize clicks.

The applications of ML stretch from analyzing large amounts of unstructured data, such as text, images or audio to finding nonlinear and complex relationships (Kaplan & Haenlein 2019, 19; Russell & Norvig 2016, 830-851; USAID 2018, 10). In this thesis, ML and other subfields of AI are all referred to as AI or "machine analysis" reflecting the language used in case study data collection workshops.

Impacts and ethical considerations of artificial intelligence

Currently various applications of AI are an integral part of what many consider everyday life. From planning travel routes to making predictions for users' convenience all the way to self-driving vehicles and chatbots - AI applications are shaping human experiences in more ways than most of us even realize. (USAID 2018, 10-12). From Google Translate to personal assistants like Alexa and recommendations systems of Netflix, the vastly growing number of different AI applications has integrated digitalization and data-driven services into the daily reality of millions of people (Berditchevskaia & Baeck 2020, 8).

Due to the combination of rapid technological advancements and the surge of real-life applications, many describe the impacts of AI as **the Fourth Industrial Revolution.** Preceded by the transformations induced by steam power, electricity and digitalization, the term illustrates the current transformational point where AI (combined with robotics) have triggered an unprecedented, global systems level change that affects the way we live. (Schwab 2016, Skilton & Hovsepian 2018.)

In 2000 less than 30 percent of the world's 700 million mobile subscriptions were in developing countries. In 2012 there were a total of 6 billion subscriptions, 75 percent of which in the developing world. (Brynjolfsson 2014, 95). Cheap parallel computation and big data coupled

with increased data storage capacity, computation power, expansion of wireless communications and improved algorithms are among the factors behind the expansion of AI. Even though calculation and identification of patterns is only one stream of AI, the capacity to analyze huge amounts of data combined with the independent learning abilities have been central for the increase of AI applications. (Boyd & Holton 2018, Kelly 2016, 38-39.)

While offering movie or music recommendations are more of a convenience, things like determining people's credit worthiness, automating online recruitment (Lee et al. 2019) or assigning risk scores that can influence detention and prison sentences (Angwin et al. 2016; Corbett-Davies et al. 2017) and recommending less care for patients of color (Obermeyer 2019) all have critical and very tangible real-life implications. If majority of machine-human interaction continues to be thought as "nerdly narrow, super smart specialists" (Kelly 2016, 42), there is a risk of having limited means to include considerations of social issues, societal contexts or potential biases arising from them.

Many at least partially already automated businesses are quickly consolidating the effectiveness gains of AI. A Global PwC study (2017a) estimates that AI contribution to the global economy could be as much as additional $15,7 trillion with biggest gains predicted to be in China and in North America. Changes are expected to be fast in sectors like retail, financial services, health care, transportation and manufacturing (PwC 2017a). In less than 30 years most truck drivers are not going to be human (Kelly 2016, 50). Practically any information intensive job could be automated. The overall employment vulnerability in countries like the UK, Japan, USA and Australia is predicted to be as high as 35–50 percent. (Boyd & Holton 2018, World Economic Forum 2016.)

Until recently aspects like utilizing the technological advancements to augment productivity and create business have tended to dominate the public discourse on AI. Addressing the ethical questions related to AI and its applications is, however, a quickly growing body of work. Questions around human-machine interaction, algorithmic and other biases, human rights issues, transparency, equality, fairness and democratization of AI are becoming more researched and relevant than ever. (Boyd & Holton 2018; Lee et al. 2019; Hadhazy 2017; Montes & Goertzel 2019; Dastin 2018; Corbett-Davies et al. 2017.)

Towards responsible and equitable AI

When AI applications are permeating all aspects of human existence, the risks are quickly becoming evident: potential to increase marginalization, inequality and breaches in personal and data security are among the identified dangers (Schwab 2016). There is a growing under-

standing on how **AI can further exaggerate the existing social challenges and power structures.** Globally it could create an even wider gap between the winners and losers, the haves and the have-nots. (United Nations 2018b).

Calls for AI to create benefits for all, to citizens, businesses and public interests alike are echoed by United Nations, the EU (European Commission 2020) and private actors like PwC (2017b) and IBM (2018) that all have their own white papers and toolkits for responsible AI or principles for trust and transparency. Common issues in them include human-centered approach to AI, trustworthiness, transparency and explainability of AI systems and data ownership, accountability and governance.

A systematic approach to **responsible AI** is adopted internationally through multilateral platforms like OECD with their Principles on AI (2019) and The Ethics Guidelines for Trustworthy AI as defined by Independent High-Level Expert Group on Artificial Intelligence (2019b). Included in the EU White Paper on Artificial Intelligence (2020) and piloted by hundreds of companies for feedback, the Ethics Guidelines put forward seven requirements for trustworthy AI. They are 1) Human agency and oversight, 2) Technical robustness and safety, 3) Privacy and data governance, 4) Transparency, 5) Diversity, non-discrimination and fairness 6) Societal and environmental wellbeing, and 7) Accountability.

Global pandemics, discoveries of electoral interference or data misuse underline that the ethical challenges of AI are universal to all countries and people. The interplay between international and national level regulations and voluntary frameworks for ethical AI and the fragmented development of AI siloed in mega-corporations (Montes & Goertzel 2019) is likely to continue. Several countries are formulating national AI strategies: Canada, China, South Korea and the United Arab Emirates have all released strategies for promotion, use and development of AI. (Dutton 2018)

Finland claims to be in the forefront and was among the first countries to release a national AI strategy in 2017. (Työ- ja elinkeinoministeriö 2017; Natalucci 2018). Finland has also announced cooperation with France for ethical use of AI, launched an AI ethics challenge for Finnish enterprises and hosted a high-level AI Forum in 2018 and a high-level conference on AI and human rights in 2019 (Council of Europe 2019; AI Finland 2018). However, even globally the discussion on how to use AI to develop new, unseen solutions for complex social challenges while safeguarding the ethical principles is very much still in its early stages.

2.3 Strategic and context specific peace technology

Peace technology is generally used to describe the multitude of ways technology can support peacebuilding outcomes. The Stanford Peace Innovation Lab defines **peace technology** in terms of its fundamental capacity to mediate people's engagement with each other. It is technology that acts *"as an intervening agent augmenting our ability to engage positively with others"* (Quihuis et al. 2015). Technology can have an important role in preventing, stopping or managing conflicts. At the same time, it can also have negative effects on conflict dynamics that need to be included in the discussion about peace technology (Dahl Jensen & Amnebjer 2019).

The potential of using emerging technologies for social impact is quickly being recognized. Using AI in the field of international development (USAID 2018), applying AI for social good (Chui et al. 2018) or for humanitarian aid (Johnson 2018) and conflict prevention (Letouzé et al. 2013) are among the examples for this. The UN Secretary General's Strategy for New Technologies (United Nations 2018a) urges the United Nations to systematically engage and support the use of emerging technologies for achieving the Sustainable Development Goals (SDGs).

Whether it is finding solutions for climate change, accelerating the achievement of SDGs, predicting the damage caused by extreme weather conditions, detecting hate-speech or forecasting migrations flows, using AI for social impact is quickly becoming something encompassing almost everything, so that it is in danger of explaining nearly nothing. The same risk applies to the relatively recent term of 'peacetech' that according to British Council (2016, 3) emerged only in 2015.

Outside military and security sectors (excluded from the scope of this thesis), the very idea of applying AI to peacebuilding is a recent one. While behind the development and humanitarian sectors, there is an increasing understanding about the dynamics of peace and conflict as well as growing involvement of technology and innovation actors to the sector. (British Council 2018) Thus, building synergies between the two – the umbrellas of technology and peacebuilding - has become an area of growing interest. Fostering pro-peace innovations by bringing together peace and conflict scholars and technology and innovation experts is seen as a promising avenue for developing new solutions to support peacebuilding efforts. (Miklian & Hoelscher 2016)

Dahl Jensen and Amnebjer (2019, 5) define **peace technology as an emerging body of work that aims to build or sustain peace in a given context**. In addition to technological components, for something to be considered peace technology, it must be of **strategic importance**

to peacebuilding objectives. (British Council 2016) This distinguishes peace technology from using technology for things like organizational or project management. The distinction between non-strategic and strategic uses of technology also links peace technology firmly to innovation. It is not just a new word for something that is already being done. (Puig Larrauri 2018)

According to British Council (2018, 4) technology serves three key functions in peacebuilding. Namely they are: 1) **data management** or collecting, analyzing and visualizing information, 2) **strategic communications** or engaging more people into conversations about peace and 3) **dialogue and mobilization** or creating new spaces for a broader variety of peace actors and actions. Quihuis et al. (2015) further define the four interconnected sub-components of peace technology to be sensors to measure human engagement, communications like Wi-Fi or phone lines, computation and lastly actuators that allow responding to the information detected by the first three components.

Some concentrate on the role of individuals and highlight the importance of peace technology triggering new behavior **at the level of human actions** (Quihuis et al. 2015). Others focus on supporting actions **at the level of policies and institutions** (Dahl Jensen & Amnebjer 2019; British Council 2018). Being able to detect and measure is central to both approaches. The existence of peace technology sub-components or functions does not, however, make any technology inherently good or bad, 'peace' or 'war'. Technology is a tool, the use of which should be evaluated critically as the very same technology can be applied to also attain the opposite objectives (Puig Larrauri 2018).

Peace technology can amplify the scope and number of different actors involved in building peace. It also helps in bringing together the peace and conflict and technology communities, to form new partnerships and collaborations across sectors. (Cottray & Puig Larrauri 2017; British Council 2018, 4.) The combination of technical and social disciplines, bringing together both formal and informal peace actors, experts, scientists, designers and developers remains a central component for success of peace technology. (Dahl Jensen & Amnebjer 2019; Miklian & Hoelscher 2016)

Existing peace technology applications include examples like the hate-speech lexicons (Peace Tech Lab 2020) that identify inflammatory expressions used in conflict countries' social media, the Redirect Method that is used to target violent extremism among potential ISIS recruiters (The Redirect Method 2020; Dahl Jensen & Amnebjer 2019, 25) and the Ground Truth Global (2020) that uses a combination of data from various sources to predict increase in conflict potential. The last example demonstrates how peace technology could be applied to commercial use.

In addition to conflict related uses, peace technology can be relevant in situations of endemic violence (e.g. mega-cities) and in creating conditions for sustainable development (e.g. creating sound business environments) (Miklian & Hoelscher 2017, 4; Dahl Jensen & Amnebjer 2019, 9-18). Similarly, commercial applications could be used as peace technology. At practical level assigning a dual use for existing commercial applications as peace technology carries the same risks as using peace technology solutions for example to military or intelligence purposes. For peace technology applications a failed roll-out or testing could mean not just a wrinkle in the business, but imprisonment or even death of users. (Miklian and Hoelscher 2017; Cottray & Puig Larrauri 2017).

Adopting technological applications to support peacebuilding efforts without accounting for ethical considerations, existing good practices or specific needs arising from local contexts can lead to the opposite effect: worsening of conflicts. (Miklian & Hoelscher 2017) Particularly when working in conflict situations conflict sensitivity is key. Following existing normative criteria by the United Nations as well as good practice needs to be incorporated into all actions. For the peacebuilding field things like Professional Standards for Protection work (International Committee of The Red Cross 2018), Human Rights Approach to Information During Crises (Signal Code 2017) and Humanitarian Charter and Minimum Standards in Humanitarian Response (Sphere 2018) are among the relevant ones. In addition to the EU Ethics Guidelines for AI (2020) the criteria to consider from the AI field include Principles for Digital Development (2015), and Ethical Guidelines for Peace Tech (JustPeace Lab 2017).

Peace technology must always be designed with the utmost awareness of the local context and existing power structures. When designing for life-or-death situations questions of context, power and ethics cannot be an afterthought, but a mandatory consideration permeating whole development process. (Miklian and Hoelscher 2017) However, according to Miklian and Hoelscher (2017, 7) there is lack of tools to preemptively consider the wider context and consequences, including the unintended ones, of peace technology.

2.4 Human-centered and systems-minded responsible design

Tackling societal challenges and “wicked problems” is not possible without an inclusive and interdisciplinary approach to design and innovation. A participatory multi-stakeholder process taking in a variety of interconnected perspectives and dimensions is needed. The question is how to do this in concrete terms? According to Mulgan (2018, 2-4) complex, collective problems require collective solutions, or using technology to link up people and machines to magnify the intelligence needed for the task at hand. AI itself can help to combine dispersed inputs from large groups of people. (Elia & Margherita 2018, 279-284).

As per definition, peace technology is closely linked with the idea of **collective intelligence** that Mulgan (2018, 237) defines to be "*the capacity of groups to make good decisions – to choose what to do, and who to do it with – through a combination of human and machine capabilities*". Levy (2005, 191) emphasizes that reality is always a collective creation, but due to the internet our thinking is now more apparently part of the same network. Instead of narrow smartness, a deeper understanding of the big picture than an individual or single organization is required to make sure that in building solutions potential errors are not amplified alongside with gains. (Mulgan 2018, 4.)

While Malone et al. (2010, 24) boil systems of **collective intelligence** down to elements of create and decide, Mulgan (2018) makes the distinction of using technologies to enhance the capacity to **observe, analyze, recollect and even to empathize** to be at the heart of collective intelligence. Elia and Margherita (2018, 281-284) highlight the need for solving complex problems through an integrated process of collaboration that is comprised of problem identification, analysis and modelling, solution definition, prototyping and solution implementation and maintenance that should happen within a community of both expert and non-expert participants. Thinking on collective intelligence should be utilized to inform the organization of the innovation process in situations where both capabilities of groups of people and of machines are needed.

Von Schomberg (2011, 9) uses the term **responsible innovation** to describe *"a transparent, interactive process by which societal actors and innovators become mutually responsive to each other with a view to the (ethical) acceptability, sustainability and societal desirability of the innovation process and its marketable products (in order to allow a proper embedding of scientific and technological advances in our society)."*

De Hoop et al. (2016) see the strength of responsible innovation in providing methods for real-life innovation processes. For them responsible innovation is about *"innovating responsibly – or not innovating at all"*. For design of AI peace technology, this is particularly pertinent criteria. Ability to compare and consider different options, their potential impacts as well as the alternative of not moving forward with the innovation is central. Rather than choosing a responsible approach to design, in peace technology it should be considered a default. **It can mean that sometimes the most responsible option is not designing at all.**

One of the key questions, however, is defining when to move forward with the design of peace technology. **The principles relevant to responsible innovation are considered applicable to the responsible design of peace technology** in this development work**.** In conflict and fragile environments, a responsible design process should always start with questions of

context, power relations, different interest groups and values, inclusive communication and stakeholder dialogue (Blok 2014, Macnaghten et al. 2014). In addition to taking context specificity seriously, finding concrete ways to preemptively consider potential impacts is essential.

According to Schlaile et al. (2018, 8-9) the research on responsible innovation is a field under development. The goal is to **make both the innovation processes and their outcomes more responsible**. The four key dimensions of responsible innovation process as per Stilgoe et al. (2013, 1570-1573) are all relevant to designing of AI applications for peacebuilding: **1) anticipation** (foreseeing risks and dangers), **2) reflexivity** (awareness of one's own role, values and responsibilities), **3) inclusiveness** (diversity of opinions and public legitimization) and **4) responsiveness** (making sure principles, norms and standards have an impact on the way the innovation process is done).

Fabian & Fabricant (2014) suggest an ethical framework for innovation to be humanistic, non-hierarchical, participatory, and sustainable. Designing with (not for) real people is central. The elements of ethical framework and of responsible innovation process are echoed in the Principles for Digital Development (2015). They are intended to help integrate best practices into technology-enabled programs and projects. The nine principles are a product of years of co-creation and now endorsed by over 200 organizations:

1. **Design with the user:** use user-centric design methods to get to know the people
2. **Understand the existing ecosystem:** consider the particular structures and needs of the country, region and community you are designing for
3. **Design for scale:** secure funding and partners to take pilot initiatives beyond their original contexts
4. **Build for sustainability:** sustainability of programs, platforms and digital tools is needed to maintain user and stakeholder support key to long-term impact
5. **Be data driven:** quality information should be made available to the right people when they need it to take action
6. **Use open standards, open data, open source, and open innovation:** An open approach enhances collaboration, maximizes resources and avoids duplication of work
7. **Reuse and improve:** don't start from zero, but take the work of global community as a starting point in order to go further than any organization go could alone
8. **Address privacy & security:** carefully consider which data is collected and how it is acquired, used, stored and shared

Though trickier than with international development projects, as technology and peace and conflict studies advance, opportunities for innovation are created at their intersection. Finding ways for meaningful collaboration of technology experts, peace scholars and innovation

researchers could hold great potential for tackling some of today's biggest global challenges. (Miklian & Hoelscher 2017) Designing social innovations can be as complex and messy as the wicked problems they are intended to solve. Instead of straightforward, clearly defined linear processes, innovation tends to happen in squiggles with dead-ends and unclear accountabilities of the numerous actors involved. (Owen at al. 2013, 33.)

The human-centered approach, often the starting point in design processes, has its shortcomings when applied to social multifaceted challenges (Both 2018). In AI peace technology design, traditional participatory methods (e.g. workshops or interviews) might not be an option as the intended end-users and key stakeholders could be party in the conflict the technology is hoping to help resolve. Understanding relationships, considering how to act in an ecosystem of expert and non-expert stakeholders are key components of the design process. Additionally, to be able to preemptively recognize short and long-term impacts within the conditions of a particular social system, a combination of human and systems-level approaches is needed. (Both 2018; Elia & Margherita 2018.)

Both (2018) suggests an integrated **framework of human-centered and systems-minded design tools and methods** that complement each other. Moving between the two design approaches means one mindset and methods can be used to help answer the questions discovered through the other. It is a dialogical approach that combines the guiding questions and tools that are relevant to both systems considerations and to human-centered design. The tools on data, insights, opportunities, and solutions and their guiding questions in the design process as per Both (2018) are combined and presented in Figure 2.

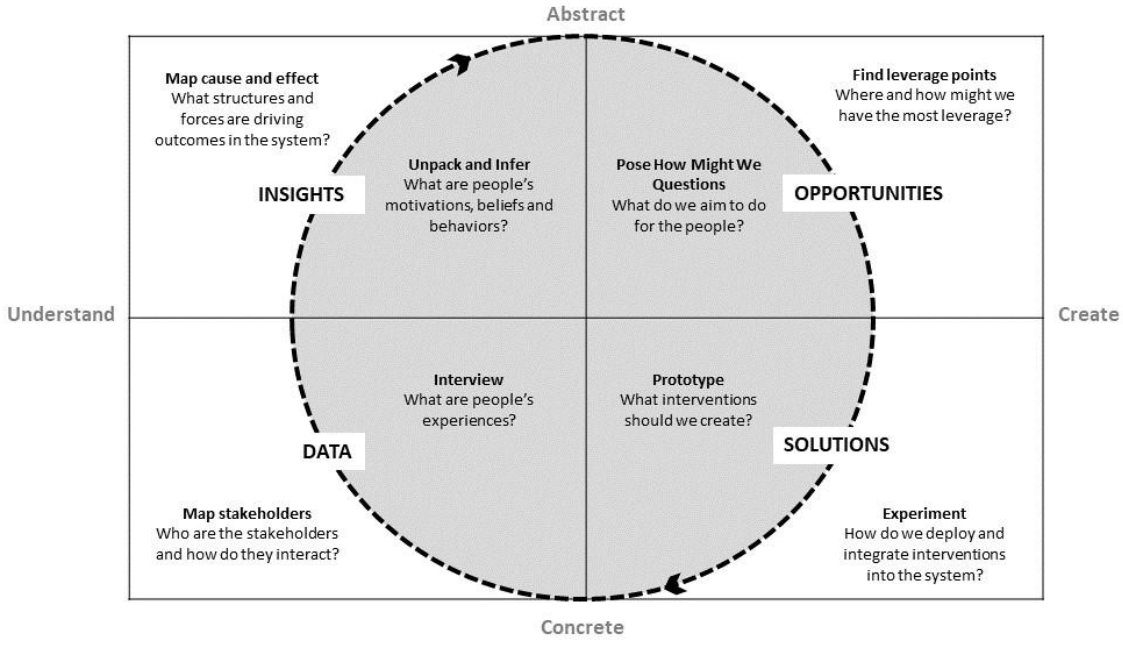


Figure 2: Human-centered systems-minded design framework (Both 2018)

Finding ways for humans and machines to combine their strengths of quantitative analysis and qualitative sense-making holds the potential for tackling challenges of extreme complexity. (Kelly 2016, 60.) In innovations where both the capabilities of people and of AI are at play, the relevant considerations are also about finding out ways to **design collective intelligence systems responsibly**. Wilson & Daugherty (2018) suggest that humans' role in systems of collective intelligence could be in training the machines, explaining the outcomes and sustaining the responsible use of AI.

According to Mulgan (2018) **organizing collective intelligence should happen through assemblies.** They combine key elements of observing things, making models and predictions, analyzing and interpreting, acting as stores of memory on what works, creating innovations and feeding into systems that make judgements about action as well as learn from those actions. Such assemblies are rare and need skillful "intelligence design" to organize and function. (Mulgan 2018, 27-30, 218-221.)

2.5 Summary of the theoretical framework

The academic discipline of peace and conflict studies sets the scene for the thesis theoretical framework. It is used to understand more holistically the context of peacebuilding – from the perspective of potential beneficiaries and of application users - than would be possible to do through data obtained in interviews or workshops. Existing research helps to evaluate the needs identified in this development work. It also lends insight on the complexity of peacebuilding as an environment and an ecosystem in which the future AI solutions should be integrated in.

Contemporary conflicts are understood as complex, dynamic and asymmetric situations that can develop and escalate rapidly and unpredictably due to the appearance of new actors or internal power struggles. They are **super wicked problems**, each unique, difficult to define, with no simple, clear nor tested solution, but rather have multiple explanations to problems that may also be symptoms of other problems. Having the same actors to both cause and solve the problem and with little time to find working solutions to replace the obsolete ones create additional complexity. (Peters 2017, 388.)

This is why **a holistic approach to peacebuilding is chosen.** It encompasses all conflict levels and dimensions from the responses of the international institutions to the local level actions and actors. Thus, peacebuilding is not limited to conflict prevention, management and resolution but is also closely connected to building a functioning society that includes everyone: state, private sector, NGOs, international institutions and communities.

Peacebuilding is used as an umbrella term to refer to the combination of different strategic interventions that happen at all stages and at all levels with the goal of reducing the potential for violent conflicts, ending them and tackling the root causes. The development work focuses on **finding areas where AI could be used to support the actions that are strategically important to peacebuilding objectives.**

The second part of the theoretical framework is to understand the capabilities and limitations of artificial intelligence. For this **a functional approach to AI is chosen** for the development work. AI is viewed as intelligent systems that can perceive and analyze their environment, interpret data and based on this decide on actions to achieve a goal without constant guidance from a human user. AI systems learn from past experiences and adjust to new information. **Machine learning** and other subfields of AI are all referred to as AI. The ability of ML systems to detect patterns in training data, analyze them into models and make predictions also for new, unseen data makes it interesting. (USAID 2018, 10-12). The central characteristics to AI are **autonomy and adaptivity**. (Russell & Norvig 2016, 2; Kaplan & Haenlein 2019, 18.)

The combination of rapid technological advancements and surge of real-life applications means that AI has profound impacts on societies and the way we live**.** While AI can enhance productivity and effectiveness of businesses, it can also further exaggerate the existing social challenges and power structures. Despite this, so far little attention had been paid to **how AI could be proactively used to tackle complex social challenges.**

Concerns of **ethics and responsibility of AI** are increasing rapidly. Issues of trustworthiness, transparency and explainability of AI systems, data ownership, accountability and governance are at the core of different principles and guidelines formulated for ethical AI. The challenges in this area are universal but have a particular relevance in AI building solutions meant for conflict environments.

The development work merges understanding on what AI is good at doing and what are its limitations to insights on conflicts as a context and the needs of peacebuilding. The analysis is used to identify future opportunity areas for AI peace technology. The evolving body of work on **peace technology** is used as an inspiration and confirmation of validity of the results achieved in a case study setting.

Peace technology is used to refer to strategic and context-specific innovations that utilize AI to prevent and end conflicts, to tackle reasons behind them and to alleviate human suffering arising from them. The characteristics that separate peace technology from others are its aim

to build or sustain peace in a given context and its strategic importance to achieving peacebuilding objectives. ((Dahl Jensen and Amnebjer 2019, 5; British Council 2016.) The distinction between non-strategic and strategic uses of technology links peace technology to **innovation**. It is not just a new word for something that is already being done. (Puig Larrauri 2018).

The thesis utilizes existing understanding on responsible innovation and design to inform the **development of tools and considerations for the responsible design of AI peace technology.** For peace technology traditional design approach and use of participatory methods might not be an option. Conflict sensitivity - making sure that the technology does not increase tensions or create new ones - is seen as a prerequisite and a starting point.

For peace technology responsible design is not an option, but a default. While sometimes the most responsible option is not designing at all, the real question for this thesis is how to decide what needs to be considered in order to move forward with the innovation. The principles of responsible innovation are applicable for peace technology. Responsibility needs to include both **responsible design process and responsible use of its outcomes**.

The thesis demonstrates that AI peace technology is based on **collective intelligence** or combining the capacity of various individuals and the capabilities of humans and machines (Mulgan 2018, 237). Thus, **the responsible design of AI peace technology is also about design of collective intelligence systems**. It is also about bringing together the expertise of both peace and conflict and technology communities. (Cottray & Puig Larrauri 2017; British Council 2018, 4.) The insight on collective intelligence systems informs the development of responsible design tools meant for situations where both capabilities of groups of people and of machines are needed.

The human-centered approach to design has its shortcomings when applied to multifaceted, social challenges (Both 2018). Together the strengths of quantitative analysis of AI and qualitative sense-making of humans hold the potential for resolving questions of extreme complexity. (Kelly 2016, 60.) In innovations like AI peace technology where both the capabilities of people and of AI are at play**, a combination of human-centered and systems-minded design approach** is used as guidance (Both 2018)**.** It is used to inform the development of responsible design tools that can help to understand relationships, to act in an ecosystem of various stakeholders and to preemptively recognize potential impacts in a particular context. The notion of “intelligence design” or how to organize collective intelligence systems through assemblies also informs the development work. (Mulgan 2018, 27-30, 218-221.)

3 Methods

This chapter explains what kind of process and methods were used and why they were chosen for this development work. It also describes what are the limitations of the study. Firstly, the purpose was to identify opportunity areas where AI peace technology could strategically support peacebuilding objectives. For this, a case study methodology including a literature review, workshops, expert interviews, design probes and a technology demo was utilized.

Secondly, based on the insights gathered during the research process, the thesis develops tools and considerations for responsible design of AI peace technology. For this objective, an iterative process using design mindset and methods such as co-creation and iteration was carried out in the development work.

The case study as a method and the reasons for choosing it are explained in the Chapters 3.1. and 3.2. Limitations for the thesis are considered in 3.3. Each phase of the case study development work process and the methods used during them are described in more in detail in the following chapter.

3.1 Case study as a method

Case study in its widest sense is often considered a method, but in fact it encompasses several research methods. So rather than a traditional development work method for a thesis, case study is an approach, or a strategy to carry out research that utilizes various sources and methods to collect insights. (Laine et al. 2007). It is particularly useful in tackling questions that are mainly “how” or “why” in nature, that focus on current, real-world context phenomena and on issues to which the researcher has little control over. (Yin 2014, 2.)

The research scope is another way to recognize a case study from other methodologies. It is an in-depth empirical inquiry investigating a contemporary phenomenon in its real context where the boundaries between the phenomenon and the context might not be clearly defined. (Yin 2014, 16; Kananen 2013, 28.) Understanding both the phenomenon as well as the conditions posed by the context are equally important parts of the research. For example, a laboratory experiment of AI applications meant for phenomena like peacebuilding would give very limited information on the real-world impacts, ethical concerns or design issues arising from them.

Case study aims to comprehensive and in-depth understanding of both the phenomenon and the multiple linkages to its context. Like other qualitative research strategies, such as action

research or development research, case study is based on dialogue between real-world practice and already established theories. A case study research, however, can be a combination of both qualitative as well as quantitative research methods in order to obtain data and gather insights. (Kananen 2013, 24; Laine et al 2007, 12.)

The case study process can be divided into four different main phases. According to Ojasalo et al. (2009, 54) the process starts with the researcher having a 1) **preliminary understanding on the research problem**, continues with 2) **focusing the research goal** though familiarizing oneself with literature and practical issues related to the phenomenon under study and 3) **collecting and analyzing data** through variety of different methods. The case study process culminates with 4) **formulation of results** into a model, use-cases or in this case, future opportunity areas for use-cases.

According to Yin (2014, 1) a case study consists of six different stages. The research is considered a linear, but iterative process. In comparison to Ojasalo et al. (2009) Yin (2014) places more emphasis on the design and preparation stages of the research and divides them into three separate phases: plan, design and prepare. Furthermore, Yin (2019) underlines the necessity of carefully first selecting and then transparently introducing the methods chosen for data collection and analysis of a case study. The case study process according to Yin (2014, 1) is presented in Figure 3.

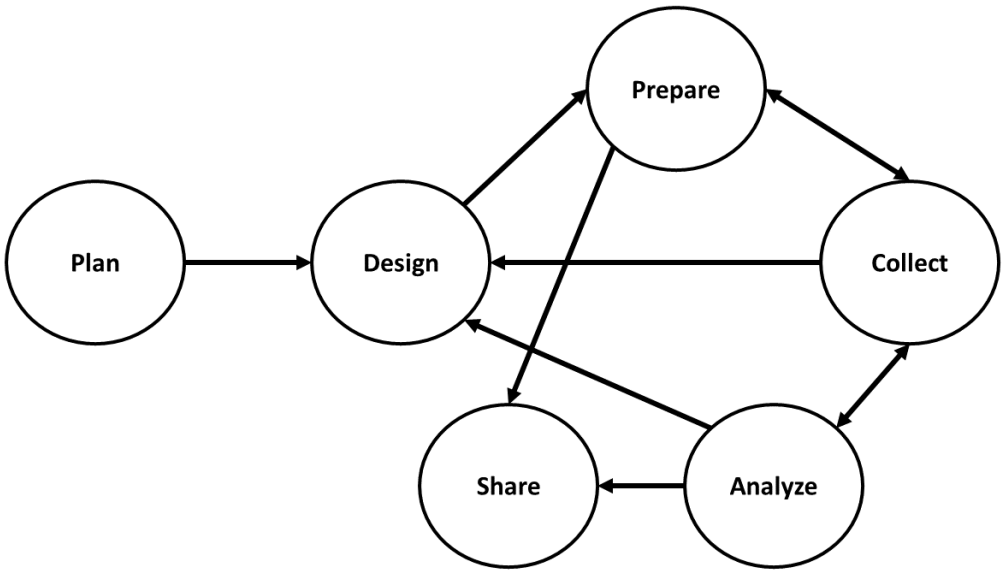


Figure 3: Case study process (Yin 2014, 1)

A case study distinguishingly tends to deal with multitude of variables: interests, data points and wide-ranging problem areas that all can have several linkages and unclear boundaries between the researched phenomenon and its context. Thus, the researcher needs to rely on

multiple sources of evidence and on converging findings through different types of triangulation. This means case study utilizes prior research and theory to guide data collection and analysis. (Yin 2014, 17; Kananen 2013, 33-36; Laine et al. 23-26.)

Case study typically uses multiple sources of information, multidisciplinary approach, and triangulation. This is to gather a holistic understanding and to look at the research topic from various perspectives. Applying multiple research methods also means that any potential gaps in insights gathered with one method can be complimented by information collected using another. (Kananen 2013, 34; Yin 2018, 126.)

Triangulation of insights improves the validity and reliability. It gives a more multifaceted and in-depth picture of the complex and contextual phenomenon under study. (Kananen 2013, 35-36; Yin 2018, 126-127.) Data and methodological triangulation are usually most practical ones to apply for a development work carried out in case study format. In addition, investigator triangulation and theory triangulation could be used. (Kananen 2013, 36; Yin 2018, 128.) Triangulation supports the overall goal of gaining an in-depth and holistic understanding of the phenomenon and its context that are being studied.

3.2 Case Peace Machine

All main criteria of a case study approach are applicable to this thesis. The development work gathers insights to formulate an understanding on future potential of AI peace technology to be used in the complex context of peacebuilding and conflict areas. **The focus is on a set of contemporary, real-world events and phenomena beyond the researcher's control.** The context includes rapid development of AI applications and opportunities created by it on one hand, and an increase of violent conflicts worldwide and the need to resolve them on the other.

Broadly the development work has a social goal of supporting peacebuilding. However, it merges expertise and insights from the fields of peace and conflict studies, peace technology and AI development as well as responsible design in order to answer the research questions set out for the thesis. The thesis does not represent a sample, but rather carries out a development work that seeks to broadly understand the phenomenon in question by using different sources of data. It uses triangulation to make analytic, rather than statistical generalizations about the future opportunities and limitations of AI peace technology, and of designing AI for social impact in general. (Yin 2014, 19-26; Kananen 2013, 24-26.)

The main goal for the thesis is to answer two broad "how" questions. Firstly, it investigates **how AI could strategically support peacebuilding objectives.** This leads to the identification

of future opportunity areas for AI peace technology. Secondly, it explores **how AI peace technology can be designed in a responsible manner.** This results in the formulation of peace technology design tools and issues to consider. Furthermore, these tools could be applicable for responsible design of AI with intended social impact in general, and for vulnerable or conflict contexts in particular.

This development work falls under qualitative case study (Ojasalo et al. 2009, 52). For the resulting opportunity areas and design tools, gaining a holistic understanding through various sources factoring in technological and design considerations as well as questions of context specificity is important. Multiple methods are used for collecting the data and formulating insights in order to understand both the phenomenon as well as the conditions posed by the context.

The thesis research is dialogical in nature. It identifies needs arising from real-world practice and context and draws conclusions from the insights gathered through different methods. Based on this, results are formulated, and their generalization suggested. The findings are reflected on the existing research and theory in research, analysis and share phases to create new information on the current understanding about the opportunities and design of AI peace technology. (Ghauri & Grønhaug 2005; Kananen 2013.)

The research design follows the logic of a case study where the researcher is an outside observer, yet interactively trying to understand the complex phenomena in its real-world setting from which the case under study is inseparable. The design evolved organically with the goal of finding suitable and attainable methods for different stages of research from data collection, analysis and sharing of results. (Kananen 2013, 11-25; Laine et al, 26).

The development work follows a linear, but iterative process adapted from Yin (2014, 1) and aligned with the process described by Ojasalo et al. (2009).

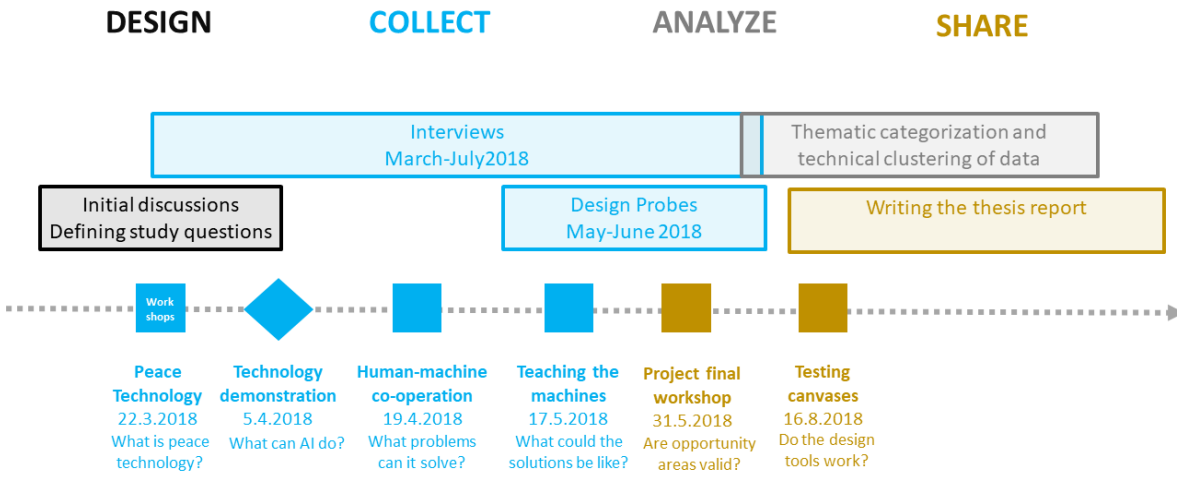


Figure 4: Thesis study process and schedule

The adapted case study model is used also to inform the report structure. The following chapters describe the case study process of data collection, analysis and sharing of the development work in more detail. The data collection was carried out between March and August 2018 while the majority of the analysis and conclusions were done between May 2018 and December 2019. The finalization of the thesis report and sharing of fully formed findings took place in spring 2020.

Case study was chosen as an approach for a service design thesis due to multiple reasons. In addition to what is described above, there was little to no opportunities to engage end-users or beneficiaries, do ethnography or other types of observation in real-life settings that is central to the insight gathering in service design. Secondly, only part of the results – the design tools developed - were prototyped and tested. Qualitative methods – including those of service design – were used. Methodological and data triangulation was applied to ensure holistic and multidimensional understanding of the case and the different aspects relevant for responding to the research questions. (Kananen 2013, 33-35; Yin 2018, 128-129; Laine et al. 23-25.)

3.3 Limitations of the thesis

The thesis focuses on the future opportunities of AI peace technology or identifying the ways AI could strategically support the peacebuilding objectives. While relevant to the topic of AI and conflicts as a context, this development work does not include specific data on the use of AI for military and security purposes. For practical reasons of access to information they are outside the scope of this thesis. Scalability and business models of AI peace technology are only discussed as part of the conclusions.

The ultimate beneficiaries of AI peace technology are people living in conflict areas and affected by conflicts. They did not participate in this development work. The reason is practical. It was not within the scope of this development work to directly engage with them: carry out interviews or workshops in a conflict setting. Instead existing research and desk study was used. In addition, potential end-users - the people working in the field of peacebuilding and living in conflict contexts – who also have understanding on the needs of beneficiaries did participate in the research via design probes, interviews and workshops.

Applying design thinking and service design methods in this study was simultaneously a natural fit and a challenge. Iskander (2018) argues that design thinking tends to be naturally biased towards preserving the status quo. When tackling questions of uncertainty and real-life wicked problems without direct access to the intended beneficiaries, one needs to be mindful about the role of design and that of the designer. According to Iskender (2018) there is a risk that they could become too dominant. In the case of AI peace technology, no individual designer can become an implicit vessel that alone understands and conveys everything that ends in the final service.

Critical consideration of pros and cons of methodological choices, awareness and transparency of personal positioning, political affiliations, biases and blind spots are never the more important than in design of AI peace technology. At its best responsible design should be open, inclusive and seek to democratize participation. (Iskander 2018) Things important for increasing innovation potential in any sector. In the case of peace technology design needs to be carefully thought out and challenged to be at its best as a default. This needs concrete tools and practices.

4 Process

This chapter describes how the case study process was carried out in this development work: what were the methods used in different phases, why they were chosen and what kind of data was gathered. The chapter also comprehensively explains how the analysis was done in dialogical manner, how the insights were formulated, and the findings tested and shared.

4.1 Overview of the process

The different methods chosen for each phase of the case study process, the purpose of each method and the outcomes obtained through it are summarized in table below.

Process phase	Method	Objective	Outcome
Design phase	Three exploratory discussions	To understand the relevance of AI in the field of peacebuilding Map what is already done	Preliminary goals for the development work Ideas for interview questions Work plan for the collect phase
Collect phase	**Workshops**		
	What is peace technology?	To understand what is considered peace technology	Mapping of peace technology ideas and solutions Mapping people's perceptions on the topic
	Human–Machine co-operation	To find AI solvable problems based on real-life problem scenarios	Mapping of stakeholders and their needs Two draft solution proposals developed
	Teaching the machines	To refine solution ideas with technical and data considerations	Solution proposals re-iterated using AI perspective
	Thematic interviews		
	8 peace and conflict experts	To understand the needs and context of peacebuilding	Qualitative insights on realities and needs in peacebuilding
	7 AI and technology development experts	To understand abilities and risks of AI and application development	Qualitative insights on capabilities and limitations of AI
	Design probes	To understand everyday realities of peacebuilding work	Direct observations on the needs and challenges encountered in peacebuilding operations
	Technology demonstration	To demonstrate machine learning's capability in speech recognition and emotional analysis	Observations on capabilities and limitations of ML as a technology and on perceptions of its use
Analyze phase	Thematic categorization of interview data	To identify and analyze key needs in peacebuilding	Thematic categories of needs and relevant issues in peacebuilding
	Technical clustering of interview data	To identify and analyze key abilities and limitations of AI in conflict situations	Clusters of technical considerations for AI peace technology development
Share phase	Project Peace Machine final workshop	To introduce emerging findings and opportunity areas for AI peace technology	Feedback on relevance of the findings and the opportunity areas identified so far
	Testing peace technology canvases workshop	To test the design tools developed	Feedback on how to re-iterate the canvas set

Process phase	Method	Objective	Outcome
	Writing the thesis report	To complete the analysis on opportunity areas and finalize the set of design tools	Opportunity areas of AI peace technology defined Tools and considerations for responsible design of AI peace technology finalized

Table 1: Phases, methods, objectives and outcomes of the development work

In order to gather cross-disciplinary insights for **five workshops were carried out with a total of 82 participants**. They were the most important part of the iterative development work process, benefitting from participation of a wide range of stakeholders: peace researchers, service designers, AI and technology developers, development policy experts, communications professionals, conflict and crises management practitioners, artists, students and start-up entrepreneurs. In addition, a total of **30 people took part in the technology demonstration** organized as part of an event. The demonstration used IBM Watson application as a test case for technological capability of natural language processing and emotional analysis.

A total of **13 thematic interviews with 15 peacebuilding and AI experts** were carried out. Representatives from the public sector, businesses, academia and civil society were interviewed anonymously, as agreed in the beginning of each interview. This was to secure a good representation of different perspectives from both AI as well as peacebuilding sectors to the study data. Additionally, for better understanding of the realities and challenges encountered in conflict settings **three design probes were carried out.** Representatives from governmental, intergovernmental and civil society organizations took part in the probes that covered different types of peacebuilding operations from Africa, Middle East and North American (HQ) contexts.

All the material gathered as part of the development work was recorded in pictures or audio that was then transcribed (not word-for-word transliteration). The workshops and technology demonstration were open and by invitation events. Their main findings were made public in the project website (Understanding Peace 2018). Excluding the public website and its related archives, all documentation is in the possession of the thesis writer for the sole use of this development work. It will be destroyed upon approval of the thesis.

The following chapters explain each case study phase and methods used in more detail. The results will be presented in chapter five.

4.2 Design

According to Yin (2014) in the preparatory phase it is essential to define the initial research goals and questions of inquiry, get to know the phenomenon under study and case study as a research strategy. Defining research questions is perhaps the most important phase in a case study. Both substance (what is the study about?) as well as form (what kind of questions is the research asking?) can point towards a suitable research method for data collection and its analysis. (Yin 2014, 11.)

Formulating specific enough goals and research questions turned out to be a longer process for this development work. It was important to understand what the needs and challenges in peacebuilding were and what kind of problems AI was good at solving. The design phase was an iterative dialogue of getting to know the realities of peacebuilding and capabilities of AI simultaneously. Acquiring insight on real-life problems and needs of people living in, working with and affected by conflicts was central for defining relevant research questions for what the study was about. Obtaining an understanding of AI abilities was important for specifying what kind of questions the development work would be asking.

The design phase constructs a logical research plan for the case to get from "here" to "there". The "here" is the set of research questions and "there" conclusions responding to those questions. (Yin 2014, 26-27.) Yin (2014, 55) also argues that in case of changes in research questions, a new research plan to restart the whole process is required. From the bundle of initial questions around AI, peace technology and peacebuilding an overall approach of gathering insights on what AI applied to peacebuilding could potentially look like was chosen as a starting point to embark on data collection workshops and interviews.

A continuous dialogue including refocusing and re-defining more specific research questions continued during the design and collect phases of this study. This follows the logic of Eriksson & Koistinen (2014) according to which defining a preliminary research question is the most important resource and enough to start the case study research process. A roughly defined initial question was redefined, particularized and re-focused throughout research in a dialogical relationship with the empirical data and understanding gathered. (Eriksson & Koistinen 2014, 23.)

Exploratory discussions

The separate preparation phases as defined by Yin (2014) were merged into one design phase that started with two informal **discovery discussions with peacebuilding experts** to scope the relevancy of AI for the sector. In addition, a discovery discussion with Timo Honkela and

two representatives from the Peace Machine network (now organized into Rauhankone ry) also took place. Based on these discussions, preliminary goals for the development work, ideas for interview questions and a work plan for the collect phase was developed.

Work plan for the research

In the design phase of this development work a research plan including themes for a **series of four workshops** was defined. Potential speakers and participants for the workshops were identified and their co-creation methods planned. **The design phase took place between March and May 2018**. Partially overlapped with the collect phase. During this time thesis goal was re-defined and more specific research questions discussed with the thesis supervisor from Laurea University of Applied sciences and work plan adjusted accordingly with the Peace Machine project leads from Futurice.

Together with the workshops, thematic interviews were agreed as primary data collection methods. The insights from interviews and results from first workshops were used to inform the research process moving forward**.** An ongoing desk study on topics around AI and peacebuilding was started to converge the case study insights with multiple sources of data. Using different types of evidence as research material and triangulating the data allows a continuous dialogue between the emerging findings. It can shape the collect phase and potentially unearth new, even contradictory viewpoints to the same phenomenon. (Yin 2014, 128; Eriksson & Koistinen 2014, 31.)

A cyclical way of analyzing and summarizing the emerging findings was also pre-defined in the Peace Machine project. Summarizing lessons learned, inviting comments and participation to the workshop was done via podcasts. The policy was to make the project open and inclusive. For this development work the continuous sharing and learning allowed any gaps in data to be addressed in the next workshop or interview. The thesis thus follows an iterative nature of explanation building by making initial explanatory propositions, comparing the data and revising the earlier propositions if necessary (Yin 2014, 180).

4.3 Collect

The collect phase took place between April and August 2018. It consisted of **three workshops, 13 thematic interviews, a technology demo and three design probes.** Contextualizing data collection methods, whether they are using existing studies, doing interviews or carrying out observation is essential (Moilanen et al. 2010, 55). In development work converging insights from more than one field of practice, this is ever the more relevant.

Being able to ensure participants' anonymity while transparently and actively communicating about the emerging findings of the project was a major part of the data collection. Particularly for those working in conflict zones, confidentiality was a key concern. Thus, all references and quotes were agreed to be distinguished only through a minimum set of identity attributes such as area of expertise or general level geographic location.

The first three workshops, the methods used in them and their key results are described in more detail in the next chapter. This is followed by description of the thematic interviews, technology demonstration and the design probes that comprise the collect phase. The last two workshops are included to the chapter 4.4. describing the share phase of this development work.

4.3.1 Workshops

The Peace Machine project had pre-defined head-line level topics for the first four workshops. The logic was to start from a general exploration of peace technology as a topic and move towards more detailed issues of AI, data, ethics and design. **The workshops were conducted in English** to facilitate more inclusive and diverse participation. In all workshops, participants included designers, data developers, AI and peace experts from private sector, academia and NGOs.

All workshops followed the same structure. They were two-hour long with an expert introductory note in the beginning, a group work phase and a joint report back discussion on key results and reflections at the end. The group work utilized an adapted Learning Café method to support free flow of ideas and dynamic discussions between experts from different fields within a short period of time. Workshops were facilitated together with Futurice project leads and the thesis writer. Canvases were tailored and re-iterated for the groups to use and for gathering the insight from each workshop. In addition to a tool for data collection, the canvases themselves were developed after each workshop and shared for comments at the project website.

Workshops helped to engage a wide range of people with different types of expertise to discuss, ideate and co-create future solutions. By including different stakeholders in the development work, it was possible to collect new ideas and develop solution proposals. This would have been difficult to do via any other data collection method. Furthermore, the workshops as a method kept the data collection constantly rooted in the reality of both fields of AI and peacebuilding. (Stickdorn et al. 2018, 391.)

The central idea was to make sure that both technical and substantive expertise were simultaneously represented in the workshops. This provided an opportunity for cross-fertilization and quick re-iteration starting from problem definition. At the same time, **the workshops were a way to gather insights on the real-life dynamics of a design process** that would need to involve a similar range of experts.

The workshop results were documented in canvasses and notes, summarized and shared via three podcasts between March and May 2018. All information produced in the four workshops is open source and available via project website or later archived in GitHub (Understanding Peace 2018). Short teaser videos to the workshops were produced for the Futurice project. (Futurice 2018). Active communication helped to extend the coverage of workshops by attracting more participants and by contributing to the cumulative knowledge of AI in peace technology.

What is peace technology? workshop

The first workshop **"What is peace technology?" was organized March 22nd, 2018 with a total of 17 participants.** It started from the definition of peace technology as technology reducing or preventing violence. Group work used Future Wheel canvas to map out perceptions of what was considered to be peace technology and to identify benefits and risks of using AI to support peacebuilding.

Figure 5: Group picture from What is peace technology? workshop

A Learning Café technique including a specified reporter allowed the discussions after group rotations to build on those of the previous group. The coverage of what different participants understood to be peace technology was wide. They included a translator to facilitate communication across cultures, an assistant for non-violent interpersonal communication, supporter for emotional skills, assistant for recognizing cultural, religious and emotional differences, emotional whistles, VR experiences, visualizations, detection of early-warning signs for conflict escalation and so forth.

The report back of the group work results and the following joint reflection discussion were documented in notes. Based on the workshop documentation (group work canvases and facilitators' notes) the workshop results were clustered under five categories:

1) Prevention through inclusion and satisfying of basic needs (e.g. amplifying voices of vulnerable groups)
2) Early warning and forecasting (e.g. making predictions using geographical data)
3) Values and ideologies (e.g. cultural difference recognition)
4) Interaction (e.g. addressing hate speech online)
5) Data analysis (e.g. creating models from data)

In addition, issues to consider for design of peace technology were discussed at the end. Participants noted that some AI peace technology would be easy to scale up or replicate (such as data visualizations). Others would only be functional at a more personal level (such as people meeting across conflict lines as "soccer lovers" rather than enemies). Local ownership of the potential applications was seen crucial for success. All levels from personal to local and from national to global should be considered as opportunities for negative of positive impact. Considerations of risk and potential for misuse were raised by participants, along with the importance of considering data reliability and existing power structures in each context.

Human-machine co-operation workshop

Human-machine co-operation workshop was organized on April 19th, 2018 and had 21 participants. A keynote on AI ethics and how humans and machines could cooperate building on their complementary strengths was used as an inspiration. To move the development work from ideation towards a more concrete and context specific co-creation, **two real-life problem scenarios were introduced as a starting point for group work.** A scenario on South Sudan (presented by Finn Church Aid) and a scenario on corporate subsidy political debate polarization in Finland (presented by Ellun Kanat).

The main points from both problem scenarios were condensed into case cards to be used in the group work. Three groups worked on one scenario each, starting from definition of concrete question to tackle. Groups used two canvases, a standard stakeholder mapping canvas to identify the parties affected or able to resolve the problem in question, and a summary Human + Machine co-operation canvas, first iteration of peace technology summary canvas. It is presented in figure 6 and included in appendix 1.

Figure 6: Human + Machine co-operation canvas used in the workshop

The solution ideas produced in groups were: 1) collaborative, easy to read news service using Wikipedia approach to combat fragmentation of data that can cause distrust 2) AI support to combining different types of data from various sources into a more truthful overall picture, and 3) utilizing AI to predict safe areas for movement of civilians within conflict zones.

The collective reflection discussion noted that in the solution ideas developed AI was utilized to gather information, check facts, verify events and create scenarios from vast sets of data. However, participants highlighted that human's role was to see how different pieces of information and analysis were connected and if the analysis provided by AI created a meaningful entity.

Teaching the Machines workshop

Teaching the machines workshop was organized on May 17th, 2018 and had 12 participants. The solution ideas developed in the previous workshop were used as a starting point. The focus was on refining them by concentrating particularly on technology concerns around the use of AI in conflicts. The main goals of the workshop were to test the insights gathered from the previous workshops and interviews so far, to develop solution ideas further to include issues of algorithms and data, and to see how this would affect the design of AI peace technology.

Concrete examples of AI for social impact were used as an inspiration. A keynote highlighted that while AI is good at finding patterns, making predictions and mimicking human substance expertise, it is up to people to interpret the information and make decisions on actions. To further focus the group work part, the specific context of South Sudan was used. In addition to the short case card setting up the context, three possible solution prompts based on the results of the previous workshop were provided for the groups. They were:

1.) How to identify spaces and locations for safe movement of people?
2.) How to have an up-to-date conflict analysis?
3.) Is it possible to build early-warning systems for predicting and preventing conflicts?

A re-iterated summary canvas was used to guide the discussion among groups with a wide variety of expertise and to document the results. Guiding questions on data use and ethics were included to the second iteration of the canvas presented in figure 7 and included in appendix 2.

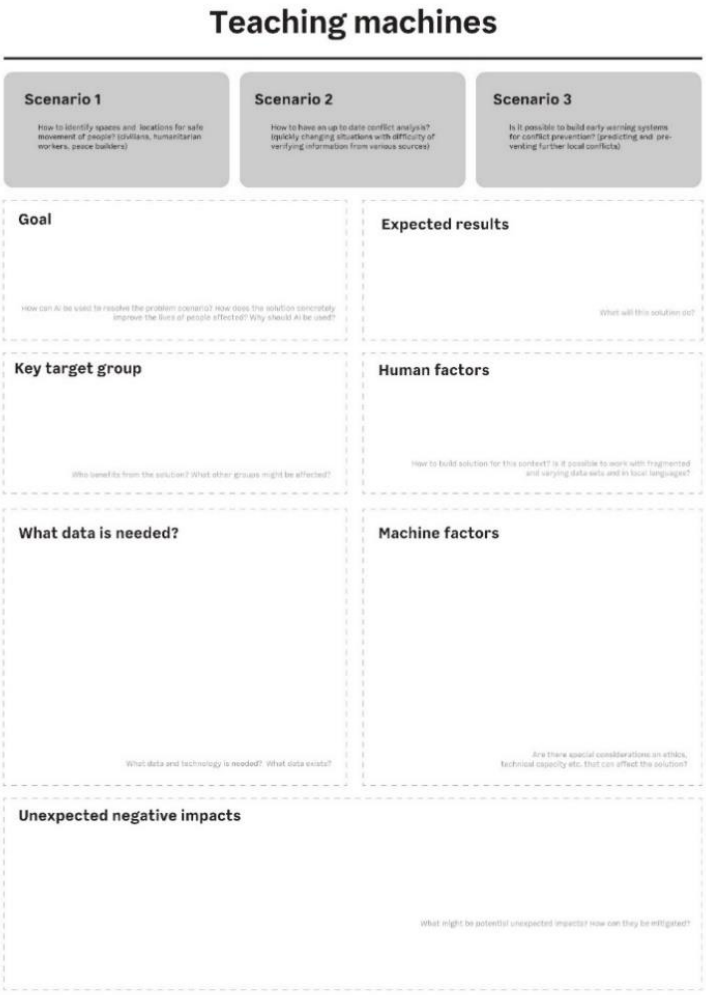


Figure 7: Teaching machines summary canvas used in the workshop

Solution proposals produced by the groups were 1) a **panic-button type of early warning system** based on machine analysis of large data sets and 2) visual system drawing together data from various sources in order to have a **more holistic and up-to-date conflict analysis.** The latter included an idea of a call-in reporting for peace mediators to collect data from community level and include it into the analysis.

4.3.2 Thematic interviews

To gain understanding on complex phenomena of AI and peacebuilding, **thematic interviews** covering a total of 15 experts from technical and substance fields were used to collect data. This was a way to gain qualitative information and insights on the realities and needs in peacebuilding as well as on the capabilities and limitations of AI.

In this development work interviews are used for collecting in-depth insights. In case studies they are seen as a good way to gather individual views and experiences on the case in question to enrich the study material (Yin 2014, 114; Moilanen et al. 2010, 95). Thematic interviews complement other data collection methods of the thesis. They are used to feed insights into and test findings emerging from the workshops.

A semi-structured thematic interview format was utilized. Its strengths include gaining insight on the meaning and real-life context of a phenomenon under study (Moilanen et al 2010, 95; Kananen 2013, 80). This method allowed flexibility in wording and order of interview questions that is important when the development work covers two very different fields of expertise. Simultaneously, an informed caution for the thesis writer not to lead on too much or use pre-defined questions selectively to re-enforce one's own views and biases was used (Kananen 2013, 80-81).

A total of 13 interviews with 15 participants were carried out between April and July 2018. All the interviews were recorded, and material agreed to be treated anonymously solely for the purpose of this development work. From the interviewed people 11 were men and 4 women, including representatives of public sector, businesses, think tanks, academia and civil society organizations. In this report the interviewees are referred to in general terms to protect confidentiality. A detailed list of interviews is presented in appendix 3.

Excluding the first two interviews that were non-structured, all interviews were carried out using a set of predefined questions. They were covered flexibly, according to the situation and expertise of the interviewee. **An average of 10 questions were covered in all interviews the length of which varied from 40 minutes to 1,5 hours.** Sometimes irrelevant questions were left out, other times specific questions were included to gather more in-depth knowledge. All interviews were discussion-like in nature. They were recorded and main points as well as specific quotes were transcribed. All but one of the interviews were carried out in Finnish. Translations used in the report are done by the thesis writer. The interview question framework is presented in appendix 4.

The interviews aimed to cover a wide range of perspectives from both fields of AI and peacebuilding. The peace and conflict experts – a total of 8 - provided insights on the needs and realities faced in the field of peacebuilding. The AI and technology experts – a total of 7 – gave insights on capabilities and limitations of AI and raised issues related to application development. This was vital for identifying where the needs of peacebuilding and capabilities of AI overlapped. The insights informed the design of the workshops creating an iterative collection phase with continuous feed-forward loops.

A coverage of 13 interviews for two such vast topics is limited. However, particularly when combined with the insights arising from the workshops, **the qualitative interview data was clearly saturated around a set of topics**. Instead of adding more interviews, the number of data points determined by the research problem and the case itself was used to determine when to stop. For qualitative research such as this, the interplay between data collection and analysis is a way to recognize when the material becomes saturated. After this, new interviews no longer bring additional perspectives to the phenomenon, but rather the responses start to repeat themselves. (Kananen 2013, 95.)

4.3.3 Technology demonstration

Futurice Peace Machine project organized a technology demonstration in *Mitä tapahtuu huomenna kohtaamiselle?* ('What happens to encounters tomorrow?') event. The event was a collaboration with Ellun Kanat, Sitra and Crisis Management Initiative. It took place on **April 5th 2018** at Vanha Ylioppilastalo. The demonstration lasted 30 minutes and included a short presentation, a live experiment to test machine learning capability in speech detection and sentiment analysis and a short reflection discussion.

The demonstration was an experiment with a random sampling of 30 people using IBM Watson based set up. Three groups of ten recorded a two-minute discussion on the topic "Is world peace possible?". Peoples' speech was detected, transcribed into text and analyzed for emotions on the spot. The text was reflected onto a big screen and intensity of sentiment analysis was added using colors. Strong emotions were shown in bright tones, more subtle ones were highlighted in dark. Full transcripts of the discussions with more detailed emotional analysis were published on the project website two days later. The detailed analysis of five different emotions - anger, disgust, fear, joy and sadness – is shown in figure 8. The technology demonstration was built as a collaboration by Digitalents Helsinki and Futurice.

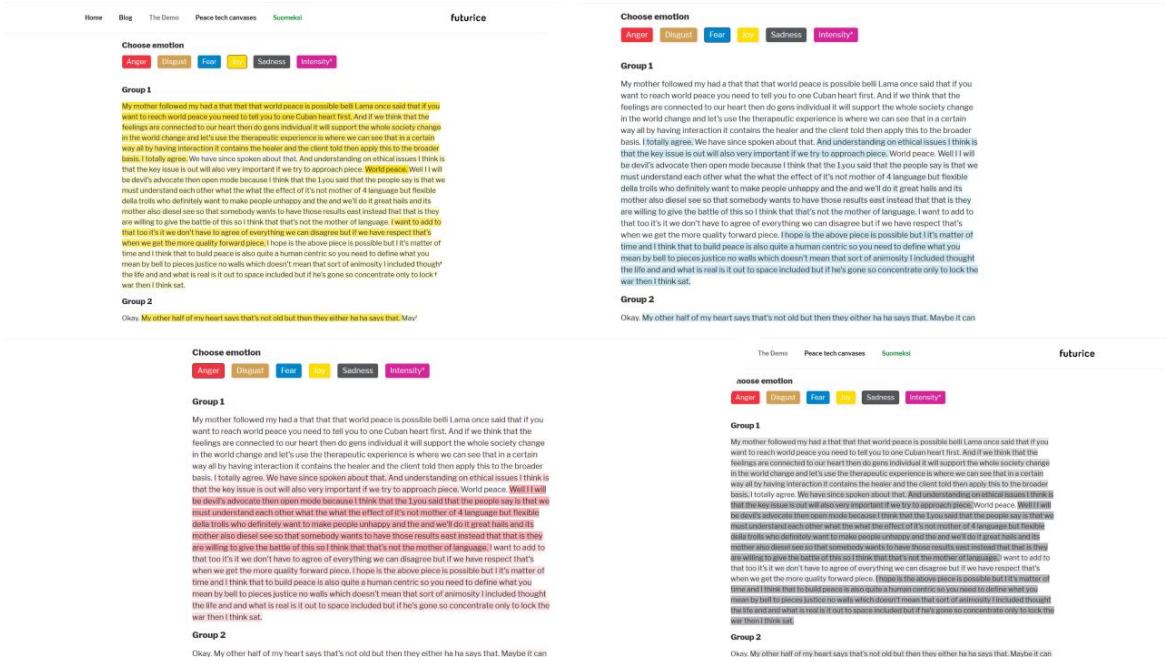


Figure 8: Sentiment analysis of group discussion transcripts

A two-minute reflective discussion on how this kind technology could be used to support peace objectives concluded the demonstration. The discussion of the same three groups was recorded for the use of this thesis. As an observer and assistant to the demonstration, the writer took notes on **peoples' reactions in a situation illustrating the capability of AI on speech and sentiment analysis.** While intrigued, the participants struggled to find concrete use for this type of technology in peacebuilding. The sentiment analysis showed little variation between different emotions and thus the test group felt it provided little additional insight to a normal discussion.

The demonstration underlined the challenges of applying the AI for peacebuilding and the risks of a technology driven approach to developing such applications. Without understanding the specific needs and individual contexts, it is difficult to design peace technology solutions. Design thinking and tools could help to bridge this gap.

4.3.4 Design probes

Design probes are a method to document potential users' everyday realities through different types of self-reflection and documentation. The probes focus on personal level experiences and auto-reflection of the surrounding ecosystem. They are used to map out the human-side of phenomena and to make sure the design process reflects actual users' pain points. In a thesis concentrating on finding potential for future solutions to complex problems, the experi-

mental nature of design probes proved to be a good match. Rather than solving existing problems, probes can be used to scope out new opportunities. By nature, they leave room for interpretation, experiments and creativity and enable emergence of new and surprising perspectives. (Mattelmäki 2006, 46.)

The data collection for this thesis was complemented by **three design probes.** They were inherently suitable for accessing the everyday realities and needs of peacebuilding on the ground. Probe participants were recruited through personal contacts and through an open call in a Facebook group. They included peace and conflict experts working at a headquarter setting (HQ) and two specific conflict countries located in Africa and in the Middle East. The participants came from civil society, governmental and intergovernmental organizations, thus **reflecting different types of ecosystems present in real-life peacebuilding operations**.

WhatsApp was used to carry out probes for a period of five working days. Making participation as easy as possible was important for the probes to work in challenging conflict settings. Anonymity and confidentiality of the probes allowed participants to document their surroundings and submit personal reflections that would be difficult to access through other methods. The material shared via probes included photos, short commentaries, on-the-spot observations and personal reflections. Majority reflects real-life situations as they happen. Personal thoughts on what works well in peacebuilding and where AI could help were encouraged. Instructions to the design probe participants are included in Appendix 5.

The probes were carried out between **May and August 2018**. After the observation period all participants were interviewed to discuss the reflections shared via the probes and to gain more in-depth understanding of the particular needs and dynamics involved in each individual setting. The goal was to understand the everyday realities of peacebuilding work and collect observations on the needs and challenges encountered in peacebuilding operations. Even with the small overall number of the probes, the goal was achieved. The qualitative data collected was unique and complemented the insights gathered through other means of data collection.

4.4 Analyze

A process of iterative qualitative data analysis was used in the thesis in order to identify future opportunity areas for AI peace technology and to develop tools for their responsible design. According to Eriksson & Koistinen (2014, 33) choosing where and how to start analysis based on the multiple types of evidence gathered is perhaps the most challenging research phase. Analysis and collection of data should be done in cyclical manner. If not analyzed simultaneously, the results could be superficial, and it would be difficult to know when material becomes saturated. (Kananen 2014, 106.)

The goals of analysis phase are threefold: 1) organizing the variety of data gathered about the case (e.g. classification or typologies) 2) doing content analysis of the data 3) interpreting the insights and results (Eriksson & Koistinen 2014, 33-34). This phase includes assigning meaning and explanations to the observations, linking different areas of understanding and formulating conclusions. Instead of mechanically assigning causalities and linkages to the evidence gathered the objective should be to simultaneously preserve the context and richness of data in the case study. (Shank 2002, 77.)

The analysis of this thesis is based on the data gathered and an **iterative approach to explanation building is adopted**. The objective is to understand the phenomenon in question by using different sources of data. The analysis merges expertise and insights from the fields of peace and conflict studies and of peace technology and AI development. Triangulation is used for analytic generalizations about the future opportunities and limitations of AI peace technology as well as its responsible design. Already in the collect phase, initial findings from the workshops and from the thematic interviews were gathered to inform the next stage of the development work and to be shared via podcasts.

Thematic categorization and technical clustering of interview data

The interview data collected was first transcribed into an excel sheet organized according to interview questions in horizontal axis and interview participants in vertical. Rather than word-to-word transcription, main points and most insightful direct quotes illustrating them were included into the table. This helped to start organizing the material collected, identifying similarities, gaps and open questions within the data. The excel sheet made the classification, comparison and clustering of data easier while simultaneously the information gathering continued. At the same time, the main themes and issues arising from the interviews were standardized by writing them into post-it notes. These were then classified into groups and the groups were assigned a title. These groupings are shown in figures 9 and 10.

The needs and realities of peacebuilding – the content issues – were used as a starting point for this first round of categorization. This thematic way of organizing the interview data helped to understand the needs and context issues of peacebuilding. The objective was to start creating organized categories from vast amount of data based on the research goals. (Strauss & Corbin 2008). The thematic categories were assigned a title and clustered freely without specific arrangement. The first round of thematic categorization was a mixture of themes and observations in order to see what insights are emerging from the data. The organization of the groups into an affinity diagram happened as the work progressed and more ma-

terial was gathered. The dark blue highlighting is added last, when writing the report. The organized affinity diagram illustrating repetitive themes of the interview data groupings and their potential linkages is presented below.

Figure 9: Affinity diagram of the thematic categories arising from the interview data

The thematic categories of the interview data were combined with key lessons learned from the first three workshops. The main findings arising at this stage of the analysis were grouped together to form preliminary suggestions of ‘emerging insights’. They concentrated around observations on the needs and context of peacebuilding, remarks on the technical abilities and limitations of AI and expectations and concerns raised around responsible design. These emerging insights were presented as drafts for discussion in the project Peace Machine final workshop. This is described in more detail in the chapter 4.5.

The author created the preliminary suggestions of emerging insights as drafts for discussion and clustered the different design considerations using the groupings shown in figure 9. The peace technology design tools were developed together, in a dialogue with the Futurice project team and finalized in the thesis report writing. The preliminary suggestions on potential

future opportunities of AI peace technology and the prototype set of three design canvases were tested in two workshops: the project's final workshop and in the testing the canvases workshop. These functioned as a way to validate and refine the emerging insights and continue the iterative analysis of the development work. As their purpose was to simultaneously share the initial results and to continue the analysis, they fall to both phases of the case study and are described in chapter 4.5.

The second round of iteration of the interview data analysis was carried in reverse order. It also included new data from three more interviews carried out in June–July 2018. Using the lenses of technical development – including the abilities and limitations of AI – the interview data quotes and observations were again organized into groups. This exercise of technical clustering used the same technique as thematic categorization: grouping post-it notes, assigning a title to each group, clustering them and organizing the clusters into an affinity diagram shown below.

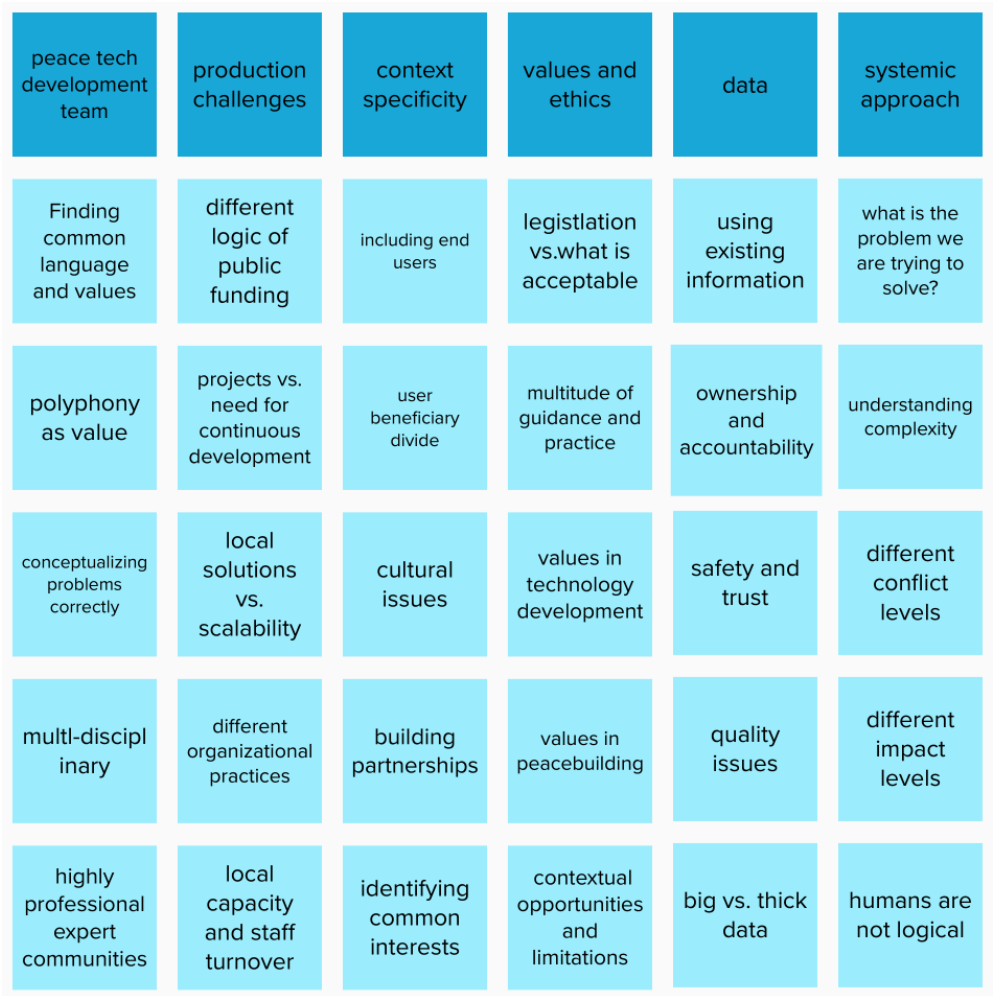


Figure 10: Affinity diagram of the technical clustering arising from the interview data

The purpose of this iterative way of analysis was firstly to organize the research data as it was being gathered and secondly to carry out content analysis from both technical and substance perspectives. Lastly, the different rounds of analysis helped to present the lessons learned at the final workshop of the Peace Machine project even when the thesis study was still ongoing. The project was essential to gathering the data for this development work. The results of this thesis are part of fulfilling the objectives set out for the project. However, the timelines and the objectives of the project and this study are complementary rather than identical.

As qualitative analysis is always an interplay between the data and the researcher, even an objective analysis includes the interpretation of data. This can always carry biases of the researcher's subjective position and reflect the choices made in data collection. The dialogical rounds of analysis subjected the classified groupings and the interpretations based on them to further scrutiny and discussion. Rounds of refinement happened through the remaining workshops and interviews as well as through publishing the lessons learned openly in the project website as they emerged.

The dialogical explanation building, and analysis continued during the report writing. The technical clustering and thematic categorization were also used to inform the development of the design tools. The workshops using various iterations of the canvases helped to hone the opportunity areas. The results are presented in their final form in chapter 5.

4.5 Sharing emerging insights and testing design tools

For case study development work sharing of conclusions means constructing the results and conclusions and bringing them into a closure. Drafting preliminary insights already at early stages rather than waiting to the end of data analysis is recommended. (Yin 2014, 176.) Following this logic, the development work continuously drafted tentative summaries and insights into textual and visual materials that were shared publicly in podcasts and on the project website. In addition, two workshops were dedicated to sharing draft opportunity areas where capabilities of AI and needs of peacebuilding meet and the emerging insights around responsible design and to testing the design tools developed.

Sharing emerging insights at project's final workshop

The final workshop of the Futurice Peace Machine project was organized on May 31st, 2018 and had 14 participants. All the participants from the previous three workshops were invited and the emerging insights of this development work as well as lessons learned from

the project were presented. The author held a presentation summarizing insights on peace-building and conflicts as a context, outlining six preliminary suggestions of draft opportunity areas for AI peace technology and presenting a protype set of three canvases for responsible design. The presentation functioned as report on learnings of the project. It utilized the first round of data analysis as described in the previous chapter.

The presentation was aligned with the objectives of the Peace Machine project. The emerging insights were presented in the form of draft opportunity areas, but they could also be described as ideas for use-cases. For the presentation they were grouped under wider headlines of “data analysis” and “language and sentiment”. The draft opportunity areas are condensed in figure 11 illustrating how the emerging insights were organized and presented at the workshop.

Figure 11: Draft opportunity areas presented in the project’s final workshop

Group work utilized the three peace technology canvases developed for this workshop. This was a way to collect feedback on the relevance of the six draft opportunity areas as well as to test the canvases. The set of canvases included 1) summary **peace technology canvas** 2) **stakeholder mapping canvas** with the dimensions of beneficiary – community – society included on one side and user – organization – ecosystem on the other, and 3) **impact canvas**

including dimensions of direct and indirect, positive and negative impacts. The canvases are included in the appendix 6.

The groups concentrated on discussing future opportunities of AI in peace technology and the role of design in it. They found the draft opportunity areas relevant and design tools mainly as a good way to make sure one is considering at least the minimum set of issues in the design of AI peace technology. The peace technology canvas was seen to function as kind of a check-list for responsible design, complemented by the impact canvas. The stakeholder mapping canvas directed the participants to think along the lines of stakeholders in a given conflict setting rather than for the potential peace technology application under development.

Testing the peace technology canvases workshop

The set of three peace technology canvases were refined based on the feedback received in the previous workshop. Draft versions of the canvases were published as a work in progress on the project website. The creative commons license made them open for comments and use. In addition, a separate workshop for testing the canvases was organized **on August 16th, 2018. It had 18 participants** from service designers, startup entrepreneurs, civil society representatives to marketing, technology and peace experts.

The workshop concentrated on collecting feedback from the participants on using the canvases for designing AI peace technology or for AI applications with an intended social impact in general. The group work used the canvases on cases picked by the groups and asked the participants to consider the following three questions: 1) are the canvases understandable and asking the right questions? 2) What is missing, unclear or needs to be changed in the canvases? and 3) Do you feel that these tools bring value-added or new ideas to the design process?

The summary peace technology canvas was re-grouped and re-worded based on the previous feedback to make the canvas easier to navigate. Under the “Problem” definition heading it included questions about the user and other stakeholders, impacts and risks. Under the “Goal” heading it included questions about machine and human factors, intended format and data needed for the solution. Issues of scalability and financing (rather than “business model” used in previous iteration) formed the final parts of the canvas.

The stakeholder mapping canvas was reformatted to be a more organic version of the previous. The impact canvas now asked to consider positive and negative, direct and indirect impacts for the potential solution rather than from the potential beneficiary perspective. The set of three canvases used in the workshop is presented in the appendix 7.

All feedback from the two workshops - sharing the emerging findings and opportunity areas and testing the draft canvases – was then incorporated into the analysis of this development work. It informed the case study report writing confirming the validity of the most important findings and suggesting issues that needed further consideration particularly on part of making the design tools easier to use for anyone. The set of canvases and the draft opportunity areas were re-iterated based on the feedback received. The results are presented in the following chapter.

5 Results

This chapter presents the results of the development work and responds to research questions. The objective of the thesis was to identify future opportunity areas where AI could strategically support peacebuilding objectives. Based on the insight gathered the second objective was to develop tools for responsible design of AI peace technology.

The overall purpose of the development work was to understand the opportunities and limitations that AI could have in supporting peace. Firstly, it was possible to identify three separate, but interlined opportunity areas for future AI peace technology applications. In these areas the needs of peacebuilding and the capabilities of AI were discovered to overlap. Secondly, the set of design tools were seen to function as a checklist of things that should at least be considered in the responsible design of AI peace technology solutions. Thirdly, similar considerations and tools could also be relevant in guiding the design of AI solutions with intended social impact at any sector.

The results presented in this chapter are divided into two sections. One presenting the three opportunity areas identified and another presenting the design tools developed.

5.1 Future opportunity areas for AI in peace technology

The results responding to the first case study question on future opportunity areas for AI in peacebuilding are condensed into three separate, but interlinked areas that hold potential for developing applications of AI peace technology in the future. Rather than an exhaustive list, they are a compilation of what are viewed as the most critical and strategically promising areas based on the analysis combining different data points of this development work. For each opportunity area main findings are presented in this chapter. The areas identified are then summarized in the next chapter using the peace technology canvas developed in this thesis.

5.1.1 AI assisted conflict analysis

"Complexity and polyphony of today's conflicts is truly a wicked problem." (Interview, Peace and conflict expert 2018)

Conflict analysis is used to "*establish an accurate understanding of the root causes, proximate causes, triggers, dynamics, and trends of conflict as well as stakeholders involved, impacts on the people and the operational environment*" (United Nations 2016, 2). A good conflict analysis should also include existing and planned responses to the conflict. It is a basic tool to understand what is going on, to identify gaps, map out options and present realistic response strategies for the conflict. (European External Action Services & European Commission 2015, 2.)

The following minimum elements should be present in all conflict analyses: 1) situation profile 2) causal analysis of conflict factors 3) stakeholder analysis 4) conflict dynamics and drivers of change. Conflict analyses are carried out when deciding on international peacekeeping operations and other interventions in conflict environments. For the analysis to be useful, it should be neutral, participatory, balanced, multidisciplinary and take issues such as gender and human rights into account. By definition, conflict analyses are always contextually specific and unique. They are resource intensive undertakings and need to be updated frequently enough (e.g. every two years). (United Nations 2016, 3-4.)

"We need to work with the reality that conflict analyses are always shitty" (Interview, Peace and conflict expert 2018)

"Conflict analysis should be the basis for everything. It is healthy to assume that your own analysis is incomplete." (Interview, Peace and conflict expert 2018)

Conflict analysis was the most mentioned individual issue by the peace and conflict experts. Rather than a static report, an improved conflict analysis was the most important need identified. Keeping an accurate, comprehensive and up-to-date picture of what is going on as well as being able to make sense of a highly complex situation presents a major challenge for peacebuilding efforts. Without an accurate picture, it is difficult to decide what are the best strategic interventions in a given situation. A thematic clustering of illustrative interview data quotes on conflict analysis is presented below.

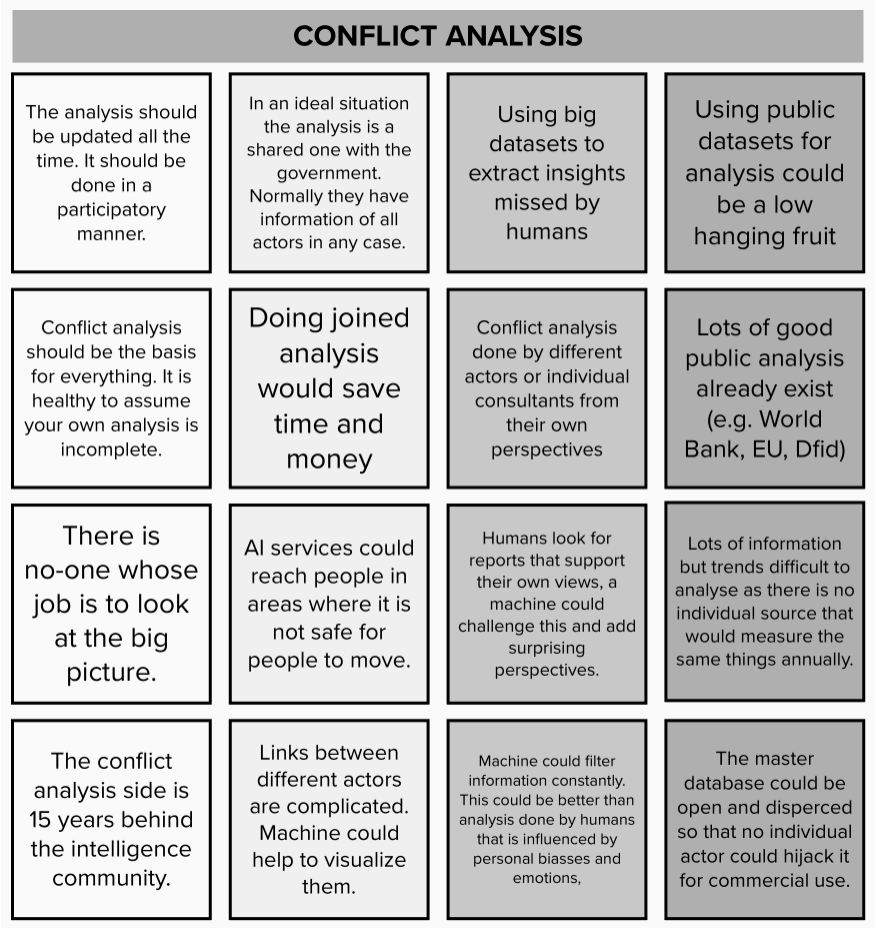


Figure 12: Affinity diagram of illustrative interview quotes on conflict analysis

Real-time information on multifaceted and quickly evolving local events is a necessity for everyone living in conflict areas and working on peacebuilding. Daily operations of conflict management, the security of staff on the ground, the ability to prevent escalation of violence were among the issues identified to benefit from an up-to-date reading on what is going. An incorrect interpretation of an individual event could carry serious implications for security of individuals, for the conflict escalation or for the peace negotiation dynamics.

The interviews highlighted that even though important, updating conflict analysis to correspond to current realities on the ground is often not allocated to an individual entity nor person within an organization working in a conflict country. Rather different **organizations tend to carry out their separate conflict analyses**. Reading a complex situation in silos is not an efficient way of working, nor likely to give the most accurate picture.

Different actors present in conflict environments, such as the United Nations, EU and various embassies and NGOs, gather a lot of information. Studies, plans, monitoring and evaluation reports and memos are common and much of this data is public. Measuring impact and comparing different interventions is however challenging, even when showing impact is key for

funding purposes. **Subjecting the existing data to AI analysis** was seen as a potential low hanging fruit to gather new insights with the help of AI and with little risks involved.

"If you compile information that is already public to an easy to digest form, what are the potential risks?" (Design probe 2018)

Utilizing the existing information cumulated from a given conflict would, however, require coordination between different actors that own and produce the information. Access, anonymization and standardization into a more AI friendly format were among the challenges mentioned. The AI expert interviews pointed out that anonymization and "cleaning" the data are major expenses and often practical hurdles in developing any AI application.

The major advantage of using AI in conflict analysis was seen to be its neutrality. Peace and conflict experts highlighted that AI's value added was not being affected by emotion, individual aspirations, "talk of town" or office politics that all were seen to have an impact on the traditional conflict analysis. Humans are prone to look for information that supports their existing perceptions, and thus machine analysis could potentially generate surprising insights. The AI expert interview data pointed to clear limitations to this expectation.

"Mining for linkages and meanings from a large set of data is extremely dangerous. For example, in the USA there has been an attempt to estimate crime renewal rates and punishments using AI assistance. The models lead to the fact that people of color are more likely to go to jail. The learning data is biased, not the system. Machine learning system will learn to be racist because that's the way the learning data is." (Interview, AI and technology expert 2018)

Among the peace and conflict expert data there were expectations that AI assisted systems could **recognize patterns that could not be detected by humans**. AI would be able to process immense amounts of data from various sources. The superior data processing capacity of AI combined with the ability to extract insights also from the flow of data or "meta-data" was among key insights arising also from the workshops. Workshop results indicated that the technical **capacity of AI to identify weak signals from large sets of data** was seen to be of strategic importance in supporting peace efforts. The role of humans was seen in interpreting and assigning meaning to the patterns recognized by AI.

Creating conflict analysis databases

For AI assisted analysis to happen in a given conflict context, creating **cumulatively learning, decentralized and open databases** is needed**.** Analysis of existing datasets combined with

improved data visualization would help to understand complexity when making both long-term strategic as well as daily operational choices. **C**ombining information from different, already existing sources was seen as a starting point to analysis that helps to follow the big picture as well as to spot new insights in an efficient manner.

Data visualization was closely linked to discussions of AI. For the hectic daily work of peacebuilding it was viewed as an important tool for making sense of complex situations, multitude of actors and their affiliations. Keeping track of who are the stakeholders, how are they involved in the conflict, which groups are active, who supports and finances who is already a challenging task itself. Clear visual presentation of situations and actors was pointed out to be crucial for making informed decisions at all levels.

A practical solution idea of a **"Peace Google",** a database analyzing existing reports of different organizations working on the same conflicts was developed in the workshops. This could maximize the use of existing information, break silos between different actors involved in the peace efforts and strengthen institutional memory in situations where a lot now rests on individual experts. Updating and maintenance of such system would have to be resolved in the future.

"In order for us to really help the peace process foreigners need to commit, at individual level, to working in this country for several years. Or we need a proper system that collects information to institutional memory and that analyzes the data factoring in prior attempts of solving the same problem. The wheel has been re-invented way too many times." (Design probe 2018)

However, as AI experts interviewed and participating in the workshops pointed out, using different types of data sets, particularly qualitative and non-standardized ones in multitude of languages means very labor-intensive preparation for any type of AI analysis. **Standardization of data collection** would be a necessary future step, requiring a new level of openness, transparency and collaboration at least among major actors involved in peacebuilding.

A dynamic, continuously updating conflict analysis database would require coordination and establishing contracts and channels for sharing data. This would all need to be agreed separately. Finding a balance between the costs and the benefits could initially be difficult. Peacekeeping missions alone involve thousands of people. A common database between different organizations would mean major changes in the way of doing business. If successful, this was seen to carry significant efficiency gains.

Establishing a more transparent and democratic cumulative memory in a situation of high staff turnover could also assist in capturing some of the silent knowledge that is now excluded from official reporting. Issues to consider when building a common database for conflict analysis included level of transparency and access. Should the results of AI analysis be open for everyone or only for the use of experts? Concerns of safety and using the information for military purposes were mentioned. However, it was pointed out that **a** database excluding potentially sensitive information would most likely still leave enough data for a useful analysis.

Hotline for community peace mediators

"How to get information from hard to reach areas is the most important question in this country right now. Who lives there? Under what conditions? How are they surviving?" (Interview, Peace and conflict expert 2018)

A practical solution proposal developed in two workshops was an idea of a report back hotline for community level peace mediators. A call-in number combined with NLP technology was suggested for collecting the valuable information gathered from local communities. The reports would cumulate anonymously into the conflict analysis database. This would assist in keeping the analysis up to date and including hard to reach qualitative insights.

A combination of "old" (phone) and "new" (NLP) technologies was seen as inspiring as it could help to tackle a common obstacle of conflict environments having no internet or satellite connections. This type of solution would require careful training in organizing the data collection while building a strong accountability loop. It was pointed out that collecting data without making the analysis available at local level, would not be acceptable. Creating an inclusive system should not be about extraction of information, but a process of co-creation and mutual gain.

"If you need to leave out all equipment and note down the data qualitatively, based on your memory, in practice it is already unusable. Unless the data collection is carefully trained in advance." (Interview, AI and technology expert 2018)

A representative and inclusive data sample is challenging to achieve anywhere. In situations where neither infrastructure nor local culture support equal participation, using a variety of sources could help to triangulate the accuracy of data. The kind of data and its sources were debated at length in the workshops. Technical ability of NLP on local language materials was among the limitations identified by AI experts. Ownership, trustworthiness and potential biases in data were also considered challenging.

Using an intermediary such as a community worker for collecting the data used in AI analysis is not an uncommon alternative. But could verbal reports of community peace mediators represent the concerns of women, disabled or other minority groups that are all relevant to peace processes? Recognizing what kind of data is suitable for AI analysis needs specific capacity. Responsible design means finding ways for a meaningful and inclusive participation to the development and use of AI peace technology.

"Garbage in, garbage out. Your model is as good as your data." (Interview, AI and technology expert 2018)

Utilizing small and fragmented datasets was recognized as a major challenge common to future applications meant for conflict areas. A lot of resources might be consumed for verifying the credibility of data and reliability of its sources. Questions of how to know what data is available, who owns the data and who can grant access to it all carry not just financial implications in conflict settings. What does it take to clean up the data? And is it worth it? These were the questions raised by the technology experts. While there are no easy answers, a thoughtful data collection design was described as imperative for security reasons.

5.1.2 Early warning systems

"So much attention is given to what is happening so that we can take action. Why something is happening and how to prevent it gets far less attention." (Interview, peace and conflict expert 2018)

AI is used to predict things like consumer preferences, trends in the stock market or even floods, famines and earthquakes (Interview, peace and conflict expert 2018). Pilots on predicting more complex social issues like the likelihood of child-protection services or probabilities of dropping out of school have also happened (Interview, AI and technology expert 2018). **Using AI to help in detecting the signs for increased conflict potential** was identified as a future opportunity area in the interviews, probes and workshops. Applying the existing technical capability of AI to peacebuilding could benefit all stakeholders from local people to international organizations and businesses alike.

Increasing human security, supporting peacebuilding and humanitarian efforts, helping businesses working in fragile environments were among the gains linked to creating better early warning systems that were commonly seen as a win-win. Conflict prevention is considered the best and most cost-effective way to tackle conflicts (United Nations & World Bank 2018). Even more than saving money, the opportunity to prevent human suffering makes a strong case for tackling conflicts in advance. However, as both the workshops and interview material

pointed out that there should be **no assumption that increase in capacity to predict conflicts would automatically lead to increase in action** to prevent them.

"Even with low-tech we already know when things start boiling and are about to explode. The real question is what do we do with this information? Would high-tech help to mobilize the necessary political will to take action?" (Interview, Peace and conflict expert 2018)

"It is not about having the information or knowing something is about to happen. The data is there, but it is not utilized or taken seriously. The real question is how do we get the data at the right time, for the right use." (Interview, Peace and conflict expert 2018)

For businesses making predictions are part of risk analysis. In a conflict situation an early warning of violence escalation would give time for measures to change the development trajectory, to stop the situation from getting worse or to find ways to mitigate human impacts. In resource scarcity the forecasting could be used to prioritize the urgency of assistance needs. It could also help to make investment decisions and report on impact. Using predictions to decide on action was seen equally challenging as deciding what data to use to make the predictions.

"What is the dataset for predicting conflicts in a given country? How can we generalize that? Could data from political elites and financial flows be used to forecast conflict escalation? How about prevention? What happens right before the conflict erupts? Is there something that could be used as an indication that something is about to happen? What is the relevant data set to analyze here? Could discussion among the diaspora be used as an indicator? (Interview, Peace and conflict expert 2018)

The datasets needed to predict conflicts were seen to be rather unique for each context. Outgoing financial flows, cross-border trade, diaspora discussions or price of camels were among the suggested indicators for conflict escalation. Similarly, as with previous opportunity area, the ability to analyze small and fragmented sets of data, including those in local languages, was seen as important, but challenging. Most conflict contexts would have no large data sets available. Partnerships between the private sector actors in possession of potentially relevant data and the civil society or public sector actors with deep understanding of local culture were flagged as something to explore in the future.

Using metadata and the lack of data to make simulations and predictions was recognized as an opportunity. At the same time concerns of **data quality** relevant also to non-conflict environments were raised in both interviews and workshops. **Trustworthiness** and finding trusted sources of data in conflict areas was a mutual concern to AI and conflict experts. Extra care

was needed to avoid deliberately altered data and to secure **ethical use of data.** Protecting privacy and detecting biases are relevant questions for all AI development, not just for peace technology.

”When you look at a lot of things at the same time, you most likely won’t find direct causalities but could still probably predict conflicts. For example, in trading there are a whole bunch of pseudo-scientific predictions being made already. One way is to do data mining and look for correlations in data, another is to identify things like surge in unemployment among young men which could lead to conflict.” (Interview, AI and technology expert 2018)

Equally important to data ethics is finding the right questions to ask. A better understanding of the overall conflict context and potential impacts planned interventions in it could prove to be more useful than simplistic predictions. Rather than a quick fix, a more systemic approach to using AI analysis was seen necessary by peace and conflict professionals. AI was expected to help in understanding the underlying complexities and dynamics of unique conflict contexts. This in turn would strengthen the ability of peacebuilding efforts to better choose actions strategically.

”We should not concentrate on predicting the next conflict but rather on understanding the context and deeper dynamics. This way we can act better when there is a chance to reinforce peace." (Interview, Peace and conflict expert 2018)

While the idea of forecasting and the technology to do it are not new, using AI in the context of peacebuilding was considered to bring new value. It would enable analysis of vast amounts of data and use of more indicators than traditional analysis could. Cumulative learning and inclusion of more diverse datasets across disciplines to break peace expert peer-group “bubbles” was seen as a key strength. **The same context specific databases needed for AI assisted conflict analysis would benefit the building of early warning systems.**

5.1.3 AI support to human communication

Perhaps closest to the original concept of Peace Machine by Honkela (2017), using AI to help people understand each other better, to communicate better and through that to support peace was an overarching theme in collected data. Particularly relevant to pre- and post-conflict settings of prevention, reconciliation and sustainable development, but also to peace negotiations. **AI support to human communication is the third opportunity area identified**.

Utilizing natural language processing (NLP) and sentiment analysis were seen to have potential to be used also as peace technology. The technology demonstration illustrated both the

ability and the limitations of NLP and sentiment analysis in a group discussion setting. As there are a multitude of existing commercial applications, this type of technology is however, likely to develop quickly. From IBM Watson type of solutions to more sophisticated versions of NLP, the expectations for commercial use are significant. The interviews and workshops highlighted that the interest in peace technology could take advantage of the commercial application technological development.

"How do we build peace negotiations based on the analysis? Who defines what? Is the dialogue design defined by a Finnish peace professional or is it a true co-creation? What do you do with whom? Who is heard? What is the focus? How do you bring all the different opinions together?" (Interview, Peace and conflict expert 2018)

Peace negotiations are among the most important steps towards ending violent conflicts. The interviews with peace and conflict experts highlighted that peace negotiators and mediators are currently a highly respected global community of experts. Supporting their work with AI in **dialogue design or in planning of peace negotiations** was seen a promising area to explore. AI assisted analysis could hold potential for **extending the inclusivity of an exclusive peace negotiation process.**

"The starting point of classic peace negotiations is getting people committed to each other and to the process. As you move forward, they have more personal level buy-in and the threshold to leave the negotiations gets higher. Here is an opportunity for methodological rethinking. How to combine the exclusivity of confidential peace negotiations and the need for inclusive participation their ultimate success requires?" (Interview, Peace and conflict expert 2018)

Finding common ground among the conflict parties and identifying potential contention areas in advance can support the successful outcome of peace talks. As each peace negotiation team has limited understanding of the complex situation, AI analysis could help to pinpoint areas that might be overshadowed due to human factors such as exhaustion, cultural differences or personal biases. The ability to consult a wider range of constituencies s prior or even during the negotiations would cultivate broader ownership of what is agreed in the negotiations table.

Studying the flow of peace negotiations was seen to have potential in **building respect among negotiating parties and in reinforcing people's personal commitment to a positive outcome**. Building a cumulative database for each unique negotiation process helps to narrow down a complex situation, the task at hand and the data set available to resolving it. Accord-

ing to AI experts, analyzing the meanings and links between words in web discussions is already widely in use. Applying that to peace talks as material or to social media for hate-speech detection could help to prevent conflicts and to support the reconciliation.

"Most likely there is little need to invent new algorithms for peace technology, but rather re-assigning the use of existing ones" (Interview, AI and technology expert 2018).

"Perhaps at the whole of society level there are bubbles and disruptive discussion habits. Algorithms could help to bring people with different views together and break those bubbles." (Interview, AI expert 2018)

Analyzing hate speech is also used for peace technology and other purposes. A test carried out during Finnish municipal elections is an example of an existing application raised in the interviews. While there are high expectations related to technology, applying AI to complex and contested issues is not straightforward. (Laaksonen et al. 2020) Using the AI to break rather than build social media bubbles was discussed in interviews and examined in the workshops. **Building alternative narratives** was seen as important part to this. AI could support people's efforts to imagine what peace looks like, to break us-them divides and to offer easy-to-digest exposure to views different from one's own.

"Where I see the strength of the peace machine type of approach is when different types of calculations are applied very locally to make different kinds of equalizing decisions. This will lessen the seeds for conflicts." (Interview, AI and technology expert 2018)

Micro-interventions based on NLP could give time to de-escalate the situation before it becomes a conflict. The design probes shed light on the ways rumors and miss-information can worsen the conflict quickly. The notion of "the other side gearing up" (Design probe 2018) is enough to increase the potential for escalating violence. The ability to quickly confirm the accuracy of information (e.g. did an explosion happen the way it is said to have happened) and providing more access trustworthy reading on the situation would help to stop rumors and tackle the escalation.

"*Conflict is not about people not understanding each other. The motivation can simply be to cause trouble"* (Interview, AI and technology expert 2018)

At the level of individual and community **nudging for peace** type of interventions were cited both by the AI and peace expert interviews. In practical terms it would be sending alerts to the person about to post hate-speech on social media or giving enough time for a community

level worker to mediate the situation before it gets violent. Different models of people's biases visualized could also support personal reflection. Like commercial versions that tell you what type of music you have listened to this year or how many steps you have taken today: observing behavior is also a way to change it. The strength of AI in this is its ability to monitor vast amounts of data and intervene immediately.

Concrete solution proposals that were developed in the workshops for the opportunity area of AI support to human communication were:

1) **Peace talk databases** to support each unique negotiation process. It could be used for alerts on contentious language in the negotiation table and for peace negotiation design by identifying common and contentious issues in advance. The databases could also be utilized for studying the negotiation flow and for training future peace negotiators and community mediators.
2) **Alert systems of micro-interventions** or "peace nudging" to tackle hate-speech in social media forums or other defined environments such as election or policy making discussions.
3) **Global "Must reads"** as targeted discussion forums with the objective of exposing people to opposing views in a manner that is tailored to their personal preferences.

5.2 Tools for responsible design of artificial intelligence peace technology

Based on the insights gathered in the development work a set of three canvases were formulated to answer a second research question about how to design AI peace technology in a responsible manner. These design tools summarize the most central questions on technical and data issues from the side of AI and on context and ecosystem from the side of peacebuilding. An iterative process was used to develop practical tools that function as a checklist of what should at least be considered in responsible design of AI peace technology applications.

From the very early stages of planning the workshops that form a part of the collect phase of this study, it became evident that existing design tools and canvases would need to be modified in order to have meaningful discussions and gather insights on the use of AI for peacebuilding. Issues of security, preemptive risk analysis and multidimensional ethics would have to be included to have a prominent role in driving the design process forward. Or if necessary, stopping it. Low threshold prototyping and re-iteration based on participation or testing by actual users are often difficult to carry out in conflict environments. This is why the importance of careful design is underscored.

"For sure move fast and break things happens also in peace work. However, the intention should be something completely different. Move fast and do no harm is the minimum." (Interview, Peace and conflict expert 2018)

The set of three canvases were initially developed to be used in project workshops. A cycle of re-iterations happened from initially using the summary peace technology canvas as a way to structure the discussions in the workshops. It directed the participants to consider issues for both the problem space and for the solution space. Changes were on the lessons learned from each workshop as well as on the insights emerging from interviews. The notions on responsible design – what one needs to consider when developing AI peace technology - emerged in the data naturally and often unprompted.

Key insights on design issues from the workshops' report back discussions were documented in notes and used to inform the analysis and particularly the development of the canvases. They are summarized the table below.

Workshop 1: What is peace technology?	**Workshop 2: Human-machine co-operation**	**Workshop 3: Teaching the machines**	**Workshop 4: Project's final workshop**
There are no universal nor large-scale solutions to today's local and fragmented conflicts. Even when similar underlying reasons might exist, resolving conflicts needs context specific solutions.	Quality, ethical use, access and privacy of data are of strategic importance for AI peace technology design and use.	Data quality is a challenge even in non-conflict settings. There needs to be ways to include considerations on quality, reliability and ethics into choosing the teaching data for AI peace technology.	Responsible design is not an alternative, but a must for AI peace technology. The design tools developed provide an early minimum checklist for what should at least be considered.
Without participation of everyone there is no sustainable peace. Therefore, AI peace technology should also be participatory, inclusive and mindful of societal and cultural considerations.	Key strength of AI is analyzing huge amounts of data. It is up to humans to evaluate the quality of data and correctness of analysis as well as assign meaning and decide on actions.	Data ethics should be a central part of responsible AI peace technology design process. Safety in data collection and protecting the privacy of sources are challenges to conflict and non-conflict settings.	The opportunity areas identified, and design concerns raised condense the relevant issues for using AI for peacebuilding. Many considerations of AI peace technology are applicable to other sectors.
Peace technology should always place the needs and problems of the people affected by conflicts at the center. Focus	Defining clear roles and responsibilities for machines and for humans at early stages of design process helps to build on	Tackling trustworthiness of data in conflict settings is a must. Extra care is needed to spot deliberately altered	The tools developed for responsible design of AI peace technology could be useful also for non-

Workshop 1: What is peace technology?	Workshop 2: Human -machine co-operation	Workshop 3: Teaching the machines	Workshop 4: Project's final workshop
needs to be on each unique case.	complementary strengths.	data that might come from politically motivated sources.	conflict environments, in combination with existing design tools.
Scalability cannot be taken for granted, nor as a starting point for AI peace technology. Establishing a business cases could be a key challenge.		Acceptable margin of error in AI peace technology is extremely small. Building and maintenance of AI applications more challenging and costly in conflict environments.	We need to differentiate peace technology from other technologies. Solutions need to be local, context specific and concentrate on making positive impact on people affected by conflicts.
		Not just data but also the lack of it could be used as an indicator for early warning of conflict escalation.	AI peace technology needs to be end-user centric yet take the impacts on the bigger picture into account. Comprehensive, society-wide risk analysis should be part of design.
		Understanding practical limitation of context (lack of infrastructure, internet connection or electricity) should be part of AI peace technology design.	Same technology should not be expected to work in different contexts. Establishing a business cases for AI peace technology can be challenging.

Table 2: Workshop insights on design of AI peace technology

The centrality of safety, ethics and risk analysis should be a holistic and cross-cutting consideration in the whole process. However, they should paralyze. The workshops demonstrated that bringing together expertise from different fields, including those of peacebuilding and AI, is necessary and simultaneously energizing. However, translating that into real-life solutions needed more concrete guidance and tools.

It was clear that responsible design and use of AI peace technology was seen as a must, rather than nice-to-have. Business as usual way of designing would not suffice. The solutions should always be context specific and they could not be expected to work in different areas. Instead

of profit and scalability, the goal should be to improve the lives of people in conflicts. Most importantly, the design would have to be serious about minimizing the risks involved in the development and use the technology. Potential impacts should be considered in advance and at multiple levels. Thus, responsible design of AI peace technology could not be done just with the user or beneficiary in mind but should always consider the wider societal and cultural context.

Being aware of complexity and being able to integrate it into the design are two different things. The canvas set developed are a practical attempt to tackle complexity that both conflicts and peacebuilding as an environment and AI as a technology bring. According to the feedback received from workshops, the canvases can help to brings together experts from different content and technical fields.

The canvases developed are intended to combine these considerations into an easy to use format to offer concrete tools on how to carry out for responsible design. The final iteration of the set includes three canvases, intended to be used at early stages to design process. They are 1) **peace technology summary canvas**, 2) **impact canvas** and 3) **solution ecosystem mapping canvas.**

Peace technology canvas

For a more concrete problem and solution definition, an initial checklist type of summary canvas was first developed for the Human-machine co-operation workshop. The summary canvas mirrors the idea of a classical business model type of canvas. The first iteration was further elaborated for Teaching the machines workshop. Three problem scenarios identified in the previous workshop were added to function as a starting prompt.

While the first iteration focused on defining the right problem to solve, the second emphasized building solutions using collective strengths of AI and people. The idea was to guide thinking about the role of human and of machine as part of the solution development. Guiding questions on data needs and expected and unexpected results evolved into consideration on impacts.

The whole set of canvases were developed for the final workshop and shared on project website (Understanding Peace 2018). They were then tested in a separate workshop and finalized. The main goal of the canvases is not to re-invent something that is already there in terms of design tools, but rather add ways to include AI peace technology specific considerations. The canvases as a set are designed to offer an easy to use and open for anyone type of framework for thinking.

Peace technology canvas should be considered as **a minimum checklist for responsible design of AI peace technology applications.** It is a summary tool that gives an overall view to the intended AI peace technology solution. To fill the individual boxes, the solution ecosystem mapping or the impact canvas should be used as per instructions in the canvas. Under the problem definition space, the canvas prompts to answer questions on user, beneficiary and other stakeholders that could be affected by the solution or needed for its development. In addition, impact is suggested to be considered for both the user and for the bigger picture. For considering the risks – both intended and unintended impacts – use of impact canvas is suggested. Ways to mitigate the negative impacts is also added as a question.

In the solution building space, the central question is considering the intended change from the perspective of the people affected. Both role and capacity issues of humans and machines are considered separately. Furthermore, questions around data and its ethics are included. Finally, the format of the solution asks to consider distribution and accessibility. The bottom part of the canvas is about scalability and financing. The canvas is presented in picture below and included in appendix 8.

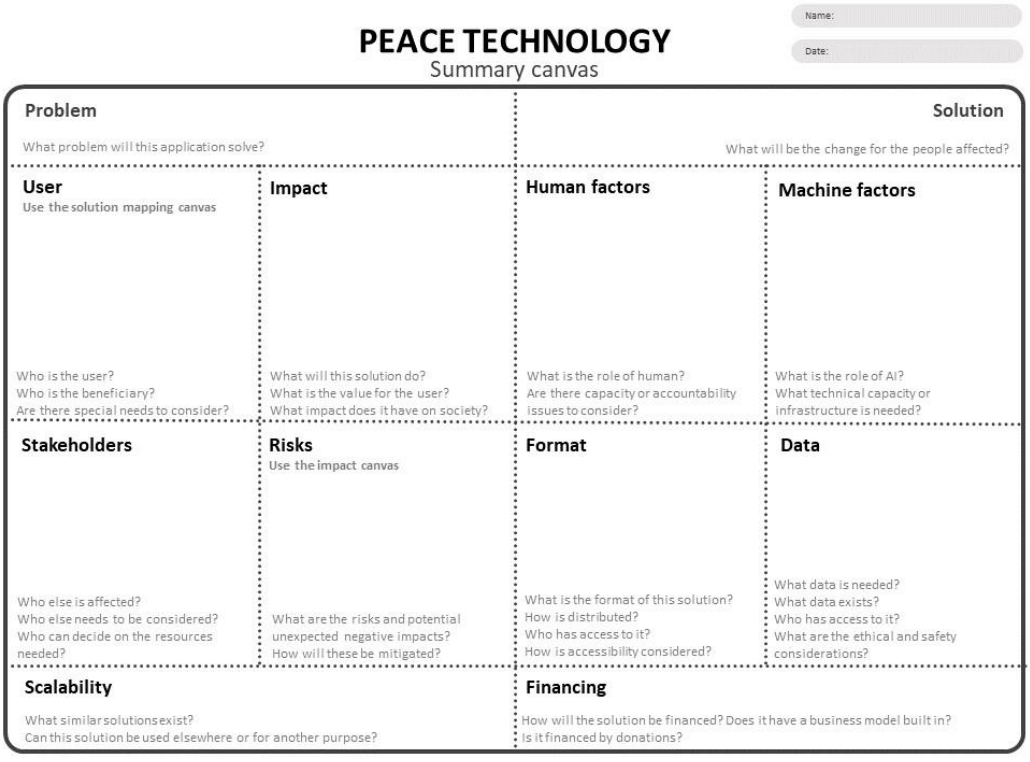


Figure 13: Peace technology summary canvas

Feedback from workshops indicated the canvas to function particularly well for **comparing potential solution ideas at their early stages.** It was seen useful in choosing between ideas

before investing in further development. The canvas could thus support decision making on where to use limited resources and in considering at concrete level what type of expertise is needed for the development and whether using AI is the best way forward. The canvas testing workshop participants saw the canvas to work for initial proof of concept phase of design, or for a light desk-top walk-through testing of a preliminary concept idea. Knowing that minimum necessary concerns are addressed gives confidence to move forward or if necessary, stop the innovation process.

Impact canvas

Both the impact and stakeholder mapping, now solution ecosystem mapping canvas were developed and tested mainly in the last two workshops. The impact canvas was a way to make consideration of positive and negative, direct and indirect impacts of the potential AI peace technology solution compulsory part of the design process. When combined with the solution ecosystem mapping, the idea was to guide thinking towards different levels of possible impact. The canvas is presented in picture below and included in appendix 8.

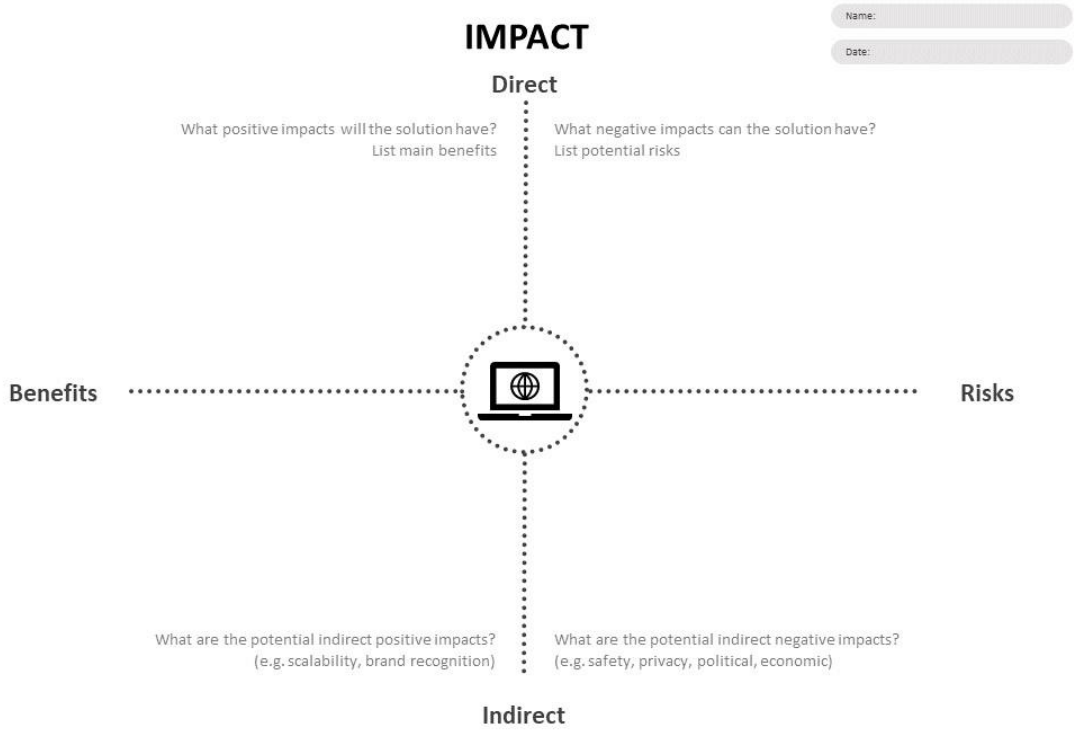


Figure 14: Impact canvas

The idea of impact canvas is simple: it **is a concrete tool to consider different levels of impact** of AI peace technology solution at early stage of the design process. The feedback received in the workshops was overwhelmingly good. Having a separate tool to think through the wider and also the unintended impacts of technology solution was seen necessary and with AI peace technology, crucial. While recognizing the variations in the terminology used in different sectors, the term impact is chosen to cover output, outcome and impact levels that should all be included in the design.

Solution ecosystem mapping canvas

What was developed as a stakeholder mapping canvas was reiterated to solution mapping canvas to depict **the need to place the AI peace technology application into its wider context**. It makes visible the potential differentiation between the intended user and the intended beneficiary, that can be present also in other sector AI applications. The goal is to include a more systemic thinking to basic stakeholder mapping and through encourage to carefully consider the potential multidimensional impacts and human implications the solution can have at different levels. The canvas is presented in picture below and included in appendix 8.

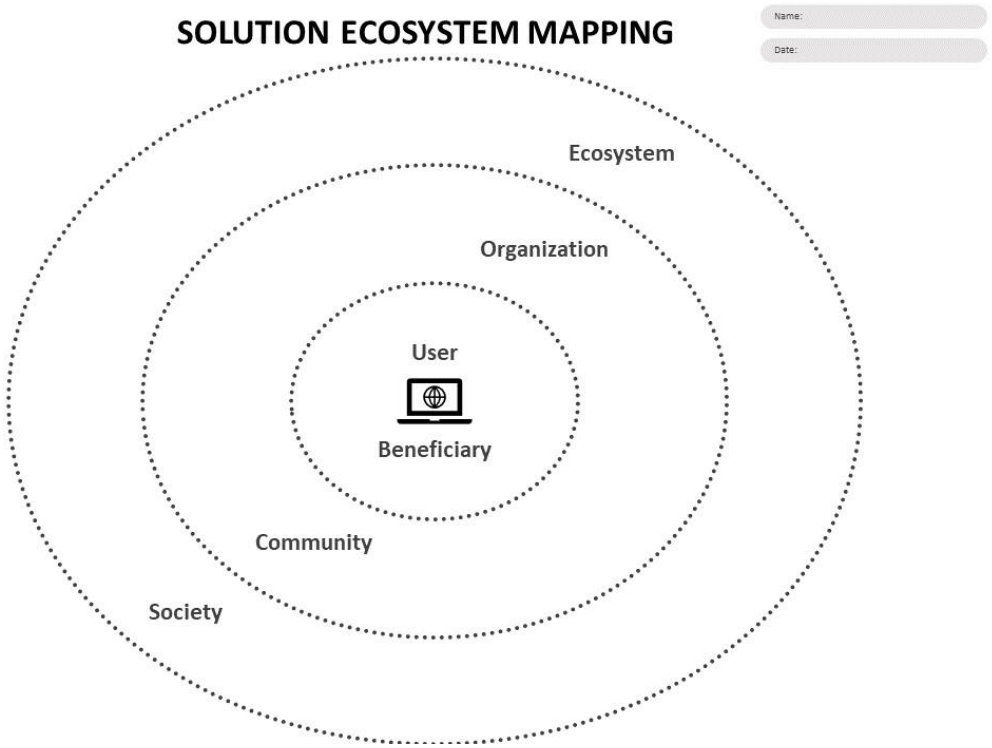


Figure 15: Solution mapping canvas

”The same user-client contradiction of interests is present in every digital project, even if you are developing a travel reimbursement system.” (Interview, AI and technology expert 2018)

Feedback from the last two workshops indicated the distinction between user and beneficiary to be a common one. But more importantly, the earlier versions of the canvas (then “stakeholder mapping”) directed the workshop participants into mapping the conflict parties rather than the stakeholders of the solution itself. This is why the solution mapping canvas was renamed and it is directed to be used in defining the potential user, considering if there is a distinction to the intended beneficiary and then placing them into their respective wider contexts and networks.

Lastly, the solution mapping canvas can support in thinking through 1) the different people and their surrounding communities that might be affected by the solution or be relevant to building it, and 2) the ecosystem where the solution needs to be integrated for its successful implementation and to which it can also have an impact.

6 Learnings and conclusions

The purpose of this thesis was to understand the opportunities and limitation that AI could have in supporting peace. The goal was to identify areas where the technological development of AI and the needs of peacebuilding meet and overlap. The development work did not set out to design individual concepts for AI peace technology applications, that might be a likely objective for a service design thesis. Rather the overall aim was to test the idea of using the technological potential of AI to solve complex social challenges, in this case violent conflicts, in a case study setting. The development work also examined the role of design and developed tools for responsible design of AI peace technology.

The thesis ended up stepping backwards to look at the bigger picture: the future opportunity areas based on understanding the needs and realities of peacebuilding and the technical capabilities and limitations of AI. Rather than exploring the inspirational idea of a “peace machine” in more practical terms, the development work gathered insights to explore the future potential of AI peace technology and usefulness of service design tools and methods. These are both steps towards responsible real-life applications.

This chapter draws together the development work conclusions and examines the results in dialogue with the theoretical framework. The development work identified three opportunity areas where AI could be used to strategically support peacebuilding efforts. The results are summarized using the peace technology canvas developed and their meaning and limitations

discussed. The canvas is used in a tailored manner as it is not designed to describe larger opportunity areas per se, but rather to be used in more concrete, context and technology specific solution development.

The set of three canvases is a way to answer the second research question of how AI peace technology can be designed responsibly. The strengths and weaknesses of the design tools created, and the considerations identified in the development work are discussed in relation to the theoretical framework. Evaluation of the work and suggestions for future research are outlined in the chapter 6.4.

6.1 AI assisted conflict analysis tackles complexity

The first opportunity area identified builds on the AI capability of analyzing large amounts of data from various datasets. In order to practically function in wicked problem like conflicts, a better and a more dynamically updating way to understand extreme complexity needed. Rather than an exercise carried out as part of peace intervention planning and updated every two years, **AI assisted conflict analysis** could provide a live tool that supports the peace efforts on the ground as well informs the decision making happening at international level. With the help of AI, a shared overall picture and ability to zoom into details could be possible.

AI assisted conflict analysis is hoped to be a combination of more concrete, working level information and of following the bigger picture type of analysis at the same time. If successful, conflict analysis could be a Google Map type of collection of individual things to zoom into and that together form an overall picture of the conflict (Mulgan 2018). It would be **an assembly of smaller applications functioning together**. Applications in this area would serve one of the three key purposes British Council (2018, 4) outlines for technology in peacebuilding: **data management, collecting, analyzing and visualizing information**. To smaller degree such living analysis could also support the dialogue and mobilization and create new spaces for a broader variety of peace actors and actions.

To build peace, a holistic and complementary mix of strategic interventions happening in all conflict stages and levels is needed. Making decisions and keeping track of this is a difficult and resource consuming task for any individual organization. Instead of everyone carrying out conflict analyses individually, with their own methods, **a joint effort would mean both effectiveness gains and standardization**. Comprehensive and up-to-date conflict analysis aided by data visualization would provide quick access to correct information and ability to situate individual events into the bigger picture of multilevel conflict squiggles. This can save lives, inform operational decisions and through that help to de-escalate situations.

AI assisted conflict analysis can help coordination between different development, security, political and humanitarian actors. This would support the integrated approach to conflict prevention promoted by actors like UN and World Bank (2018). Working in silos is not enough to tackle modern day conflicts. Thus, a shared picture, aided by AI analysis that is perceived more neutral or guided by emotions and political interests like humans, would offer a shared basis joint decision making. The same AI augmented reading of the events, stakeholders and their interlinkages that is vital to human security and carrying out daily operations, can be of strategic importance to achieving sustainable peace through bringing together the various levels of interventions.

For AI assisted conflict analysis to be more than another viewpoint of what is going on, it **must place the peace objectives at heart.** This means gather information relevant to peace efforts and making analysis from the perspective of what can best support peace in a given conflict situation. It also means making sure the information is safe and it will not be utilized to further aggravate the situation. This is not possible without humans assigning meaning and making decisions both on how the analysis is carried out and for what objective as well as for what the analysis means and what kind of decision should be made based on it. In practice expertise on peacebuilding needs to be in the lead when defining the right questions to ask. Collaboration with technology experts is needed to make sure those questions are suitable for AI to resolve.

So rather than concentrating on AI, such conflict analysis needs to be an effort of **collective intelligence** to work in real life. An assembly of different things – big and thick data, different types of visualizations and collection of analysis carried out with the help of AI – would help to form an overall picture, making complexity simple enough for humans to tackle. (Mulgan 2018, 27-30.) Individual datasets present a partial picture. Using the ability of AI to bring together more data and points of analysis than single human or organization could do is the strategic role of AI in conflict analysis. The role of human is to assign meaning to the information, or in other words form a reading on what separate parts of analysis put together indicate and then decide on action.

The combination of data, analysis, visualization and assigning meaning can help to understand complexity in a way not possible for individual human to grasp. To develop and use such a solution, the **human-centered and systems-minded approach** (Both 2018) would have to be put to use. As a solution it could assist in the vital task of understanding relationships, interlinkages, different levels of intervention and impact. AI ability would thus help human capacity in choosing actions to prevent, stop, manage and mitigate violent conflicts. An ongoing development and updating is a requirement for success, something very familiar to technology development and more unfamiliar to peacebuilding organizations.

The AI assisted conflict analysis has potential to be used if not in all, at least in many different stages of conflict to support responses relevant to transformation, settlement or containment. (Ramsbotham et al. 2018, 16.) Or efforts to prevent and end violence as well as to tackle issues that create the need for it. Furthermore, such analysis would support moving to rebuilding a functioning society through economic and democratic development interventions that are part of normalization and reconciliation phases (Ramsbotham et al. 2018, 16).
In order to capitalize on AI assisted conflict analysis **existing datasets** have real potential to be used as a starting point to build databases on individual conflicts. Taking advantage on the multitude of information that already exists, much in public domain, would be in line with Miklian & Hoelscher (2016) that urge creation of knowledge bases for culturally sensitive peacebuilding. They can help to integrate local user needs, customs, beliefs, values and marginalized groups to be part of peacebuilding.

In real-life much of the existing data is fragmented, not standardized for machine analysis and owned by individual organizations or businesses. A model of how to bring together an overall set to function as a database for an individual conflict would be next step to test in practice. What kind of data is suitable? Who owns it? What kind of agreement are needed for that data to be accessed? Is it worth the effort? Building conflict databases needs more than individual pilots. It means actual and ongoing collaboration and sharing of information between organizations that are not necessarily used to doing so. The role of peacebuilding and business heave weights is central, yet they are likely to be slowest to embrace a radical change in the way of doing business.

To be applied in practice, any such solution needs to be integrated into the ecosystem of peacebuilding. An AI assisted conflict analyses needs commitment of large organizations to make the initial investment in gathering and anonymizing the data as well to develop a model for it. A set of criteria specific to each unique context for standardizing the gathering of data in the future should be agreed together with different organizations taking part in the effort. These results could be tailored or used as inspiration to building other context specific databases.

Individual pilots are not enough to resolve complex conflicts. **An ongoing commitment is needed between actors and within organizations.** Investing resources for the initial building of solutions needs to be backed up by ongoing support for maintenance. Commitment at the level of leadership is needed for securing the continuity. Commitment at the working level is needed for capacity building and changing ways of working to gathering and using data in a way that lends to AI analysis in the future.

A human-centered approach and inclusion of different stakeholders could be supported by AI technology. For future solutions in this opportunity area – for building of peace analysis databases or for applications that utilize it - considerations of data safety, privacy, access, accountability and ethics will be a matter of life and death. These apply to both to the design and use of future applications that by definition would have to be carried out responsibly. Systems-minded considerations and collaboration between experts is needed in order to tackle those concerns in practice and include criteria and lessons learned from both AI and peacebuilding sectors.

The main considerations for this opportunity area are condensed into the peace technology summary canvas that also combines the considerations of the next opportunity area. It should be noted that the canvas is meant for design of context specific solutions and its suitability to be used in presenting a wider opportunity area is limited.

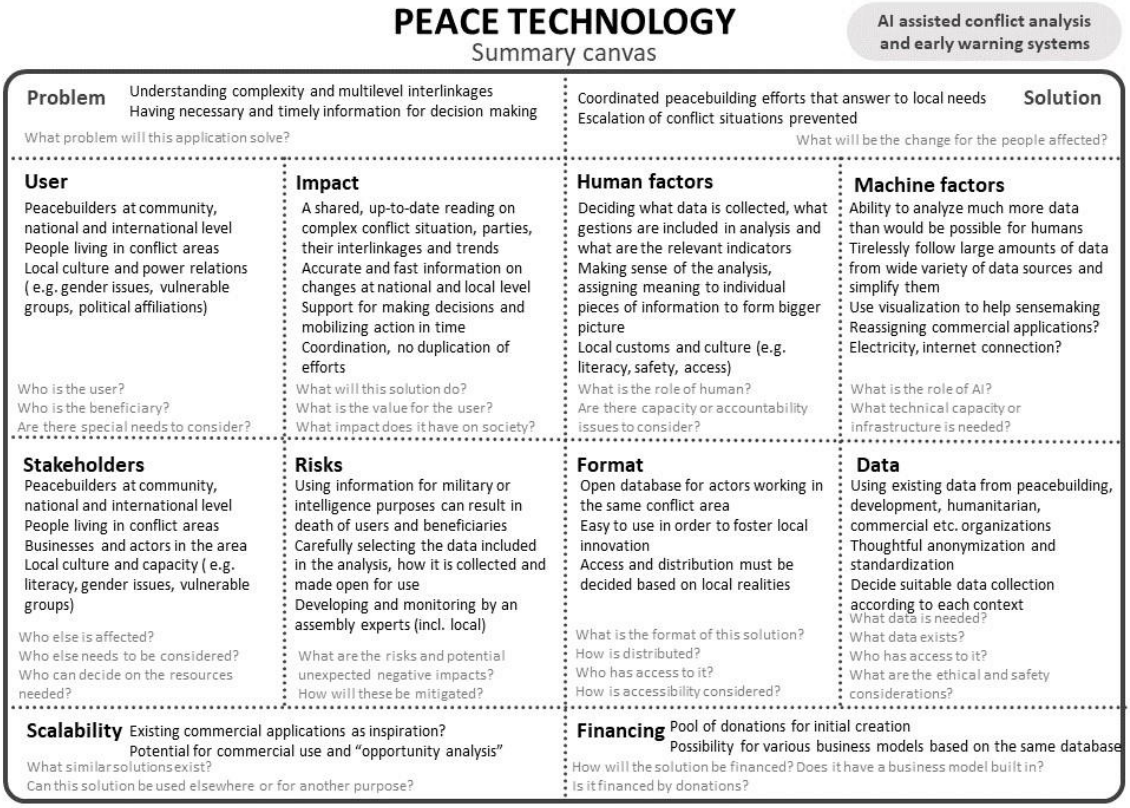


Figure 16: Summary canvas of AI assisted conflict analysis and early warning systems opportunity areas

6.2 Early warning systems mobilize action

The second opportunity area identified could be based to the same assembly of solutions and databases that are required for AI assisted conflict analysis. It could also benefit from the

third opportunity area and include analysis on human communication as an indicator. **Early warning systems** as the second opportunity area is about using different types of data and selecting indicators to identify increase in conflict potential. If the previous opportunity area solutions were to form a collection or an assembly, potential early warning systems could be an integral part of it.

Prevention is considered the most cost-effective and least human suffering alternative to tackle conflicts. (United Nations & World Bank 2018). A functioning early warning system can give time to act to in order to prevent a situation from spiraling into violence. Using AI to help detect the signs for increased conflict potential was viewed as a win-win for reducing human suffering, doing risk analysis that can also benefit private sector actors and for carrying out impact assessment for humanitarian of peace interventions.

A reliable early warning system could potentially turn what is now risk analysis into an opportunity analysis. Equitable distribution of resources and economic development are important building blocks to sustainable peace. Turning risks into opportunities would be a real game changer. Thus, early warning systems could support peace objectives at the stages all stages of conflict but are particularly relevant in preventive measures and in normalization and reconciliation needed to prevent recurrence of conflicts.

For this opportunity area clear definition of responsibilities of human and AI is even more crucial. Building on collective strengths is the only way AI assisted early warning systems could be more than existing individual efforts combined. **Ability to include much wider sets of indicators and data (texts, videos, pictures) to be monitored tirelessly is the strength of AI.** It can help in repetitive tasks that are difficult to humans, such as analyzing material from violent events and quickly identifying patterns and gaps from large datasets. Instant alerts would be crucial in situations where quick action is needed. In conflict situations this means saving lives. As with previous opportunity area, AI analysis carries and can even strengthen the biases from the data. However, it does not look for events that reconfirm existing analysis, as was seen to happen with humans.

The role of human is that of quality and safety assurance in data collection and in using the resulting analysis. Understanding the unique conflict context is important for successful data collection and to making sense of the patterns and indicators detected by AI. Assigning meaning and deciding on actions should be done with the overall big picture in mind. Similarly deciding relevant indicators for escalation cannot be done without understanding the uniqueness of each context. It is likely that a single reliable indicator is found so reading too much into the AI analysis is a risk. As with any technology, early warning systems would only be a tool, the use of which is the responsibility of humans.

Striking a balance or “finding a symbiosis” of which decisions need to be done by humans and which can be automated for AI to make unsupervised is important. With more automation, potential for mistakes increases. More supervision means more resources spent. Thus, the value of using AI for early warning might come into question. It was pointed out that conflict experts are just as likely to know when things are about to “get hot”. Monitoring much larger amounts of data and ability to alert on changes immediately would be the benefits of AI. **Careful consideration of access** is key also for this opportunity area as potential for misuse towards military purposed could be high.

Concentrating on big picture, functioning as institutional memory in situations of high staff turnover and doing wide ranging trend analysis could help people to answer the why-questions important to understanding larger development trajectories. When it is no-ones and everyone’s responsibility to follow the big picture, having the assistance of AI would free up human resources to do other things. Trend analysis was seen as a safer option than including operative level information into the databases. This is yet another one of the same risks that need to be considered all of the opportunity areas.

This study highlights that in early warning systems, as with all AI peace technology solutions, the value that comes from utilizing AI needs to be considered carefully. If the costs or safety risks are greater than the benefits, it is difficult to justify spending limited resources to build AI systems. Possible effectiveness gains should be complemented by benefits that are of strategic importance to achieving peace objectives.

This study finds a common perception that AI is able to detect patterns that humans would miss. Using gaps in data and metadata for more detailed analysis is a clear strength. Assigning the right questions to ask from the data and meaning to the patterns detected remains the task of humans. AI can support decision making. Early warning systems can give enough time and help to mobilize the political will and resources to act. Even with AI **there is not guarantee that the ability to predict conflicts would automatically increase the willingness and actions that prevent them.**

As the main elements for early warning systems as well as the main risks and needs for integration to the peacebuilding ecosystem are same with the first opportunity area, they are summarized in the same peace technology canvas that is presented in picture 12.

6.3 Better human communication for sustainable peace

The third opportunity area identified is **AI support to human communication.** It is perhaps the widest and most elusive, yet closest to the initial inspirational idea of a "peace machine" that helps people to better understand each other. With the commercial applications of technologies like NLP and sentiment analysis developing swiftly, having more practical ways to utilize AI in this area also to support peace objectives might be around the corner.

The technology demonstration that tested NLP and sentiment analysis put the high expectations in this opportunity area into perspective. AI ability to register speech and emotions can help in acquiring new data sets from conflict environments that could then be used for analysis. When little big data exists, finding ways to include "thick data" are important. This opportunity area links and supports the previous two opportunity areas. How to capitalize this into applications that are of strategic importance to peace objectives remains much fuzzier question. The array of local language data sets in conflict areas is additional challenge. Humor, irony, changing nuances in language and the fact that people are not logical all add to it.

Potential future solutions in this opportunity area would support two of the three key functions for technology in peacebuilding as defined by British Council (2018, 4): strategic communications and engaging more people into conversations about peace, and dialogue and mobilization and creating new spaces for a broader variety of peace actors and actions.

However, any AI support to dialogue and communications (e.g. Facebook, Google translate) is not peace technology. The concrete ideas where AI was identified to have strategic importance to peacebuilding objectives were 1) supporting peace negotiations, 2) hate speech detection and building alternative narratives, and 3) micro-interventions or "peace nudging alerts" that would also support the reflection of one's own biases. The last two solution ideas are central at all conflict phases, but prevention and reconciliation in particular would benefit from future applications in this area.

The AI support to peace negotiations is perhaps the most classic example of strategic importance to peace. Solutions in this area would benefit not just individual negotiations but also the learning of future peace negotiators. Reinforcing personal commitment to negotiations while making them more inclusive by consulting constituencies that are normally hard to reach could hold potential for future applications that support sustainable peace. Technology would thus help to bridge the gap of "who gets a seat in the table" where decisions that concern whole societies, countries and regions are made.

The main issues in this opportunity area are summarized in the peace technology canvas presented below. The same limitations on the use of canvas apply as it is initially designed for development of concrete and context specific applications.

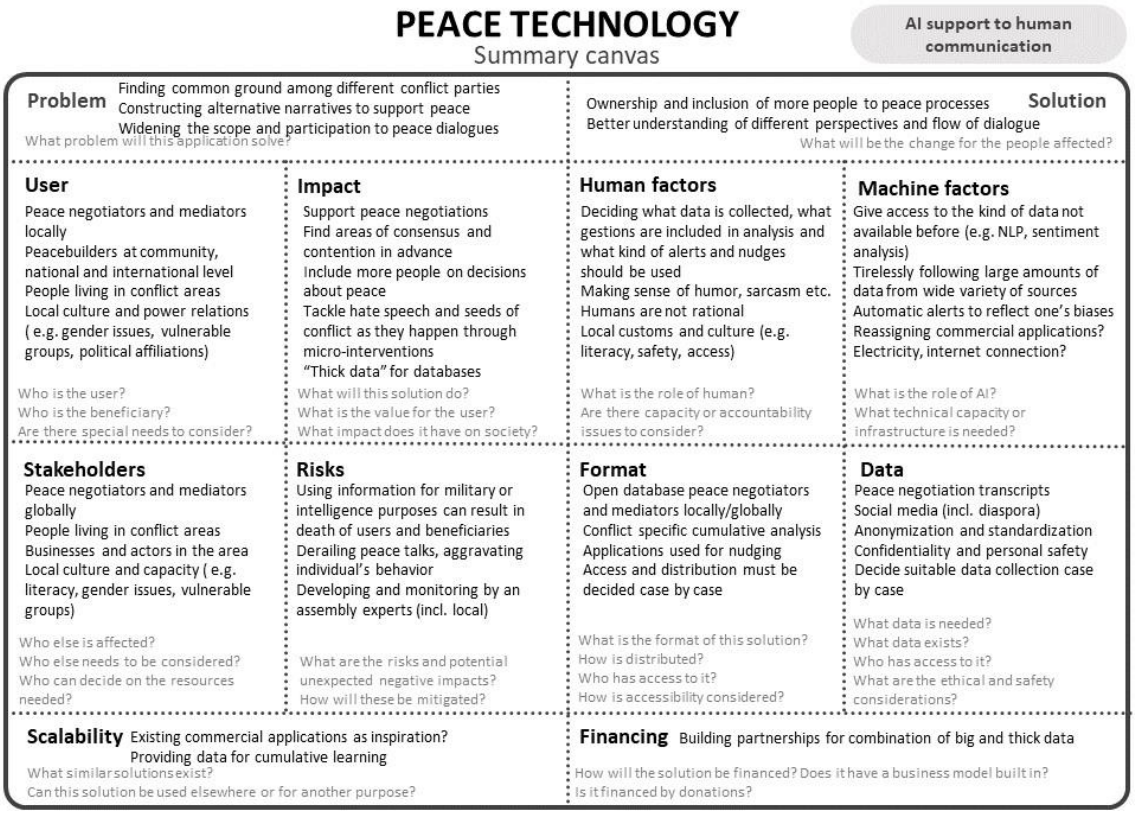


Figure 17: Summary of AI support to human communications opportunity area

6.4 Responsible design of AI peace technology needs concrete guidance

The second development task was to create design tools and considerations for responsible design of AI peace technology based on the insights gathered in a case study setting. Responsibility was treated as a must rather than an option for design of any peace technology. This study tested the usefulness of design-thinking and tools in creating technology applications with the aim of resolving complex social challenges, such as modern-day conflicts. It quickly became clear that existing design tools and methods were not adequate for life and death situations and for processes where inclusion of the end-user and beneficiary was limited.

The canvas set developed aims to provide concrete tools to tackle the complexity involved, to move from cross-silo cooperation approach to collaboration within an ecosystem of experts approach and to push the development process from technology driven towards a collective

intelligence driven design. For responsible design of AI peace technology, human-centered approach needs to be the core, yet it alone falls short. Combination of human and machine capabilities and their respective wider environmental concerns is needed. Similarly considering impact at various levels – individual, community and society – in advance is necessary. The set of canvases are an attempt to operationalize a **human-centered and systems-minded approach** (Both 2018) for the design of AI peace technology.

While not always common, starting from in-depth risk analysis in technology development is not difficult. Making sure innovation is done responsibly in terms of stopping the innovation process if needed holds some challenge. Gathering together different criteria and good practices from two sectors is easy. The bigger question for AI peace technology design, however, is what should be considered in order to move forward. How does that collection of knowhow and good practices translate into concrete ways of responsible design that produces responsible peace technology? This is the meaningful collaboration between communities of experts that holds great promise according to Miklian and Hoelscher (2017).

The study shows that assemblies could be used organize the ecosystem of experts needed. This is not to substitute design with user, but rather help in considering what are the suitable and safe methods to do so. With little to no possibility to do light testing and re-iteration based on failure, two layers of collective sensemaking is needed. First to understand what problems are worth solving in a given context, and secondly what are the potential solutions and their impacts. Third task is to assign methods for each round of sensemaking that secure inclusion of both systemic and human concerns and guiding questions. Substantive understanding of the conflict context, of peacebuilding ecosystem and of good practice principles related to them are as important as is the technical knowledge on AI abilities, limitations and good practice principles related to this sector.

The design tools help to find a common language for the assemblies of experts needed for both the problem definition as well as for the impact sensemaking. The assemblies of experts should be understood in wide sense of the word, to include thematic, areas, community and so on expertise when needed. Similarly, assemblies are likely needed for assigning meaning to the AI analysis and individual pieces of information. No AI application alone solves complex conflicts, nor forms a bigger picture, but rather provide partial solutions and suggestions for humans to collectively make sense and decide on.

The insights gathered in the case study setting were used to create a set of issues that should at least be considered in responsible design of AI peace technology. Dialogical nature of the process allowed the three canvases developed to be tested and re-formulated so that they include the main questions to ask from understanding AI as technology to conflicts as a context

and peacebuilding as an ecosystem. The canvases should be viewed not as an exhaustive list but rather a minimum set of requirements for responsible design of AI peace technology. The feedback received suggests that similar considerations are also relevant for responsible design of AI applications with intended social impact in other sectors.

To realize the potential of AI peace technology, a multidisciplinary approach and building of partnerships between different actors is a must. Practical guidance is needed to move from the initial enthusiasm of meeting experts from different backgrounds (as in the study workshops) towards identification of common interests to build the partnership on (not to mention towards building a functioning AI solution). According to feedback received, the set of canvases helps in this task. Only an ecosystem of experts working together can answer questions like what is relevant data, how can one access it, can the solutions be built and operated in a cost-effective and safe way, do they have the necessary buy in at organizational and political level, what does the analysis produced mean and what are the appropriate actions to take in a given context.

This thesis demonstrates **that design can play a role in combining substance and technical expertise to create context specific innovations that help to solve complex social challenges.** In the case of AI peace technology, design needs to be more than a common language. It should also function as a gatekeeper to responsibility both in developing the applications and in securing their responsible use. For this purpose, the set of design tools developed integrate the most central elements of responsible innovation according to Stilgoe et al. (2013):

1) **anticipation** (foreseeing risks and dangers) that is done in the form of separate impact canvas
2) **reflexivity** (awareness of one's own role, values and responsibilities) that is done in the form of considering the differences intended end-user and beneficiary and their respective environment in the ecosystem solution mapping canvas that places the technology into its wider context
3) **inclusiveness** (diversity of opinions and public legitimization) done through formulating the design tools to guide dialogue in assemblies of experts and though making their development process and future use open for everyone
4) **responsiveness** (making sure principles, norms and standards have an impact on the way the innovation process is done) that is done in the form of including concerns of data quality, AI ethics and human considerations relevant peacebuilding into the peace technology summary canvas and using the canvas set to prompt two levels of understanding and sensemaking among communities of experts

A fifth element of responsibility - **context specificity** - is added to the list above. Responsible design of AI peace technology can only be done with an in-depth understanding of each unique conflict context and of the different people in it. This is the only way to build AI peace technology in a conflict sensitive way, making sure it does not worsen the conflict. However, for peace technology, **conflict sensitivity is a point of departure**. Do no harm as a principle needs to be followed at all times. But it is not sufficient as a goal for solutions aiming to do good and build peace.

Understanding the context includes the power relations, culture and multilevel ethical issues related to it. They need to be at the core of the design process. The set of canvases responds to Miklian and Hoelscher (2017, 7) search for concrete tools to consider the wider impacts in a given context as opposed to the technology issues dominating the developing of peace innovations. **Ethics and human safety are the very first and the very last issues to consider in responsible peace technology design.** The canvases developed are useful for an early stage idea testing and for mid-process decision at a proof of concept phase. They are there to complement the existing design tools that should be used to cover other phases of design.

Decisions on scalability and resourcing do not happen in a vacuum, nor can they be done by individual actor for AI peace technology. The tools developed should help in making the decisions on whether it is worth using limited resources to build AI solutions. This is one small step towards supporting peacebuilding and AI actors in forming the coalitions needed to create the ecosystems and to establish conflict specific databases**. If these pools of data are made open, they could foster local innovations and creativity.** This could in turn encourage developing applications that would not happen due resources needed for the initial investment of gathering, anonymizing and standardizing the data needed. There is also potential in **AI peace technology to boost local level economic development and innovation in conflict areas.**

As funding in peacebuilding tends to be attached to projects and programs, long term commitment needed for building and maintenance of AI peace technology is a challenge. Necessity to demonstrate impact not just on lives of the people affected, but also to regional or global security, is not uncommon. In real life peace technology applications need to function in organizations where there is little experience on continuous development or culture of agile re-iteration that is needed. The question to tackle in responsible design is thus twofold: a human perspective one on what kind of interventions are created and a systems-level one on how they are integrated into the overall ecosystem (Both 2018). This is why both systemic and human-centered considerations need to be part of the responsible design of AI solutions not just in peace technology, but also in other sectors.

The thesis suggests that responsible design of AI peace technology is about design of collective intelligence. For this, a human-centered and systems-minded design approach to collective intelligence is required. Organizing collective intelligence through assemblies is rare and needs skillful design (Mulgan 2018, 218-221). Assemblies are a way organize the multidisciplinary collaboration in needed for AI pace technology to function into ecosystem of experts. AI can help in all three key functions of technology in peacebuilding, that according to British Council (2018, 4) are data management, strategic communications and dialogue and mobilization. When viewed as collective intelligence, which AI peace technology should always be, the three strategically important functions for peace are: 1) support to human capacity in processing immense amounts of data and through that 2) augment human ability to understand complexity and interlinkages, and 3) suggest alternative interpretations and avenues for action for people to decide on.

6.5 Evaluation of work and prospects for future research

With much attention on how AI will affect our lives, much less interest focusses on how we could use the rapid development of AI application to tackle some of the most urgent challenges of our time. The landscape of increasing complexity and number of conflicts is among the most pressing ones. This thesis is a limited case study sample that looks at a real-life super wicked problem of fragmented, interlinked and fast-changing conflicts in today's world. From an inspirational idea of a "peace machine" the scope of this thesis was to identify areas where AI could support peace and how the applications in those areas could be designed responsibly.

The thesis manages to answer both of its research questions. Firstly, by identifying three separate, but interlinked opportunity areas where the capacity of AI and needs of peacebuilding overlap. Secondly, by using the insight collected to formulate a set of design tools and considerations for responsible design of AI peace technology. While the development work has been in process, the questions around how to use AI to help us tackle complex social challenges (such as pandemics) have become even more pertinent. Along with this, the issues of responsible design and ethics of AI are equally relevant.

The study relies on multiple sources of evidence and converges the different types of information through an ongoing process triangulation and analysis. This was a useful approach in gathering a holistic understanding on a research topic covers many different fields of expertise. Having an in-depth understanding of not just one, but various real-life phenomena and their contexts was challenging. However, combining the data points and information from different sources led to saturation of material and formulation of results based on dialogical research process and an interplay between the data collection and the emerging findings.

The validity of results was tested in two separate workshops. The feedback from participants and from the Peace Machine project leads at Futurice was positive. The issues identified were seen as relevant and additional to the existing knowledge on AI peace technology and its responsible design. Furthermore, Futurice indicated that they considered that the project had achieved its objectives and that all the relevant stakeholders were included to the workshops that form a major part of the study material. Collective intelligence of groups of people, of technology through the demonstration and of individual experts through the interviews and probes were used for the development work. Considering the responsible design, confidentiality issues and ethical questions also benefitted greatly from the discussions with Futurice project leads.

Designers are often characterized to like challenges. They crave real-life problems to build solutions that can change people's lives, relish opportunities to come up with new solutions and develop novel approaches. This was echoed by the enthusiasm and high energy that was present in all the workshops organized. Using the potential of AI to solve real-life wicked problems, like conflicts, is a challenge worthy for the designer community to take on. Merging experiences and expertise of people from different backgrounds is exciting but needs practical guidance in order to create actual AI solutions. The design tools are a way to do that.

The development work demonstrates that there is potential in using AI for building future applications of peace technology that can be of strategic importance to peacebuilding, even in complex environments of conflict "squiggles". Bringing experts and expertise together is not enough to responsibly develop and use AI peace technology. Design plays a vital role and for that to be successful, concrete tools and methods are needed. Considerations relevant to responsible design of AI peace technology can be applied to other sectors. However, as the tools developed are only tested with peace technology in mind, the need for sector specific tailoring remains unexplored.

All the expert fields of peacebuilding, AI, technology development, responsible design and innovation have their own discussions, good practices, principles and norms. A more comprehensive summary of the most relevant existing criteria and good practices from these sectors would be a good addition to responsible design and the tools developed. This would particularly help in situations where substance or technical experts are not present. The set of canvases functions as a minimum checklist and through that hopefully raise awareness of brining AI experts into peacebuilding discussions and vice versa. Condensing normative and practical frameworks from AI and peacebuilding into a workable, easy to understand guide would be an interesting prospect for future research.

One gap in the development work is establishing business cases for future AI peace technology solutions. While the area is discussed in terms of raising questions on financing, the workshops had little time to explore the issues of scalability and funding. Creating ecosystems and databases around specific conflict areas is a step towards that goal. It could help to foster local innovations and create business models that are not yet imaginable. Rather than setting up individual pilots, big international peacebuilding actors could play a strategic role in forming multi-expertise alliances and networks and maintaining the collaboration needed. They could form an ecosystem that nurtures new ideas on how to combine the superpowers of technology and peacebuilding. This might be worth exploring.

References

Blok, V. 2014. Look who's talking: Responsible innovation, the paradox of dialogue and the voice of the other in communication and negotiation processes. Journal of Responsible Innovation, 1, 2, 171-190.

Boyd, R. & Holton, R. J. 2018. Technology, innovation, employment and power: Does robotics and artificial intelligence really mean social transformation? Journal of Sociology, 54, 3, 331–345.

Brynjolfsson, E. 2014. The second machine age: Work, progress, and prosperity in a time of brilliant technologies. New York, United States: W. W. Norton.

Corbett-Davies, S. & Pierson, E. & Feller, A. & Goel, S. & Huq, A. 2017. Algorithmic decision making and the cost of fairness. Ithaca, United States: Cornell University.

Corbin, J. & Strauss, A. 2008. Basics of qualitative research: Techniques and procedures for developing grounded theory, 3rd edition. Los Angeles, United States: Sage.

De Hoop, E. & Pols, A. & Romijn, H. 2016. Limits to responsible innovation, Journal of Responsible Innovation, 3, 2, 110-134.

Doty, A. 2016. Bringing Peace to Life? A Narrative Analysis of Finnish Development Intervention in Conflict-Affected Nepal. Acta Universitatis Tamperensis 2209. Tampere, Finland: Tampere University Press.

Elia, G. & Margherita, A. 2018. Can we solve wicked problems? A conceptual framework and a collective intelligence system to support problem analysis and solution design for complex social issues. Technological Forecasting & Social Change, 133, 279-286.

Eriksson, P. & Koistinen, K. 2014. Monenlainen tapaustutkimus. Helsinki, Finland: Kuluttajatutkimuskeskus.

Eronen, O. 2018. Rauhanvälitys murroksessa: miten tulemme toimeen jatkuvan muutoksen kanssa? In Siirtola, R. & Palm, A. & Yli-Vakkuri, L. (eds.) Yhdessä enemmän: Kriisien hallintaa kokonaisvaltaisesti. Helsinki, Finland: Laajan turvallisuuden verkosto WISE ry. 123-131.

Galtung, J. 1969. Violence, Peace, and Peace Research. In Journal of Peace Research, Vol. 6, 3, 167-191.

Galtung, J. 1976. Three Approaches to Peace: Peacekeeping, Peacemaking, and Peacebuilding. In Peace, War and Defense: Essays in Peace Research. Vol. II. Copenhagen, Denmark: Ejlers. 297-298.

Ghauri, P. & Grønhaug, K. 2005. Research methods in business studies: A practical guide. 3rd ed. Harlow, United Kingdom: Pearson Education.

Gould, J. & Lewis, C. 1985. Designing for usability: Key principles and what designers think. Communications of the ACM 28, 3, 300–311.

Honkela, T. & Kaaro, J. 2017. Rauhankone: Tekoälytutkijan testamentti. Helsinki, Finland: Gaudeamus.

Kahl, A. & Puig Larrauri, H. 2013. Technology for Peacebuilding. Stability: International Journal of Security & Development, 2, 3:61, 1-15.

Kananen, J. 2013. Case-tutkimus opinnäytetyönä. Jyväskylä, Finland: Jyväskylän ammattikorkeakoulu.

Kaplan, A. & Haenlein, M. 2019. Siri, Siri, in my hand: Who's the fairest in the land? On the interpretations, illustrations, and implications of artificial intelligence. Business Horizons, 62, 15-25.

Kelly, K. 2016. The inevitable: Understanding the 12 technological forces that will shape our future. New York, United States: Viking.

Laaksonen, S. & Haapoja, J. & Kinnunen, T. & Nelimarkka, M. & Pöyhtäri, R. 2020. The Datafication of Hate: Expectations and Challenges in Automated Hate Speech Monitoring. Frontiers in Big Data, 3.

Laine, M. & Bamberg, J. & Jokinen, P. 2007. Tapaustutkimuksen taito. Helsinki, Finland: Gaudeamus.

Letouzé, E. & Meier, P. & Vinck, P. 2013. Big Data for Conflict Prevention: New Oil and Old Fires. In Mancini, F. (ed.). New Technology and the Prevention of Violence and Conflict. New York, United States: International Peace Institute. 4-28.

Lévy, P. 2005. Collective intelligence, A civilisation: Towards a method of positive interpretation. International Journal of Politics, Culture, and Society, 18, 3-4, 189-198.

Liesinen, K. 2018. Kokonaisvaltaisen kriisinhallinnan ihanne ja välineet. In Siirtola, R. & Palm, A. & Yli-Vakkuri, L. (eds.) Yhdessä enemmän: Kriisien hallintaa kokonaisvaltaisesti. Helsinki, Finland: Laajan turvallisuuden verkosto WISE ry.19-33.

Macnaghten, P. & Owen, R. & Stilgoe, J. & Wynne, B. & Azevedo, A. & de Campos, A. & Chilvers, J. & Dagnino, R. & di Giulio, G. & Frow, E. & Garvey, B. & Groves, C. & Hartley, S. & Knobel, M. & Kobayashi, E. & Lehtonen, M.& Lezaun, J. & Mello, L.& Monteiro, M.& Pamplona da Costa, J. 2014. Responsible innovation across borders: tensions, paradoxes and possibilities. Journal of Responsible Innovation, 1, 2, 191-199.

Malone, T. & Laubacher, R. & Dellarocas, C. 2010. The collective intelligence genome. MIT Sloan Management Review, 51, 3, 21–31.

Marmar, C.R. & Brown, A.D. & Qian, M. & Laska, E. & Siegel, C. & Li, M. & Abu-Amara, D. & Tsiartas, A. & Richey, C. & Smith, J. & Knoth, B. & Vergyri, D. 2019. Speech-based markers for posttraumatic stress disorder in US veterans. Focus on: The Impact of Trauma on Mental Health: PTSD and Beyond. Depression and Anxiety, 36, 7, 607-616

Mattelmäki,T. 2006. Muotoiluluotaimet. Helsinki, Finland: Teknologiateollisuus ry.

Miklian, J. & Hoelscher, K. 2016. A Blueprint for Pro-Peace Innovation. PRIO Policy Brief 25. Oslo, Norway: Peace Research Institute Oslo.

Miklian, J. & Hoelscher, K. 2017. A new research approach for Peace. Innovation and Development Journal, 8, 2, 189-207.

Montes, G. & Goertzel, B. 2019. Distributed, decentralized, and democratized artificial intelligence. Technological Forecasting & Social Change, 141, 354-358.

Mootee, I. 2013. Design thinking for strategic innovation. What they can't teach you at business or design schools. Hoboken, United States: John Wiley & Sons Inc.

Mueller, H. 2017. "How Much Is Prevention Worth?" Background paper for the United Nations-World Bank Flagship Study Pathways for Peace: Inclusive Approaches to Preventing Violent Conflict. Washington DC, United States: World Bank.

Mulgan, G. 2006. The Process of Social Innovation. Innovations: Technology, Governance, Globalization, 1, 2, 145–162.

Mulgan, G. 2018. Big mind: How collective intelligence can change our world. Princeton, United States: Princeton University Press.

Obermeyer, Z. & Powers, B. & Vogeli, C. & Mullainathan, S. 2019. Dissecting racial bias in an algorithm used to manage the health of populations. Science, 366, 6464, 447-453.

OECD 2015. States of Fragility 2015: Meeting the Post-2015 Ambitions. Paris, France: OECD.

Ojasalo, K. & Moilanen, T. & Ritalahti, J. 2009. Kehittämistyön menetelmät: Uudenlaista osaamista liiketoimintaan. Helsinki, Finland: WSOY.

Owen, R. & Stilgoe, J. & Macnaghten, P. & Gorman, M. & Fisher, E. & Guston, D. & Bessant, J. 2013. A framework for responsible innovation. In Owen, R. & Bessant, J. & Heitz. M. (eds.) Responsible innovation: Managing the responsible emergence of science and innovation in society. Chichester, United Kingdom: John Wiley & Sons. 27–50.

Peters, B. G. 2017: What is so wicked about wicked problems? A conceptual analysis and a research program. Policy and Society: Understanding policy problems: Tame and wicked problems reconsidered, 36, 3, 385-396.

Pettersson, T. & Högbladh, S. & Öberg, M. 2019. Organized violence, 1989–2018 and peace agreements. Journal of Peace Research, 56, 4, 589–603.

Puig Larrauri, H. & Jung, Y. 2017. Reimagining Peacebuilding Through Innovation. Session Report no. 14. Stockholm Forum on Peace and Development. Stockholm, Sweden: SIPRI.

Quihuis, M. & Nelson, M. & Guttieri, K. 2015. Peace Technology: Scope, Scale, and Cautions. Building Peace, 5, 14-16.

Ramsbotham, O. & Woodhouse, T. & Miall, H. 2016. Contemporary Conflict Resolution: The Prevention, Management and Transformation of Deadly Conflicts. Fourth edition, fully revised and updated. Cambridge, United Kingdom: Polity Press.

Rittel, H. & Webber, M. 1973. Dilemmas in the general theory of planning. Policy Sciences, 4, 155–169.

Ross, A. 2016. The industries of the future. New York, United States: Simon & Schuster.

Russell, S. & Norvig, P. 2016. Artificial intelligence: A modern approach. Third edition. Global edition. Harlow, United Kingdom: Pearson Education Limited.

Schlaile, M. & Mueller, M. & Schramm, M. & Pyka, A. 2018. Evolutionary Economics, Responsible Innovation and Demand: Making a Case for the Role of Consumers. Philosophy of Management, 17, 1, 7–39.

Schwab, K. 2016. The fourth industrial revolution. Geneva, Switzerland: World Economic Forum.

Shank, G.D. 2002. Qualitative Research: A personal skills approach. Upper Saddle River, United States: Pearson Merrill Prentice Hall.

Skilton, M. & Hovsepian, F. 2018. The 4th industrial revolution: Responding to the impact of artificial intelligence on business. Cham, Switzerland: Palgrave Macmillan.

Sphere Association 2018. The Sphere Handbook. Humanitarian Charter and Minimum Standards in Humanitarian Response. Fourth Edition. Geneva, Switzerland: Shortrun Press.

Stickdorn, M. & Lawrence, A. & Hormess, M. E. & Schneider, J. 2018. This is service design doing: Applying service design thinking in the real world: a practitioner's handbook. First Edition. Sebastopol, United States: O'Reilly Media Inc.

Stilgoe, J. & Owen, R. & Macnaghten, P. 2013. Developing a framework for responsible innovation. Research Policy, 42, 9, 1568-1580.

United Nations 2008. Peacekeeping Operations. Principles and Guidelines. Department of Peacekeeping Operations. Department of Field Support. New York, United States: United Nations Secretariat.

United Nations 2010. UN Peacebuilding: An Orientation. United Nations Peacebuilding Support Office. New York, United States: United Nations.

United Nations 2018a. UN Secretary-General's Strategy on New Technologies. September 2018. New York, United States: United Nations.

United Nations 2018b. The World Economic and Social Survey. Frontier Technologies for Sustainable Development. United Nations Department of Economic and Social Affairs. Economic Analysis and Policy Division. New York, United States: United Nations.

United Nations & World Bank 2018. Pathways for Peace: Inclusive Approaches to Preventing Violent Conflict. Washington DC, United States: World Bank.

UNHCR 2019. Global Trends. Forced Displacement in 2018. United Nations High Commissioner for Refugees. Geneva, Switzerland: UNHCR.

Von Schomberg, R. 2011. Towards Responsible Research and Innovation in the Information and Communication Technologies and Security Technologies Fields. European Commission. Research and Innovation Policy. Luxembourg: Publications Office of the European Union.

Yin, R. 2014. Case study research: Design and methods. 5th ed. Los Angeles, United States: SAGE.

Yin, R. 2018. Case study research and applications: Design and methods. Sixth edition. Los Angeles, United States: SAGE.

Electronic Sources

AI Finland 2018. AI Forum 2018. Accessed 10.4.2020. https://www.tekoalyaika.fi/en/ai-forum-2018/

Alliance for Peacebuilding 2013. Selected Definitions of Peacebuilding. August 2013. Accessed 20.4.2020. http://www.allianceforpeacebuilding.org/2013/08/selected-definitions-of-peacebuilding/

Angwin, J. & Larson, J. & Mattu, S. & Kirchner, L. 2016. Machine Bias. ProPublica. Accessed 10.4.2020. https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

Berditchevskaia, A. & Baeck, P. 2020. The Future of Minds and Machines: How artificial intelligence can enhance collective intelligence. Nesta. Accessed 24.5.2020. https://www.nesta.org.uk/report/future-minds-and-machines/

Both, T. 2018. Human-Centered, Systems-Minded Design. Stanford Social Innovation Review. Accessed 24.5.2020. https://ssir.org/articles/entry/human_centered_systems_minded_design#

British Council 2016. Build Up. Innovative Peacebuilding in Syria. A Scoping Study of the Strategic Use of Technology to Build Peace in the Syrian Context. Accessed 24.5.2020. http://creativeconomy.britishcouncil.org/media/uploads/files/Peacetech_Report_Web_En.pdf

British Council 2018. Innovative peacebuilding in Syria II. An update on the strategic use of technology to build peace in the Syrian context. Accessed 10.4.2020. https://howtobuildup.org/wp-content/uploads/2016/04/INNOVATIVE-PEACEBUILDING-IN-SYRIA-II.pdf

Chui, M. & Harryson, M. & Manyika, J. & Roberts, R. & Chung, R. & van Heteren, A. & Nel, P. 2018. Notes from the AI Frontier. Applying AI for Social Good. McKinsey Global Institute Discussion Paper. Accessed 11.4.2019. https://www.mckinsey.com/featured-insights/artificial-intelligence/applying-artificial-intelligence-for-social-good

Copeland, B.J. 2018. Artificial Intelligence. Encyclopedia Britannica 2018. Accessed 20.6.2019. https://www.britannica.com/technology/artificial-intelligence#ref219078

Cottray, O. & Puig Larrauri, H. 2017. Technology at the service of peace. SIPRI Guest Blog. Stockholm International Peace Research Institute. Accessed 10.4.2020. https://www.sipri.org/commentary/blog/2017/technology-service-peace

Council of Europe 2019. Governing the Game Changer – Impacts of artificial intelligence development on human rights, democracy and the rule of law. Conclusions from the Conference. Accessed 10.11.2019. https://rm.coe.int/conclusions-from-the-conference/168093368c

Dahl Jensen, C. & Amnebjer, S. 2019. #Peacetech. The Role of Tech in Creating or Sustaining Peace. Danmission. Accessed 25.5.2020. https://issuu.com/danmission/docs/peacetech_report_-_danmission?e=7378243%2F69638873

Dastin, J. 2018. Amazon scraps secret AI recruiting tool that showed bias against women. Reuters. Accessed 10.4.2020. https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G

Dutton, T. 2018. Artificial Intelligence Strategies. Politics + AI. Accessed 10.11.2019. https://medium.com/politics-ai/an-overview-of-national-ai-strategies-2a70ec6edfd

Elements of AI 2019. University of Helsinki & Reaktor. Accessed 12.10.2019. https://course.elementsofai.com/1/1

European Commission 2020. On Artificial Intelligence - A European approach to excellence and trust. White Paper. Accessed 24.5.2020. https://ec.europa.eu/info/publications/white-paper-artificial-intelligence-european-approach-excellence-and-trust_en

European External Action Services & European Commission 2015. Guidance Note on the Use of Conflict Analysis in Support of EU External Action. Accessed 12.10.2019. https://europa.eu/capacity4dev/public-fragility/minisite/support-and-guidance/guidance-note-use-conflict-analysis-support-eu-external-action

Fabian, Chris & Fabricant, Robert 2014. The Ethics of Innovation. Stanford Social Innovation Review. Accessed 10.3.2019. https://ssir.org/articles/entry/the_ethics_of_innovation

Futurice 2018. Accessed 10.4.2020. https://vimeo.com/futurice

Ground Truth Global 2020. Accessed 10.4.2020. https://www.groundtruthglobal.com/

Hadhazy, A. 2017. Biased Bots. Artificial-Intelligence Systems Echo Human Prejudices. Princeton University. Accessed 10.4.2020. https://www.princeton.edu/news/2017/04/18/biased-bots-artificial-intelligence-systems-echo-human-prejudices

Harvard Humanitarian Initiative 2010. The Signal Code: Human Rights Approach to Information During Crises. Accessed 10.4.2020. https://signalcode.org/

IBM 2018. IBM's Principles for Trust and Transparency. Accessed 10.3.2020. https://www.ibm.com/blogs/policy/trust-principles/

Independent High-Level Expert Group on Artificial Intelligence 2019a. A Definition of AI: Main Capabilities and Disciplines. Definition Developed for the Purpose of the AI HLEG's Deliverables. European Commission. Accessed 15.3.2020. https://ec.europa.eu/digital-single-market/en/news/definition-artificial-intelligence-main-capabilities-and-scientific-disciplines

Independent High-Level Expert Group on Artificial Intelligence 2019b. Ethics guidelines for trustworthy AI. European Commission. Accessed 10.4.2020. https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai

International Committee of the Red Cross 2018: Professional Standards for Protection Work. Accessed 10.4.2020. https://www.icrc.org/en/publication/0999-professional-standards-protection-work-carried-out-humanitarian-and-human-rights

Iskander, N. 2018. Design Thinking Is Fundamentally Conservative and Preserves the Status Quo. Harward Business Review. Accessed 11.4.2019. https://hbr.org/2018/09/design-thinking-is-fundamentally-conservative-and-preserves-the-status-quo

Johnson, S. 2018. A basis for an ethical AI framework for humanitarian response. Accessed 10.4.2020. https://medium.com/@Simon_B_Johnson/a-basis-of-an-ethical-ai-framework-for-humanitarian-response-bc2938b99f80

JustPeace Lab 2017. Ethical Guidelines for Peace Tech. Accessed 10.4.202. https://justpeacelabs.org/blog/2017-03-27/released-ethical-guidelines-peacetech/

Lee, N. & Resnick, P. & Barton, G. 2019. Algorithmic bias detection and mitigation: Best practices and policies to reduce consumer harms. The Brookings Institution Center for Technology Innovation. Accessed 11.4.2020. https://www.brookings.edu/research/algorithmic-bias-detection-and-mitigation-best-practices-and-policies-to-reduce-consumer-harms/

Marr, B. 2018. 5 Important Artificial Intelligence Predictions (For 2019) Everyone Should Read. Forbes. Accessed 11.4.2020. https://www.forbes.com/sites/bernardmarr/2018/12/03/5-important-artificial-intelligence-predictions-for-2019-everyone-should-read/

McCarthy, J. & Minsky, M. & Rochester, N. & Shannon, C. E. 1955. A proposal for the Dartmouth summer research project on artificial intelligence. Stanford University. Accessed 15.3.2020. http://www-formal.stanford.edu/jmc/history/dartmouth/dartmouth.html

Natalucci, M. 2018: Finland makes bold bid to be at forefront of European AI leadership. Timeline. Accessed 10.11.2019. https://www.governmentcomputing.com/brexit-eu/comment/finland-makes-bold-bid-forefront-european-ai-leadership

OECD 2019. OECD Principles on AI. Accessed 10.4.2020. https://www.oecd.org/going-digital/ai/principles/

Peace Tech Lab. 2020. Peace Tech Lab's Hate Speech Lexicons. Accessed 10.4.2020. https://www.peacetechlab.org/toolbox-lexicons

Principles for Digital Development 2015. Accessed 10.4.2020. https://digitalprinciples.org/

Puig Larrauri, H. 2018. Is technology a tool, or is it tooling us? Opening remarks for the Build Peace 2018 Conference. Accessed 10.4.2020. https://medium.com/@howtobuildup/is-technology-a-tool-or-is-it-tooling-us-opening-remarks-for-the-build-peace-2018-conference-d4f2c0caf4e2

PwC 2017a. Sizing the Price. What's the real value of AI for your business and how can you capitalise? Accessed 10.4.2020. https://www.pwc.com/gx/en/issues/data-and-analytics/publications/artificial-intelligence-study.html

PwC 2017b. Responsible AI Toolkit. Accessed 10.4.2020. https://www.pwc.com/gx/en/issues/data-and-analytics/artificial-intelligence/what-is-responsible-ai.html

SAS 2018. Artificial Intelligence. What is it and why it matters. Accessed 12.12.2018. https://www.sas.com/en_us/insights/analytics/what-is-artificial-intelligence.html

The Redirect Method 2020. Blueprint for Bypassing Extremism. Accessed 10.4.2020. https://redirectmethod.org/

Työ- ja Elinkeinoministeriö 2017. Key issues paper for discussion on new approaches to digitalization and artificial intelligence. Accessed 20.6.2019. https://tem.fi/documents/1410877/5926493/Key+issues+digi+non+paper+Finland.pdf/00f03d80-2fe7-486b-9b42-156c5780580d

Understanding Peace 2018. Chilicorn Fund project website. Futurice. Accessed 10.11.2019. www.understandingpeace.net

UNICEF 2018. Industry Toolkit Children's Online Privacy and Freedom of Expression. Accessed 11.8.2019. https://www.unicef.org/csr/files/UNICEF_Childrens_Online_Privacy_and_Freedom_of_Expression(1).pdf

United Nations 2016. Conflict Analysis Practice Note. Accessed 10.4.2020. https://undg.org/wp-content/uploads/2016/10/Conflict-Analysis-Practice-Note-13-May-2016-Version.pdf

UNOCHA 2020. Global Humanitarian Overview 2020. Monthly Funding Update January 31st, 2020. Accessed 11.4.2020. https://reliefweb.int/report/world/global-humanitarian-overview-2020-monthly-funding-update-31-january-2020-enfrar

USAID 2018. Reflecting the Past, Shaping the Future: Making AI Work for International Development. Accessed 11.10.2019. https://www.usaid.gov/digital-development/machine-learning/AI-ML-in-development

Vuolteenaho, S. 2018. "Voi hyvänen aika", huudahtaa virkamies – Tekoäly löysi 280 tekijää, jotka ennakoivat lastensuojelun asiakkuutta. YLE. Accessed 17.2.2020. https://yle.fi/uutiset/3-10413353

Wilson, H. & Daugherty, P. 2018. Collaborative Intelligence: Humans and AI Are Joining Forces. Harward Business Review. Accessed 11.4.2020. https://hbr.org/2018/07/collaborative-intelligence-humans-and-ai-are-joining-forces

World Economic Forum 2016. The Future of Jobs. Report to Annual Meeting. Accessed 21.1.2019. www.weform/reports/the_future_of_jobs

Figures

Tables

Appendices

Appendix 1 Human + Machine co-operation canvas. First iteration of summary canvas

Human + Machine co-operation

The problem/need

What is the problem that needs to be solved?

Description

Key Characteristics

Who is affected?

What does the situation mean for elderly / disabled / youth?
Who are the people that can take decisions on this matter?
Who do these decision makers listen to?

Human factors

What special needs can affect the situation?

Goal

How can human machine cooperation help the situation?
How does it concretely improve the lives of people affected?

Key target group

Who benefits from the solution? What other groups might be affected?

Potential solutions

What do machines do better than humans? How can humans and machines complement eachother?

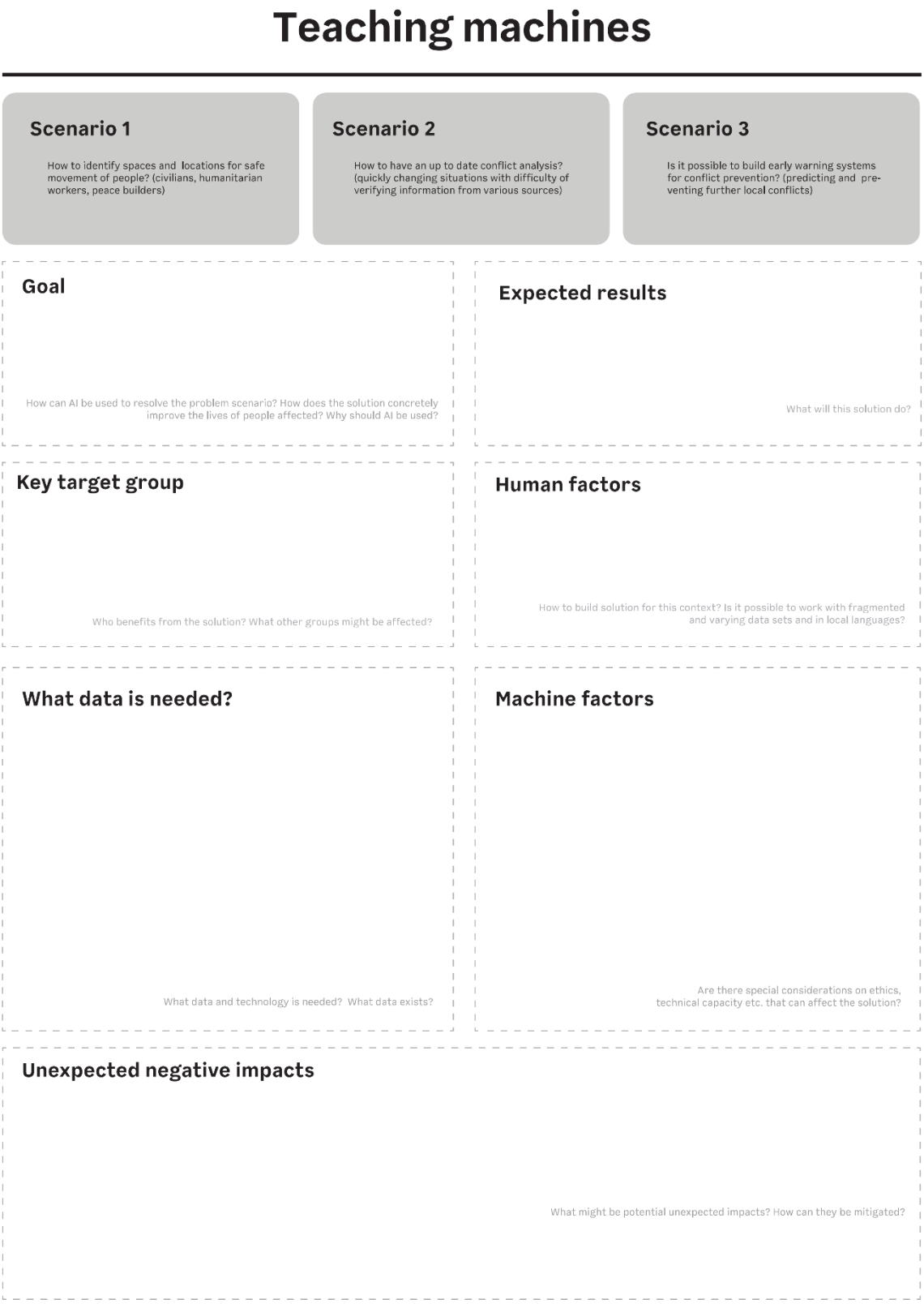
Expected results

What are the concrete and immediate results?

Unexpected results

What might be potential unexpected impacts? How can they be mitigated?

Appendix 2 Teaching machines summary canvas. Second iteration of summary canvas



Appendix 3 List of interviews

1. Three AI experts, private sector 18.4.2018
2. Peace and conflict expert, think tank 24.4.2018
3. Peace and conflict expert, public sector 26.4.2018
4. Peace and conflict expert, private sector 2.5.2018
5. Peace and conflict expert, civil society 7.5.2018
6. Peace and conflict expert, civil society 7.5.2018
7. Peace and conflict expert, civil society 8.5.2018
8. Peace and conflict expert, civil society 8.5.2018
9. Peace and conflict expert, civil society 11.5.2018
10. AI and technology expert, public sector 22.5.2018
11. AI and technology expert, think tank 5.6.2018
12. AI and technology expert, private sector 27.6.2018
13. AI and technology expert, academia 3.7.2018

Appendix 4 The interview question framework

Haastattelun teemat: tekoälysovellukset rauhan rakentamisessa

1. Mitä tällä hetkellä tehdään tekoälyn, nousevien teknologioiden, innovaatioiden tai digitalisaation saralla teidän organisaatiossanne?
2. Mitkä ovat teidän työssänne tai näkökulmastanne suurimmat ja ajankohtaisimmat ongelmat ja kysymykset?
3. Ketkä ovat aktiivisia toimijoita uuden teknologian ja tekoälyn sovellusten kehittämisessä?
4. Ketkä nostavat esiin positiivisia puolia? Entä huolenaiheita?
5. Mitkä sidosryhmät tai toimijat ei ole mukana ja miksi?
6. Mitkä kysymykset jäävät katveeseen? Mihin teemoihin tai kysymyksiin tulisi mielestäsi panostaa lisää?
7. Missä kysymyksissä tiedät, että omalla allasi käytetään jo hyväksi tai harkitaan tekoälyn hyödyntämistä?
8. Onko jotain ratkaisuvaihtoehtoja tai sovelluksia jo testattavana? Mitkä näistä vaikuttavat mielestäsi lupaavimmalle?
9. Mitkä ovat mielestäsi polttavimmat ongelmat ja konkreettiset haasteet, joihin teknologian ja keinoälyn kehitys voisi tuoda uusia ratkaisuja?
10. Missä teemoissa itse näet eniten potentiaalia tekoälysovelluksille tulevaisuudessa?
11. Mitkä ovat mielestäsi suurimmat huolenaiheet/rajoitteet/eettiset kysymykset keinoälysovellusten hyödyntämisessä omalla sektorillasi / yhteiskunnallisten ongelmien ratkaisemiseksi?
12. Mitä en ole vielä kysynyt?

Appendix 5 Instructions to the design probe participants

KIITOS OSALLISTUMISESTASI LUOTAINTUTKIMUKSEEN!

Tämän luotaintutkimuksen avulla kartoitetaan, minkälaisia haasteita rauhan rakentamisen ja kriisinhallinnan kentällä ja/tai hauraissa tilanteissa työskentelevät kohtaavat arkipäivässään. Tutkimus on osa palvelumuotoilun lopputyötä, jonka tavoitteena on selvittää keinoälyn ja teknologian mahdollisuuksia vastata tunnistettuihin haasteisiin.

Luotaintutkimus koostuu viisi arkipäivää kestävästä tutkimusjaksosta, jonka aikana osallistujat lähettävät WhatsApp-sovelluksen kautta kuvia ja kommentteja työssään ja arkielämässään kohtaamistaan tilanteista, käytännön ongelmista ja haasteista. Lopputulokset raportoidaan nimettömästi ja valmiissa työssä huolehditaan, että vastauksista ei voi tunnistaa luotaintutkimukseen osallistuneita ihmisiä.

Luotainjakson jälkeen osallistujat haastatellaan. Haastattelu kestää noin tunnin ja se voidaan tehdä mahdollisuuksien mukaan joko kasvokkain tai Skypellä. Haastattelussa käydään läpi muun muassa luotaintutkimuksen aikana esiin nousseita asioita.

OHJEET

1. Luotaintutkimus toteutetaan WhatsApp-sovelluksen (tai erikseen sovitun muun sovelluksen) avulla
2. Tutkimusjakson kesto on viisi arkipäivää jokaisen osallistujan oman aikataulun mukaan sovittuna ajanjaksona.
3. Tutkimusjakson aikana osallistujat lähettävät WhatsApp-keskusteluun rauhan rakentamiseen ja kriisinhallintaan, ratkaisuun tai ennaltaehkäisyyn liittyviä:
a. Kuvia
b. Lyhyitä kommentteja ja ajatuksia
4. Kuvat, kommentit ja ajatukset voivat liittyä mihin tahansa omassa työssäsi tai arjessasi kohtaamiisi haasteisiin tai havaitsemiisi mahdollisuuksiin. Myös huomiot siitä, mikä toimii jo hyvin, ja miten keinoälyllä / teknologialla näitä asioita voitaisiin mielestäsi entisestään parantaa, ovat hyvin tervetulleita.

PROJEKTIN TAUSTAA

Kehityspolitiikan ammattilainen Niina Mäki tutkii keinoälyn mahdollisuuksia rauhan rakentamisessa ja kriisinhallinnassa osana palvelumuotoilun opintojaan. Laurean ammattikorkeakoulun lopputyönä julkaistava tutkimus pyrkii kartoittamaan mitä palveluja ja sovelluksia tälle sektorille voitaisiin tulevaisuudessa kehittää. Lopputyön kieli on englanti ja se tehdään osana Futuricen Chilicorn Fundin Rauhankone-yritysvastuuprojektia. Lisätietoja projektista löytyy nettisivuilta www.understandingpeace.net.

Opinnäyteprojektin ensimmäinen vaihe koostuu haasteiden ja mahdollisuuksien kartoittamisesta WhatsApp-luotaintutkimuksen, haastattelujen ja kirjallisuuskatsauksen avulla. Osana Rauhankone-projektia toteutetaan myös neljä työpajaa, jotka järjestetään huhti-toukokuussa 2018. Projektin etenemisestä kerrotaan nettisivuilla ja tuotettu tieto on avointa.

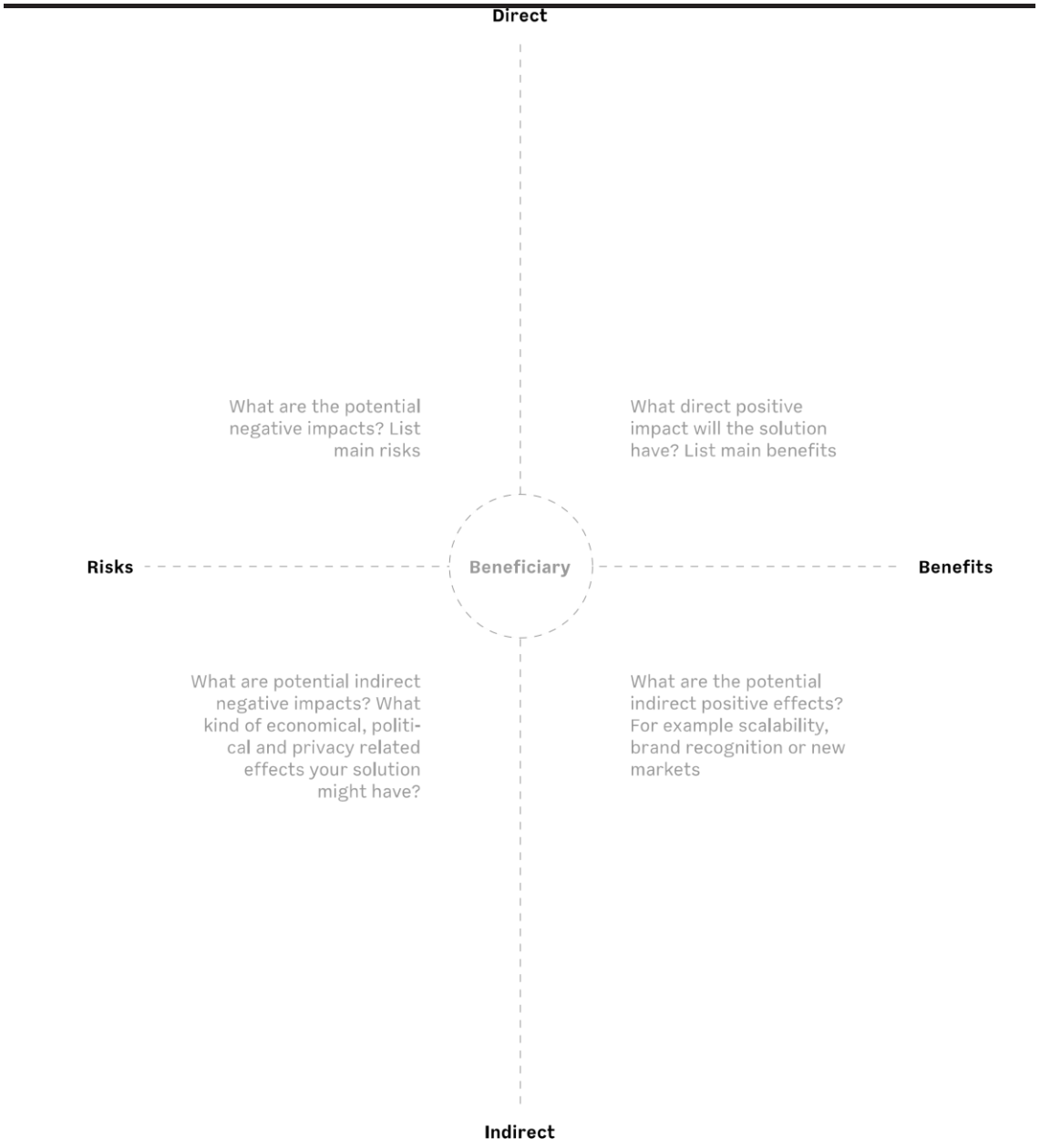
YHTEYSTIEDOT

Niina Mäki, niina.maki@student.laurea.fi, 044 526 5454

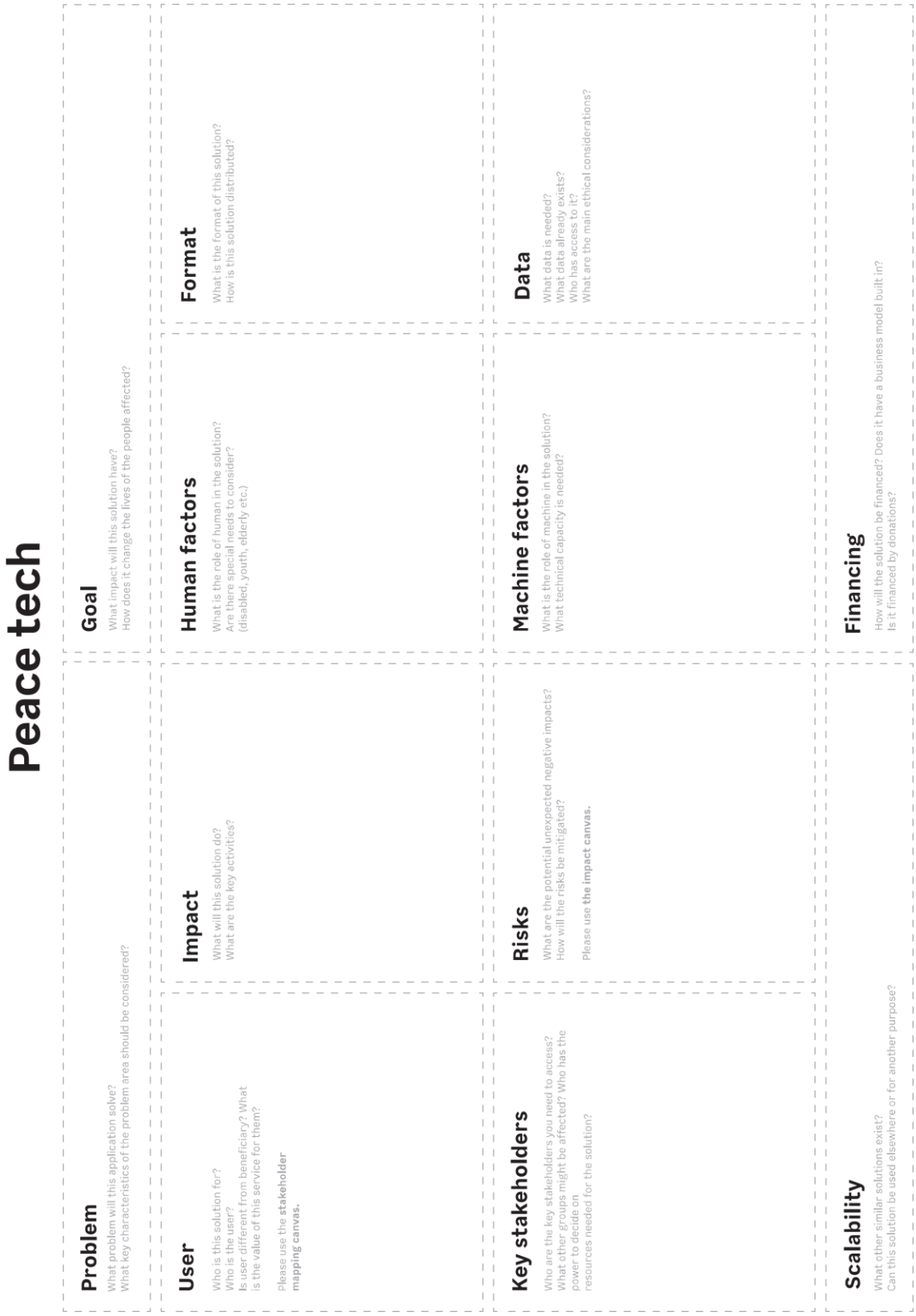
Appendix 6 Set of three canvases used in the Project Peace Machine final workshop

Stakeholder mapping

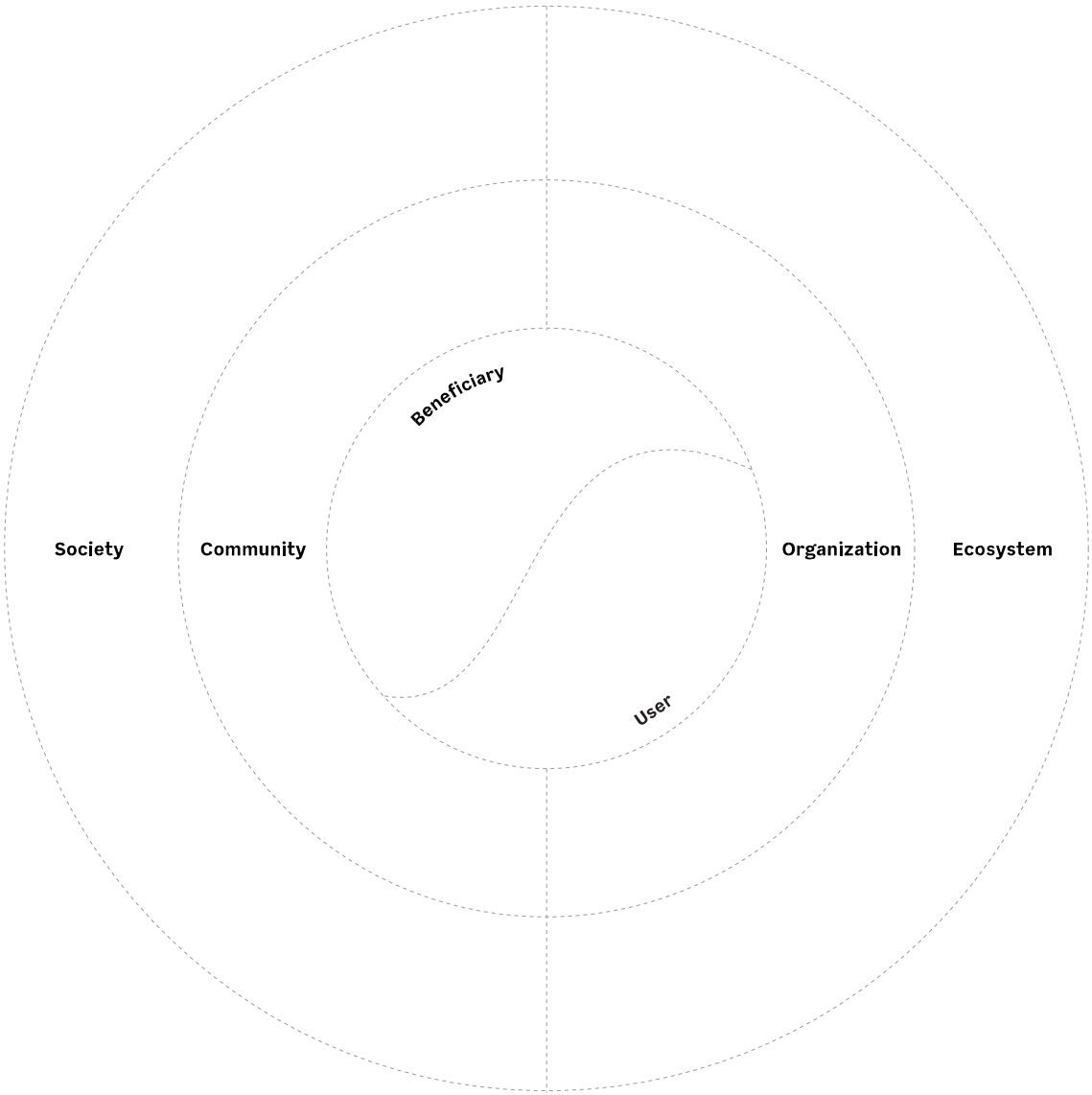
Impact


Direct
What are the potential negative impacts? List main risks
What direct positive impact will the solution have? List main benefits
Risks
Beneficiary
Benefits
What are potential indirect negative impacts? What kind of economical, political and privacy related effects your solution might have?
What are the potential indirect positive effects? For example scalability, brand recognition or new markets
Indirect

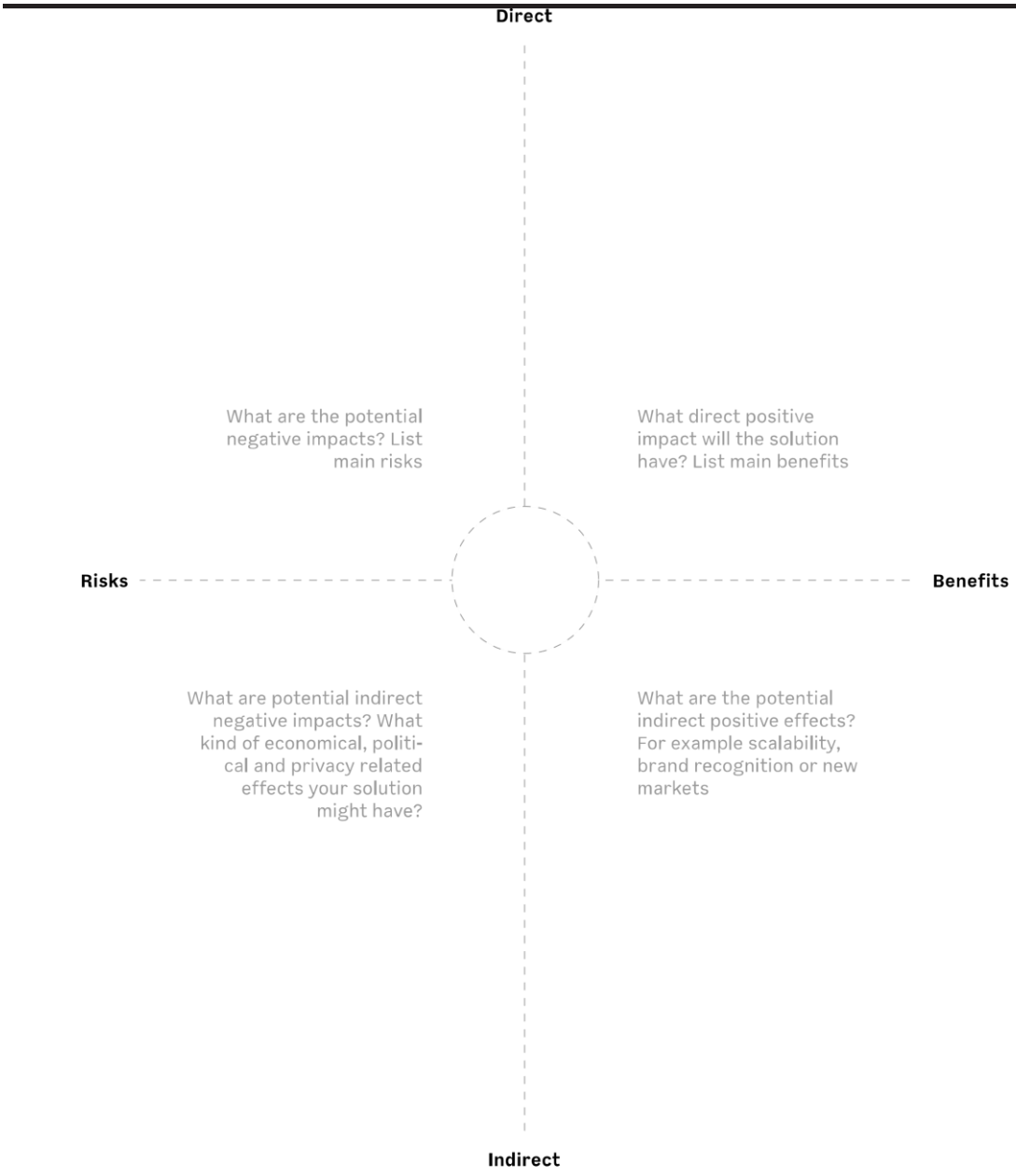
Appendix 7 Set of three canvases used in the testing canvases workshop



Stakeholder mapping



Impact


Direct
What are the potential negative impacts? List main risks
What direct positive impact will the solution have? List main benefits
Risks
Benefits
What are potential indirect negative impacts? What kind of economical, political and privacy related effects your solution might have?
What are the potential indirect positive effects? For example scalability, brand recognition or new markets
Indirect

Appendix 8 Final set of three canvases for responsible design of AI peace technology

PEACE TECHNOLOGY

Summary canvas

Name:

Date:

Problem

What problem will this application solve?

Solution

What will be the change for the people affected?

User

Use the solution mapping canvas

Who is the user?
Who is the beneficiary?
Are there special needs to consider?

Impact

What will this solution do?
What is the value for the user?
What impact does it have on society?

Human factors

What is the role of human?
Are there capacity or accountability issues to consider?

Machine factors

What is the role of AI?
What technical capacity or infrastructure is needed?

Stakeholders

Who else is affected?
Who else needs to be considered?
Who can decide on the resources needed?

Risks

Use the impact canvas

What are the risks and potential unexpected negative impacts?
How will these be mitigated?

Format

What is the format of this solution?
How is distributed?
Who has access to it?
How is accessibility considered?

Data

What data is needed?
What data exists?
Who has access to it?
What are the ethical and safety considerations?

Scalability

What similar solutions exist?
Can this solution be used elsewhere or for another purpose?

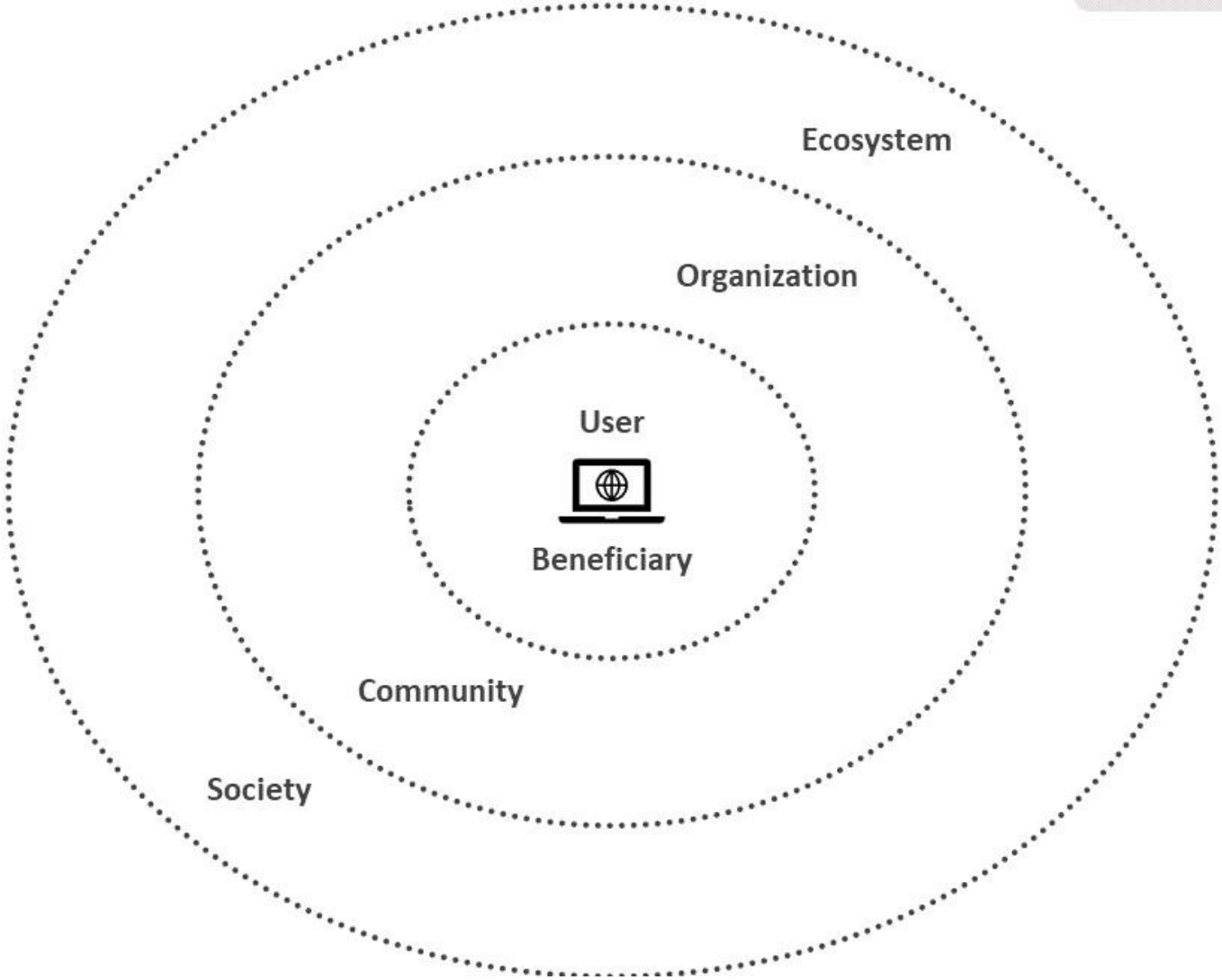
Financing

How will the solution be financed? Does it have a business model built in?
Is it financed by donations?

SOLUTION ECOSYSTEM MAPPING

Name:

Date:



Name:

Date:

IMPACT

Direct

What positive impacts will the solution have?
List main benefits

What negative impacts can the solution have?
List potential risks

Benefits

Risks

What are the potential indirect positive impacts?
(e.g. scalability, brand recognition)

What are the potential indirect negative impacts?
(e.g. safety, privacy, political, economic)

Indirect