Steffen Dean Turnbull

How does the use of technology in innovative public services influence the way we relate to the world around us? A case study on the use of artificial intelligence in Labour Market services

Master Thesis

at the Chair for Information Systems and Information Management
(Westfälische Wilhelms-Universität, Münster)

Supervisor: Dr. Amirouche Moktefi
Co-Supervisor: Prof. Dr. Anu Masso

Presented by: Steffen Dean Turnbull
[redacted]

Date of Submission: 2022-04-27

Content

Figures

Abbreviations

AI	Artificial Intelligence
ADM	Automated Decision Making
ATM	Automated Teller Machine
EUIF	Estonian Unemployment Insurance Fund
GPS	Global Positioning System
ICT	Information and Communication Technology
OECD	Organization of Economic Co-operation and Development
OTT	OTsustusTugi
PIN	Personal Identification

1 Introduction

1.1 Motivation

The world that we currently refer to live in tends to be dynamic. Besides the unforgiving and continuous natural change, this dynamic is often referred to a likewise continuous and in contrast rapid change of our social and technological ecosphere. This might lead us to the question, if this dynamic can be, as merciless to our human understanding, as our laws of nature are to our physical surroundings. We can certainly assume that the omnipresence of information and communication technologies (ICTs) has not only had a major impact on our planet and society, but that it fundamentally changes the way we as humans experience the world (Levin and Mamlok 2021). The Internet is one of the most prominent examples. It is not only a tool that effectively helps deliver services of any means and format, but what it created was a completely new environment that yielded a differentiated set of our basic assumptions, principles, and habits, compared to citizens that lived in a pre-digital age (Sacasas 2020).

The influence that modern digital technologies have on our individual as well as our community lives, is not only evident but arguably unmatched. The contemporary development in the field of computing has led to a thriving progress of technological and communicational convergence. From stationary to more and more mobile, ICTs have become an “intrinsic element of today’s society and each of its members’ lives” (Levin and Mamlok 2021, p. 1). The results have not only gradually altered the reality of present lives and experiences but are drastically different from the historical customary perspective that embodied the majority of humankind’s reality (Levin and Mamlok 2021). Digital technology is reshaping the human experience at both an individual and a social level and if we assume that social innovations in our cultures are aimed at creating public values and solving problems based on principles, then the use of technology is inclined to focus on what it is to do and how it will do it. These standards are set by us and serve to achieve what we argue to be fundamental values. Assessing and evaluating technological innovations however shows that we mostly aim to measure the technical parameters of the solution ranging from efficiency, precision, truth, to economic strength, and so on (Levin and Mamlok 2021).

Supposed that technological solutions do actually improve modern social problems, like traffic, environment or health concerns, the question that persists is what it implies for our human experience. Is there a price that we pay as individuals in order to attain what we as a technological modern civilization value to be our societal needs? The German philosopher Emmanuel Kant once said that “freedom actually means also freedom from

goals, which you are not told how to get happy, but you as a citizen decide how you are happy" (Drechsler and Kostakis 2020, p.1).

Speaking of philosophers and technology, we see that throughout the past centuries a distinct discipline of philosophy of technology has correspondingly and steadily emerged. The philosophical discussion of a technology, which was principally reflected as a process that was not only remarkable but predominant, focused on the explanation of its relations to other philosophical fields such as morality but also to the political world and the truth itself (Schomberg and Blok 2019 based on Jonas 1984; Winner 1980, 1983 and Heidegger 1977). Over the years this branch of philosophy asserts an alternative way of analyzing the relation of philosophical matters and "must make technology a foreground phenomenon and be able to reflectively analyse it in such a way as to illuminate features of the phenomenon of technology itself" (Ihde 1993, p. 38). Taking this into account, academics have aimed to make use of empirical methods in order to analyse particular artifacts. Philosophers of technology have tried to focus on the concrete usage and the context these artifacts are used in (Verbeek 2005).

If we look at specific examples of Bruno Latour, "Speed Bumbs help us make the moral decision not to drive too fast near a school. Ultrasound scans help us to ask and answer moral questions about the lives of unborn children. Energy-saving lightbulbs take over part of our environmental conscience. Coin locks on supermarket pushcarts remind us to return each cart neatly to its place. Turnstiles tell us to buy a ticket before boarding a train. Current developments in information technology show this moral significance more explicitly. With the development of ambient intelligence and persuasive technology, technologies start to interfere openly with our behaviour, interacting with people in sophisticated ways and subtly persuading them to change their behaviour [...]" (Verbeek 2011, p.1–2). The relation between humans and the world reconciles with particular technologies and becomes a factor in our moral decision making of day-to-day life. Therefore, a framework that conceptualizes this significance could help in understanding the use of technology and its implications for citizens (Verbeek 2011).

Another example of the rapid integration of technology in our surrounding environment and reality are Smart cities. They have become one central component of the common understanding of typical public sector innovations, but just because professionals in the field of technology and innovation have developed something that can arguably be of value to the citizen, it does not "imply happiness and goodness, which are rarely seriously addressed" (Drechsler and Kostakis 2020, p.1). This goes back to the idea of social innovation. While the recent focus is rather on what the innovation does, we might have to think more about what innovation means in a conceptual way. Obviously, we tend to

accept it as a technological innovation. This however denotes not only a technological but also a commercial implication (Schomberg and Blok 2019). Building on the original concept of innovation, which was political and therefore fundamentally referred to as a factor of change to the established political order. “Only with the rise of mainstream economics did the concept of innovation become understood in the light of an intrinsic relation between technology and the market, particularly in the period of the last 60 years” (Schomberg and Blok 2019, p.311).

Innovations are commended as the definitive solution to every type of challenge (Godin 2008), which means that they have become an essential aspect of modern processes targeted towards tackling anything we tend to do (Este 2013). This becomes evident in the recently launched initiative of the European Union called the innovation Union, which noticeably underscores the importance of innovations in the goal of increasing the European prosperity. The stimulation of economic growth while assuring ecological and environmental sustainability might be the biggest challenge of our time and for this reason, innovations are set to define not only emerging technological artifacts and services but will shape the entire era we live in (Schomberg and Blok 2019, Blok 2019).

1.2 Problem Definition

Taking a closer look at the latest thought about innovation, this idea in depth changes the way we could think about the human-world relation, which is no longer simply arbitrated by particular innovations, but could also be influenced by the specific techno-economic view we have, that has been widely accepted of innovation in our digital age. This means if particular technologies differ from one another in different ways, the human-world relation is thus also mediated in their own way (Schomberg and Block 2021). Therefore, taking a closer look at possible use cases and technologies this paper will later focus on.

We are just witnessing the birth of a new phase when it comes to data. The advanced research and efforts in the field of artificial intelligence (AI) not only changes and improves the way organizations are handling the vast amount of data that any public organization in the digital age is dealing with, but it also produces in some sense completely new ways of harm, as for example inflicting inequality. The exploration of automation in public service provision has shown that the use of artificial intelligence “disempowers the poor and vulnerable by removing the personal knowledge and empathy of social workers from the bureaucratic environment, by ensuring that errors persist and are passed along the bureaucratic chain of care provision, and by making it harder for people to correct errors where they are identified” (Taylor 2021, p.8, Eubanks 2018).

Algorithm-based optimization of processes could therefore show severe disadvantages to citizens referred to as non-users of technology. Digital Inclusion is often a central point of digital public agendas and still the issue regarding participation can hardly be ignored when it comes to artificial intelligence. An example of this where "those who use the right devices to communicate their needs become visible at the expense of those who do not is how Uber's optimization decreases support for public transport systems, which in the US affects the mainly lower-income people who rely on them" (Taylor 2021, p.8, Kulynych et al. 2020).

The other important aspect of this problem is that artificial intelligence related technologies that public organizations might use are severely dependent on private sector interventions and the power of commercial firms. This means that the use of private systems to handle public service functions could frequently break the important link that exists between citizens and public authorities while simultaneously creating a potential domination of power relations and questioning the legitimacy of public acts (Taylor 2021). Automated methods can additionally have serious implications on the user's reliance and relation to them. The initial trust towards the decision and thereby the technology itself can easily be obstructed but single error incidents can change the relation between the user, technology and coworkers completely (Dzindolet et al. 2003, Kuziemski and Miscura 2020).

A recent study concerning Street-Level Algorithms and artificial intelligence in bureaucratic decision-making showed that from the workers perspective the use of artificial intelligence can be considered valuable in a management decision making context when the specific case requires additional resources but also has some positive effect on the receiving individuals by strengthening their cause in relation to the individuals received service (Fluegge et al. 2021). Meanwhile the raised ethical concerns and long-term impacts on public-citizen interactions and its philosophical instruction generates their own problems because the use of technology encourages different behaviors and a more impersonal approach to helping vulnerable citizens apply for social welfare services. This is why this study on the post phenomenological perspective suggests that's the perception of artificial intelligence decision making software is not only accepted by its affordances but is also multi-variant across users and circumstances.

The special case of Estonia derives from its pioneer position in the field of digitalization. The commitment towards IT-infrastructure is a steppingstone for modern digital public services and e-government. Especially the Estonian labour market and related policies have shown to represent the liberal spirit of a deregulated space where digital technologies can be integrated successfully. However, studies have also identified the strong separation

between the digital world of highly skilled IT personal and the less IT literate citizens. The later have also been identified as mostly rural and older citizens that tend to represent the undesirable high numbers on long term unemployment figures in Estonia. The most recent neo-liberal pursuit of resolving these very problems with the assistance of modern AI Technology, establishes the fundamental difficulty of Innovation driven policy regarding social services and the people's relation to the public sector. Questions concerning the complex process of integrating disruptive new technologies to handle and unravel these and many similar challenges, motivates this work to express on the potential problems related to the expenses and influences users might experience or perceive. In other words, how does AI technology inflict certain non-neutral change to the user when thinking of the impact the technology has on their reality and therefore their own personal "world" (Buhr et al. 2016). With the more or less recently introduced OTsustusTugi (OTT) decision-support system the Estonian Unemployment Insurance Fund (EUIF) has committed to introduce artificial intelligence to raise the quality of services provided to the unemployed citizens of Estonia and increase the efficiency of organizational processes itself. Now that the system has been in use for roughly two years, the developers and the EUIF claim consummate success and present a possible approach of how modern digital services have been changed to the better. Nevertheless, it remains interesting to evaluate the system beyond the existing studies that have addressed current impacts of the system and to introduce are more philosophical perspective to the influence it has on the users and society (Nortal 2021).

Consequently, guided by workspace observations, public official interviews and qualitative post phenomenology analyses, a severe and deeper look into the underlying personal effects of artificial intelligence technology ethics and philosophy from a user's perspective could help elaborate a better understanding of this problem while common business ethics and the private laws are not designed to answer these questions, which are primarily political" (Taylor 2021, p.1).

1.3 Research Questions

Derived from the two previous subchapters and the implications of this paper's motivation and problem definition, following research questions can be named in order to address these.

- Q1: How do users at the Estonia Unemployment Insurance Fund (EUIF) apply artificial intelligence to facilitate their work and offer social services to citizens?

- Q2: How do the users at the EUIF perceive the influence of Artificial Intelligence on their working practices and interactions with citizens or colleagues?

- Q3: What are the ethical considerations of AI in providing social welfare services for employment seeking citizens?

This study is supposed to invoke a closer examination of public officials' current experiences, practices and use of automated software, the ethical foundations that affect public officials' uptakes, habits and decisions regarding the technology implementation, and the co-constitutive human-technology relations that equally reinforce and constrain various aspects of the service decision making experience.

1.4 Study Structure

This paper begins with a literature review addressing the recent evolution of ethical and perceptional studies of artificial intelligence technology in decision making, with a main emphasis on public services followed by a short overview of the most recent automated approaches for labor market management. The next chapter discusses the proposed research method for this study and the field of use by explaining post phenomenology and its conceptual framework as well as the methodological approach in the field of philosophy of technology. Following this the next chapter outlines the research design used in this paper to conduct this study of artificial intelligence technology in decision making. A presentation of the data findings in addition to a discussion and implications of the findings are outlined in the next chapter. An at the end a short description and the main insights of this research study are concluded.

2 Literature Review

2.1 Introduction

A possible perspective on artificial intelligence technologies refers to devices that have the ability to maximize the chance of success on taken actions toward a specific goal based on the computational perception of environmental intakes and analysis (Russel & Norvig 2016, Sun and Medaglia 2018). This means that when AI technologies are mentioned during this research it includes technologies such as machine learning, rule-based systems, natural language processing, and speech recognition (Eggers et al. 2017, Sun and Medaglia 2018).

A popular word among the most recent technological dynamics has been transformation. In the private and public sector there is much talk about transforming entire organizational structures, business models, services, mindsets and so on. The reason seems to comply with the newest economic and societal needs. But why do we feel the need to transform everything? This is but a seminal question of our time that advances us to ask what are the metrics of our collective success? While the world itself is in constant change our global disenchantment seems to grow as persistently, especially with the digital state of the art represented by our economy, the single foundation of agreement seems to be a deeply rooted intrinsic need to innovate. A dynamic world always appears to create new problems every moment and irrespective of the problem or its nature the answer seems to be innovation. What can be observed is some form of defensive mechanisms that governments have inherited which implies all public functions and decisions are but under the pledge of innovation in order to appear economically dynamic, politically legitimate and suited to the unique moment in history (Pfotenhauer and Jasanoff, 2017, Kuziemski and Misuraca 2020).

With this notion of innovation, the spread of AI has reached a level of penetration that includes a wide range of different application areas including the public sector (Desouza 2018). Possible examples of AI systems that are adopted in the public service sphere are algorithmic based support systems which are predicting different outcomes like risk of youth for targeting interventions or language processing tools allowing better and faster written transcripts on policy debates (Chandler et al. 2011). This transformation of the public sector was also accompanied by unwanted uncertainty and a certain degree of hesitance that meant a situation where the private sector deliberately shaped the world of AI around us. The use of such technologies can enable many benefits that include efficiency and effectiveness of formerly unproductive bureaucratic processes by the automation of “cognitive labour, freeing up high-value work, augmenting predictive

capabilities for decision-making, and improving services to citizen queries" (Eggers et al. 2017, Sun and Medaglia 2018, p.1). While possible challenges and negative side effects of introducing AI technologies are related to the "destruction of jobs caused by automation, the infringements of privacy caused by digital surveillance (The Economist, 2016), and the reinforcement of biases in policymaking caused by algorithmic governance" (Janssen and Kuk, 2016, Sun and Medaglia 2018, p.1).

This means that as the pressure on the deployment of automated decision-making systems in the public sector increases, academics have acknowledged the importance to examining "how machine learning and bureaucracy have both become generalizable modes of rational ordering based on abstraction and deriving authority from claims to neutrality and objectivity" (McQuillan 2019, Kuziemski and Misuraca 2020, p.2). Further literature showed that there is a limitation on research that simply focusses on the implications of algorithmic bias, which means that it could be of interest to consider a more politicized discourse of AI and other basic frameworks in analyzing algorithmic accountability and extends them. (Dave 2019, McQuillan 2018, Pasquale 2019, Kuziemski and Misuraca 2020).

We can therefore see that there are still many critical and ethical research questions open that need to be answered. While in general the introduction of digital technologies in the public sector tends to be primarily depicted as favorable to the users and recipients of services there are challenges that are not as predominantly addressed. Some include the processes that happen under the slogans of democratization, convenience, and imply the idea of more choice to citizens while serving they are not always transparent but praised purpose. "Digital technology as disguised attempts to strengthen the grip of control over the citizens? In other words - is AI facilitating the power shift between the public sector and citizens or merely intensifying existing distribution? Is the use of AI in the processes of governance changing the way power is exercised?" (Kuziemski and Misuraca 2020, p.2). These are but a few of the critical questions we encounter when studying the controversial side(s) of AI technologies. What becomes clear is that these technologies and the resulting issues are definitely not neutral to the users, citizens and the public organizations (Kuziemski and Misuraca 2020).

To achieve a better understanding of the problem and framing the research objective of this study, it was necessary to identify, classify and consolidate the current scientific knowledge that has been produced on the topic of this paper. The literature search itself has been performed between December of 2021 and March 2022. The used method of choice is a structured literature review using the PRISMA framework. In the first phase during a preliminary search, the scope of the studies of interest were determined. This

step contributed to the definition of the research objectives and the inclusion and exclusion criteria for the literature review. The Inclusion criteria can be described as follows:

- Studies published in the English language since 2000.
- Academic or commercial research focusing on the use of Artificial intelligence in public sector related organizations or institutions on public service delivery in all geographical locations.
- Academic or commercial research focusing on the implications of adopting AI including ethical, cultural, individual, organizational, and economic levels.

The exclusion criteria on the other hand are described by:

- Academic or commercial research focusing on the private sector or no specific mentioning of public sector AI use.
- Academic or commercial research focusing on the technological aspects of AI and Design.

The research itself was conducted using the three following databases, Science direct, EBSCO and Google Scholar. The initial selection was merely based on the title as well as abstracts that appeared while using the following search terms and structures (“public sector” OR “public service” OR “government” OR “public administration”) AND (“AI” OR Artificial Intelligence” OR “Automated Decision Making” OR “Algorithmitimatization”). And while this search was not aimed at achieving the exhaustive representation of the public sector AI research field, the found results were used to create a better overview of the current knowledge and show the importance of a post phenomenological study on the implications for users and data subjects.

The research came up with over 200 initial results. After screening the articles and reports for eligibility, using the defined criteria the number was reduced to 102 remaining papers of relevance that might help outline the current problem definition and research objective of this study. The complete overview is shown in the form of the created Prisma flow diagram (See Figure 1). The findings of the literature review showed that over the past twenty years a steady incline of research that primarily focuses on the field of public sector AI occurs, as can be seen in figure 2. Among the found results the literature helped demonstrate the more detailed attention of the most recent research. Studies that have been selected, which were published during the last two years, showed a significant rise towards the ethical, and social implications of AI in the public sector. While five papers

have also been looking at concerns regarding transparency the main area of research that was identified in the most recent literature was still concerned with the existential tradeoffs that come with integrating AI into the public sector, including efficiency and workflow matters (See image 3).

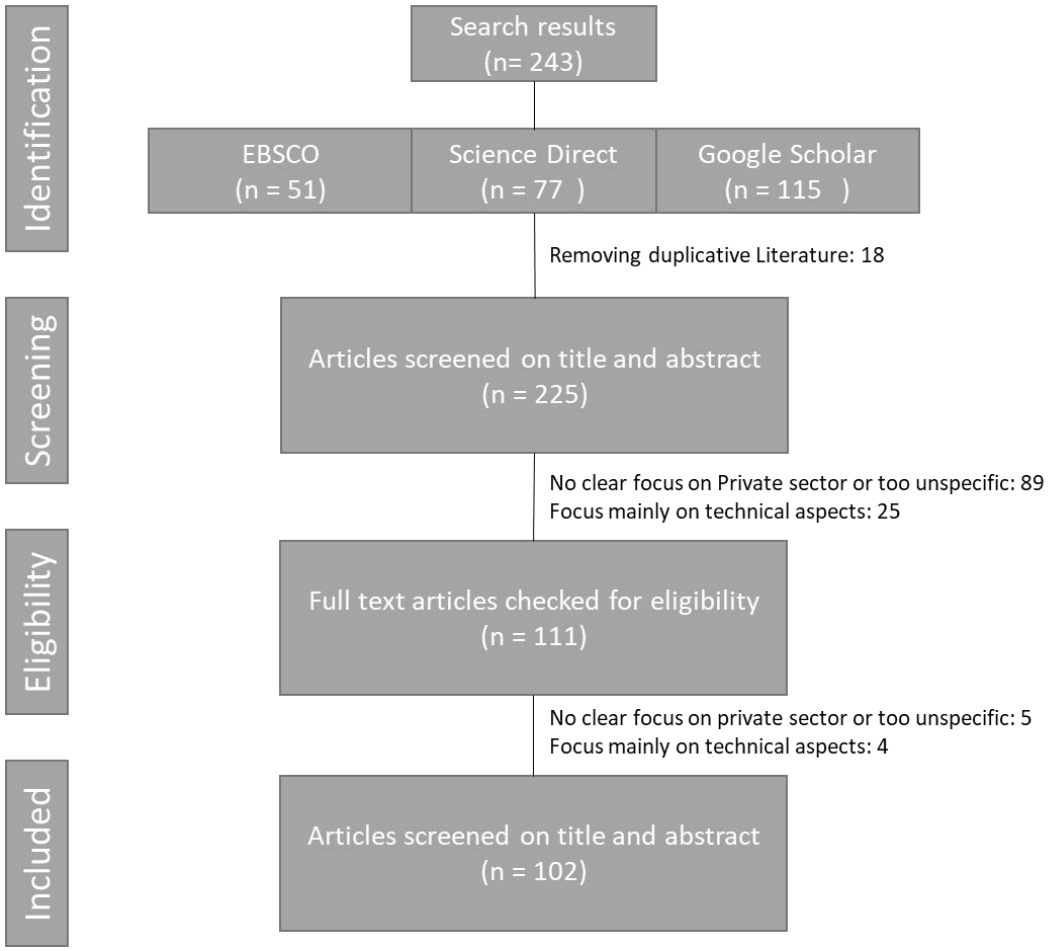


Figure 1. PRISMA Flow Diagram

The main contributions on AI impact research in the public sector that is of relevance to the consideration of a post phenomenological case study exploring the experiences of AI users delivering services in the public sector can be further categorized into, Ethical implications of AI, Transparency implications of AI, Perception and Experiences towards AI, Trade-offs of AI, and other perspectives of public sector AI. The following subchapter will elaborate on each of these contributions and outline the importance of this knowledge for the following research. Finally, this process could help understand the use of automated decision-making applications such as the later introduced OTT system from the Estonia Unemployment Insurance Fund and define its primary function of assisting the career counsellors in providing their services and tracking citizens employment behaviour. The public sector and social service providing institutions have employed various methods of engaging with employment seeking people and therefore handling

important labor market situations for citizens and associated institutions, including the use of automated decision making (Fluegge et al. 2021).

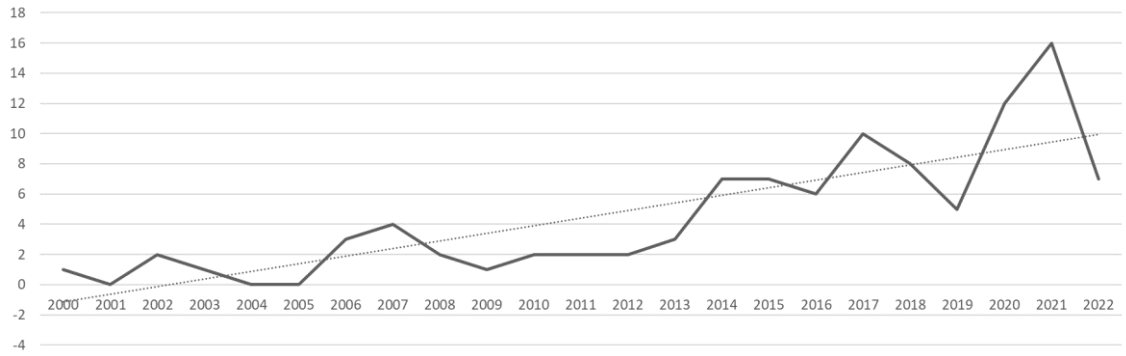


Figure 2. The Evolution of Public Sector AI Research

Figure 3. Research Distribution since 2020

However, the next chapter also shows that the rapid advancement and ubiquity of technology such as AI based automated decision-making tools might be the foundation for a more comprehensive or at least more systematic approach of mediating public to citizen interactions that is based on data. As part of this review, it became clear that these applications have certain characteristics but also affordances which appeal to the public sector and therefore also the case of the Estonian Unemployment Insurance Fund seeking for more advanced and innovative ways of engaging and helping vulnerable citizens. This post phenomenological study later claims that with each technology however, certain features are being amplified while others might be reduced. Examples of this can be showcased by OTTs ability to provide the career consultants the capacity to oversee and the information of each employment seeking citizen and track his resulting behaviour, due to the storage of larger amounts of data that might also raise concerns of privacy or detrimental use. This review will also focus on these potential detrimental implications and explain them in more detail in the course of the literature findings. In light of this, this chapter begins with a more general examination of the AI landscape in relation to the

ethics of such technology, followed by the progression of automated decision making in the public sector towards labor market and social service application cases.

2.2 AI Ethics and Moral Agency

The prospect of AI can be described as the universal helper of humanity. The confrontations that it is aiding are often urgent and exciting but at the end there is still a legitimate reason to worry. The occurrence of new and rapidly changing technology comes with a steep learning curve that does not only imply innovation but also means mistakes and miscalculations that are leading to equally important impacts whether they are anticipated or harmful. Therefore, AI technologies are no exception of this and the attempt to include ethics and safety may be essential to the development of AI, but more importantly it is indispensable for the creation of public benefits. Ethics and its definite implications should be integrated to the underlying design process at every stage of the AI system delivery. This goes hand in hand with a more collaborative effort between different direct stakeholders and indirectly involved people as an attempt to align the development of the technology with the ethical values and insights derived from public and academic discussions (Leslie 2019).

While AI studies integrate the thought of a fundamentally moral principle, the techno ethics might help focus on the social embeddedness of the OTT technology and ethics, with special regards to the use and misuse of the system in the wider range of society in addition to the beliefs and the form of governance and policies that are of influence. This means that that ethics of technology can be considered as something interdisciplinary whilst the application to AI technology manifests its own form. And from the phenomenological perspective the work on ethics is a possible way of revealing the obscure attitudes of the AI technology. Hence, when this study later on tries to investigate how the use of the OTT system can frame the users' experiences and his behaviours, AI ethics helps raise questions, assumptions about the system it's innovations and the underlying evolution of this technology (Introna 2017). Combined with the help of post phenomenology the concept of techno ethics is extended by highlighting the possible technologically mediated relations to the user and his world (Verbeek 2009).

But why are there more and more considerations concerning the ethics of AI technology? The discussion on this topic did not emerge out of nowhere, nor is it present for the sole reason of confrontation, but it is rather a direct response to explicit experiences. The range of societal but also individual harms can be derived from misuse, abuse, design failures or general unintended and negative impacts as an immediate consequence of the AI system and its utilization. Potential harms that have been identified in the past include the bias and discrimination done by AI algorithms. This means that by gaining their

information from real world data representing the structures and dynamics of society, models created on the basis of any data driven technology can and often will reproduce, reinforce, and amplify real world challenges like marginalization, discrimination, and inequality. The designers bias and preconceptions also have to be considered and can likely be replicated by the system. (Leslie 2019).

When citizens are the main subject to the decisions, predictions and classifications made by a technology, the citizens might fail to hold accountable the entity, whether it is in the form of the system or a human, responsible to the impacts caused by the decisions, predictions and classifications that were done. Humans were previously perceived and attributed as exclusively accountable agents that perform cognitive functions. Now, we must think of new agents in the form of AI that are automating these very cognitive functions but leaving the question open of who is responsible for their behaviour? (Leslie 2019).

In some cases, the outcome of an algorithmic process may have direct affects to the decision's subject and while in some cases a missing or weak explainability can be accepted, other scenarios will cause problematic implications. Supplementary harmful outcomes and issues arise by the potential invasion of the citizen's privacy. The online availability of sensible information has led to a severe threat that individuals and organizations are facing. The cumulative digital world has become of great importance to the training and operation of AI systems that are designed on data structures and as part of their development and deployment processes require large amounts of data. This also subsequently entails that it will often rely on personal data, that the citizen might not want to share. Countless times data and information are captured and extracted without a common consent and therefore privacy invasion as a consequence, might lead to individuals no longer being able to pursue their interest free from unwanted influence or interference (Leslie 2019).

Beyond that, ethical disputes also arise around the isolation and disintegration of social connections. This means that AI systems are often developed and carried out with promises related to enhancing or enriching a person's experience and the possibility of personalizing digital services in general. However, this promise does not only imply that the system can vastly improve the citizens life or the service delivery process, but it holds the risk that automation might reduce the exposures of individuals to real life events and experiences. When we consider replacing human to human interaction with human to technology alternatives and adding a sense of hyper personalization of their digital space, the subject's world view and his actual "world" might be polarized to certain relationships. As a direct consequence there could be a societal shift towards lower levels

of trust, empathy, and mutual understanding. (Leslie 2019). If we apply this current knowledge on the ethical implications of AI to this papers research, it becomes clear that there is a profound gap of certainty especially in the public sector concerning the awareness and integration of AI ethics and that is why the outlined research questions try to include the users' perceptions in order to analyse the ethical implication of the chosen case.

2.3 AI Transparency

2.3.1 AI as Sociotechnical

This works motivation to see AI from a different perspective is broadly related to its impact on individuals' behaviour. The literature and general perception on AI have long begun to examine and research other effects of such technologies that could have detrimental effects (Sabanovic 2010, Shneidermann 2020, Vaughan and Wallach 2020, Ehsan et al. 2021). AI technologies are positioned in a social context and their development is carefully taking into consideration social, organizational and cultural factors that have a severe influence on the usage and government of the system. This means that in contrary the public sector or any other applicable area could risk deploying a socially un-integrated and individual AI system to the organizational workflow and therefore potentially introduce misuse, mistrust, or profound ethical risks with unintended consequences, specifically for socially weak and marginalized groups as was mentioned in the chapter before. (Makarius et al. 2020, Wolf and Blomberg 2019, Yang et al. 2019, Yang et al. 2016, Mohamed et al. 2020, Sanchez-Monedero et al. 2020, Suresh and Guttag 2019, Ehsan et al. 2021).

What the literature showed, was that while using a scientific technological lens on this topic, scholars started to critically reflect and question the underlying assumptions that are being made by AI using algorithms and neural network systems. This insight was followed by suggestions that were concerning the application of a more decolonial approach to enable AI technologies and focusing on vulnerable groups that are endangered by the negative consequences of the specific technological innovation (Suchman 1987, Mohamed et al. 2020, Ehsan et al. 2021). Other scholars have "diagnosed the dominant mode of AI algorithmic reasoning as "algorithmic formalism" – an adherence to prescribed forms and rules – which could lead to harmful outcomes such as reproducing existing social conditions and a technologically-deterministic view of social changes" (Green and Viljoen 2020, Ehsan et al. 2021, p.8-13). This means that while addressing potential harms it requires attention on the internal limitations of the algorithm itself and the socially related concerns that are outside the algorithmic formalism context

(Green and Viljoen 2020). In addition to this the literature has started to question the direct implications of algorithmic abstraction that is the fundamental part of machine learning systems (Selbst et al. 2019).

The literature has shown that this abstraction of the broader social context has several possible unintended outcomes when it comes to the development and application of AI technologies. These traps are called “Framing, Portability, Formalism, Ripple Effect, and Solutionism” and described as follows (Ehsan et al 2021, p.8-13).

- Framing: This is the failure to model the entire system over which a social criterion, such as fairness, will be enforced.
- Portability: This is the failure to understand how repurposing algorithmic solutions designed for one social context may be misleading, inaccurate, or otherwise do harm when applied to a different context.
- Formalism: This is the failure to account for the full meaning of social concepts such as fairness, which can be procedural, contextual, and contestable, and cannot be resolved through mathematical formalisms.
- Ripple effect: This is the failure to understand how the insertion of technology into an existing social system changes the behaviors and embedded values of the pre-existing system.
- Solutionism: This is the failure to recognize the possibility that the best solution to a problem may not involve technology.

However, the consequence of this can be expressed by a need to extend the abstract restrictions and include social factors in addition to the purely technical side. And in addition to this the studies on algorithmic fairness frequently showed that profound involvement in a usable and ethical algorithmic system is non-technical which follows that the localized community solutions have more subjective value to them than scalable ones (Ehsan et al 2021, Katell et al. 2020, Kyung Lee et al. 2019).

2.3.2 Social transparency

Deriving from a social perspective of Artificial Intelligence another concept called social transparency could help understand the profound issues of AI technologies in a social and individual context. The sociotechnical concept that was discussed before is based on the previous work in the field of social transparency, which relates to human-to-human interaction concepts. Obtaining from this, the literature showed that transparency plays a

as crucial role in human computer interactions and computer-Supported cooperative work (Ehsan et al 2021, Star and Strauss 1999, Suchman 1995). And when extending this thought academics also proposed a perspective on design principles that ensure a degree of social translucence. This means that other person's social hinds and present activities can be observed as part of the digital system, giving the user a foundation on which to apply their familiar social rules that facilitate effective communication and collaboration (Ehsan et al. 2021).

Therefore, resulted the idea to a possible operational design framework for AI solutions based on group awareness and distributed team support principles, where the literature has studied the inclusion of these very elements. Which indicates that knowledge that constitutes group awareness includes awareness about "Who, What, and Where to support awareness related to the present, and How, When, Who, Where, and What for awareness related to the past" (Ehsan et al 2021, p.8-13, Gutwin and Greenberg 2002). Henceforth theory building has partially shifted towards new design features and spaces that are inspired by social and collaborative technologies. The concept of social transparency develops upon more social translucence and awareness to support information and network exchange. The notion is that one's visibility towards others as well as the effect of their interaction is made visible and helps encompass the user's role as an observer of others. As a result of this, social interferences are made visible in terms of perceived similarity and accountability with the aim of utilizing this approach in order to design and analyse various social learning features and impacts (Ehsan et al. 2021).

2.3.3 Transparent AI

The importance of transparency to the field of AI can be seen in the previous chapters elaborating on the potential harms of AI, its ethics and the social transparency. Transparency therefore should not only give the user the ability to know how and why the model performed certain things in every context specific moment, but it also means that there is a rationale behind every decision and or behaviour of the system that can be understood and analysed. This incentive also involves the justification of ongoing processes of design measures and the implemented outcomes as well as a content clarification, intelligibility, and explicability. Outcome transparency and offering explanations to the involved and affected stakeholders is the foundation of showing the underlying processes and complex outcomes in a plain language understandable to non-specialists of the field. This includes providing answers to why certain models performed a specific decision or behaviour. Being able to clarify and communicate the rationale of the decisions and behaviours therefore has a socially meaningful implication that lets mathematical and often complex technological terminology be translated into everyday

language building stronger social factors around the technology. Professional and institutional transparency can help demonstrate that a specific decision or behaviour of the system is ethically permissible, non-discriminatory/fair, worthy of public trust and safety securing (Leslie 2019).

2.3.4 Attitude towards AI

The use of automation tools can be conceptualized by the instances in which the underlying algorithms or the AI system is utilized to perform functions such as collect, process, model and use data to suggest a specific decision. The system is constantly improving by using its own feedback but the socio technical perspective that was introduced shows that this concept goes beyond the technical aspects of the technology (Araujo et al 2020, Kitchin 2017, Elish and Boyd 2018). The idea is to extend the previous conception of fixed and encoded procedures fed by data to come up with the desired outcomes based on inner calculations (Araujo et al 2020, Gillespie 2014) and try to conceptualize the evolution of a system that is in continuous exchange with societal, institutional and human contexts. The result of this research path has shown that algorithms can also be seen in various different perspectives such as technically, computationally, mathematically, politically, culturally, economically, contextually, materially, philosophically, ethically and others (Araujo et al. 2020, Kitchin 2017).

Studies on the use of technology have shown that the general trend among participants was that so called expert systems have long been subject to a higher level of perceived objectiveness and rationality than human alternatives would (Araujo et al. 2020, Dijkstra et al. 1998). The origin of this tendency derives of the assumption related to the extraordinary results of statistical methods in numerous fields and their comparison to human judgements (Araujo et al. 2020, Dawes et al. 1989). This concept called algorithmic appreciation implies that human users prefer judgement or recommendations done by algorithms over human ones (Araujo et al. 2020, Logg et al. 2018) and is associated to the notion of machine heuristics that suggest that "…the less a user anthropomorphizes an interface, the more she or he will consider its decisions and selections to be objective and free of (ideological) biases" (Araujo et al 2020, p.613). However, the general attitude towards algorithmic decisions can also be very mixed. While concerns about risks are in general quite mutual the emerging points of fairness and usefulness have shown to be divided or at least brought forward mixed opinions. Araujo et all phrase it as follows, "domain-specific knowledge, as well as belief in equality and online self-efficacy were associated with more positive general attitudes about the usefulness, the fairness, and the risk of decisions made by AI, whereas increased levels of privacy concerns had a negative association" (2020, p.621).

2.4 AI Trade Offs

2.4.1 AI in the Public Sector

Different scenarios within the literature expressed the urge to closely examine the benefits and misdemeanours, resulting from the direct use of AI systems in the public sector. The importance of a lively academic and public debate is clear when considering that governments in most countries cannot simply expand the deployment of associated technologies in order to provide public services without justification. Outcomes and impacts are of general interest to a wider range of people than just public officials. What the study also demonstrated is that governments on all continents are actively engaging with research institutions with the aim of combining different capabilities, capacities, and opinions on the development of technology that might help solve modern world problems.

Although policies and ethical implications are a reoccurring topic when going through the literature on AI use in the public sector, the concerns are still on a narrow path that is defined by cultural and social issues on the client's side of the system. The majority of studies have analysed and focused primarily on the presentation of new solutions to very practical issues while leaving out any social, cultural or philosophical consequences to the actual user and facilitator. The answer to this concern lies within the nature of AI and its obvious benefits. Based on "the techniques it provides the necessary resources for the development of the solutions, which involve algorithms and methods to process information and provide answers. These responses are materialized in predictions and alerts, calculations, analysis, and learning, which offer services that can improve the performance of the public sector and value to the functions of government" (Sousa et al. 2019, p.7).

But performance is not the only distraction from philosophical and socio ethical questions, The use of AI also contributes to the public sectors internal ability of control. Massive amounts of information and data can now be easily processed, which means that assessments of managerial functions and actions are also performed with more efficiency across various public sector organizations. The overall results of the research done by Sousa et al (2019) have shown that the most frequently recurring public functions belonged to the field of general public services as well as economic affairs and environmental prediction cases. Other relevant field such as defence, culture or education have in contrast been neglected to the academic examination.

The general perception is that AI systems will change the nature of computing in a significant way and there are obvious benefits to the private and public sector, but it also requires institutions to adapt in novel and challenging ways in order to manage these

technologies and address the various ethical and security related concerns that were identified (Desouza and Dawson 2019).

While AI is but a synonym for the great digital transformation that seems to be ongoing in our current world, the potential of changing our society and economic realities is not only a mere narrative. The benefits that come along in the form of new and exciting technologies have as much potential to provide beneficiary outcomes as they likewise create risks. These risks must be managed and in a very democratic sense governed that not only respect certain values but also the fundamental human rights. The European Union demonstrated this fact by defining the aim of developing only trustworthy AI in coherence with ethical and societal values from the European Charter of Fundamental Rights (Kuziemski and Misuraca 2020).

As Kuziemski and Miscura state the public sector plays a vital role in the development and uptake of AI. However, most of the debate tends to place governments either in the role of an "regulatory actor" or at best that of an "facilitator", setting out the framework conditions for private actors and citizens to use AI in an ethical manner. This leaves the alternative role of the public sector as "first buyer" and direct beneficiary of AI take-up and implementation rather obscure, if not neglected. In other words, the current policy discourse focuses on the governance "by" AI, far less on the governance "with" AI (2020, p.3). And what the literature shows is that legislators are committed to contain joined approaches, forces and engagement in a common policy definition between EU member states. The aim of this joined endeavour is leveraging the achievement and investments of AI research across Europe towards a social, economic, ethical, and legal appropriate development while cooperating with business and private research in an appropriate manner of communication. This ambiguous attempt implies a more challenging and intense policy design work and needs to document the national willingness of following the firm intentions of the European level (Kuziemski and Misuraca 2020).

Despite these ambitious aims, the current and past years have shown the adoption of placed AI in public institutions are but a mirror of the acceleration that comes from the private sector and focus only on the enormous benefits that they provide. These benefits can be expressed as economically driven pledges in relation to efficiency and effectiveness, resulting in "faster" and "better" policy making as well as service delivery for citizens businesses and intergovernmental exchange. The goal of this process is of course enhancing the satisfaction level in addition to trust and the overall quality of governance and public services. This means governments are a popular and trending subject to the use of AI technologies, but have however, not yet shown to generate the equal amount of research and examination of its impacts in relation to the actual use.

Therefore, many research and literature is still hung up upon the narrow perception that government is only the regulator (Kuziemski and Misuraca 2020). Nevertheless, machine learning algorithms that are combined with the increasing amount of data sources that are available can be used to radically improve the public sector. If we assume this is done in a responsible way, advanced AI technology seems to pave the way to pro-active public service delivery models and processes while simultaneously offering great relive to the constrained institutions and the responsible personnel of mundane and repetitive tasks.

The observation of the continuous collection and analysis of big amounts of data through the use of AI associated automated systems in the public sector and by governments naturally creates some form of concern in regard to the underlying power relations. These concerns are deeply rooted in our understanding of political legitimacy and on the one hand focus on the unique relation between governments and its citizens and on the other hand embody the rising"…opportunities for citizens to contest recommendations and results of the AI systems used for public services, with citizens finding it unacceptable to use algorithms to make decisions with real-life consequences for humans, especially when it comes to: violations of privacy; lack of fairness; removal of the human element from important decisions; as well as the inability to capture the nuance and complexity of the human nature" (Kuziemski and Misuraca 2020, p.3, Smith 2018).

The public pursuit towards novelty can not only be a powerful driver for governments to adopt new technological solutions and innovations but can also inflate our understanding of what is seriously necessary and considerable. As Kuziemski and Miscura argue, "with the rise of predictive analytics and automated decision making, the very nature of a knowledge base is changing towards overwhelming quantification" (2020, p.3). The implication of carving through the vast amount of data that is created and keeps on getting created leaves algorithms obtain insights as well as certain correlations that may not bet obvious to the human perception in the first place. And while these technological insights or findings are often alleged to be objective and neutral, the consequence of this leads to everyday bureaucrats embracing the risk-based assessment systems based on its tempting promises without thinking about any hidden consequences. The assignment of numeric values "…to any activity and then hiding behind the machine produced 'evidence' shields imperfect humans from the accusations of bias, misdemeanour and inefficiency. At the same time, the drive towards rivalry and international benchmarking increases data collection" (Kuziemski and Misuraca 2020, p.3).

The result of increased data collection from citizens is represented by the data granularity and may significantly increase the overwatch potential and become a central point of interest when talking about tools to support global governance and its related challenges

(Johns 2017, Kuziemski and Misuraca 2020). Therefore, data mining seems to be the basis of any future action and the foundation for anything from sense and meaning making to the more general acquisition of important knowledge related to anything in the world and brought to our collective attention (Kuziemski and Misuraca 2020). Nevertheless, there are various barriers that automation systems face when being subject of implementation in the public sector, deriving from a lack of digital maturity to more context specific challenges like costs. This finding is completed by several negative experiences from related and previously studied public sector use cases (Kuziemski and Misuraca 2020). This can be explained by some common patterns that have been observed when relying on automated methods. First of all, it might happen that stakeholders who "initially consider decision aids trustworthy, after observing it making errors happen to distrust even its reliable applications" (Kuziemski and Misuraca 2020, p.4). Another reason is that the premature adoption of a possible faulty application has shattered the trust in the system itself. Another example can be the introduction of "new resource-intensive processes inside the public sector - especially if they require reskilling and a lot of taxpayer dollars - enters the logic of path-dependency - it is much harder to abandon a flagship and politically salient project that has promised to "revolutionize" a given sector" (Kuziemski and Misuraca 2020, p.4). In the light of this papers research topic there are a rising number of studies that have made machine learning algorithms subject to their examinations and found that they are not only questionable in terms of abstraction but fail to make suitable interpretations of the interrelation between technical systems themselves and the social world around them (Kuziemski and Misuraca 2020, Selbst et al. 2019)

2.5 Bureaucratic AI Application

At the centre of public sector utilization of AI, the algorithms and their underlying strategy are designed to make decisions or assessments that are usually done by and at the level of every day or so-called street level bureaucrats. Building on this idea of publicly used AI on the street level and its implications, recent studies involve matters concerning the application of automated quality control in online spaces, algorithmic bias in the justice system as well as automated housing allocation for homeless people. In the one case the system was used to support the decisions made by the ruling judge and in the case of automated housing allocations it is obvious that the system was used to determine automatically who was in the most urgent need for accommodation and matched the housing opportunities with homeless applicants based on their eligibility criteria (Pääkkonen et al. 2020, Eubanks 2019, Fluegge et al. 2021). So as Fluegge at al. accentuate, "these street-level algorithms either supported or automated decisions usually made by street-level bureaucrats, are called street-level algorithms, as they perform the

tasks traditionally held by street-level bureaucrats, although these algorithms have also been applied outside traditional bureaucratic settings" (2021, p.6).

Knowing the area of application, the question that remains is what can be said about the repercussions on people's lives and their reality as a direct result of street level algorithms. The literature emphasizes on these issues, that more frequently concern private platforms, running powerful algorithms. The critique is that the privately running systems make decisions and assessments that have a major impact on people's lives but are on the one side not facing the same level of inquiry that related public used algorithms might face in order to avoid misdemeanour. In addition to regulatory disputes another concern of latest research has been the right to an "individual process", meaning the right of the citizen to have a free and fair trial in the courthouse just as a free and fair trial in a social welfare providing system at the job centre. The notion is that human cases are not always as simple as they might seem and have not only unique characteristics but a degree of novelty that cannot be encoded into most algorithms (Fluegge et al. 2021, Alkhatib & Bernstein 2019).

Developing an AI system that is supposed to be the perfect digital or computational twin to the works and persona of a street level bureaucrat result in a different perspective of bureaucracy and the public sector. Public services can often be characterized by their limited consensus in accordance to means and ends of decision making (Fluegge et al. 2021, Suchman et al. 1999). Therefore, it is quite custom to focus on context specific public services understanding when designing and developing suiting AI systems. The literature showed that a popular topic of discourse related to this, can be described as to how automated public services focus on the transformation of human responsibility (Fluegge et al 2021, Petersen et al. 2020). Coming back to the example of automated housing allocation, the research showed that the algorithms redistributed the discretionary control of uncertain locations because of the challenges associated to prediction or control outcomes of action with these places. In this case the user's responsibility is supported by the algorithmic assessment and decision making in these places (Pääkkonen et al. 2020).

Similar to this finding, the case of job placements and welfare services showed that responsibility of the caseworker is the relevant factor and only supported by the algorithms, as the humans are still the ones that make the decision on what is recorded and the outcome. This means that the literature emphasizes on the fundamental aspect, how AI should be supporting the work of humans (Fluegge et al. 2021, Petersen et al 2021). Other studies have found that the use of AI in automation and support systems can also be described as "co-bureaucrats". This means that instead of substituting the role of the human bureaucrat, the algorithm indicates the alteration of the human role from a

decision maker to a mediator of the decision (Fluegge et al. 2021, Whilborg et al. 2016). Linking the value of AI utilization to the direct impact and degree or level of responsibility can help public sector institutions in assessing its feasibility (Fluegge et al. 2021, Young et al. 2019). The result is that the algorithmic impact on the human level or responsibility has been subject to a rising number of studies and strengthened the academic or general public concern in regard to the use of AI in public sector organizations (Fluegge et al. 2021, Christin 2020).

The justification of using AI to enhance or provide certain public services is often argued on the basis of making public service more efficient and effective while simultaneously becoming less contingent and subject to judgements (Fluegge et al. 2021, Brayne and Christin 2020). Another argument is that algorithms ensure the fairness in traditionally obscure decision making within public organizations and its inflexible practises. Therefore, the result should be better decisions as well as mitigating the individual bias or prejudices of human bureaucrats. Many studies have been done in this fashion related to a wider range of public use cases including, the prediction of child harm, predictive policing, determining eligibility of social welfare support and automated decision making in asylum and integration systems showing that AI is already implemented in numerous different public services (Fluegge et al. 2021, Saxena et al. 2020, Brayne and Christine 2020, Eubanks 2018, Human Rights Watch 2020, Molnar 2019).

The main take away from the literature on AI and automation systems in the public sector including the social welfare services and the labour market domain is that the central collaboration between the user and the system is still quite delicate. As Fluegge at al. phrase it, “Caseworkers' decision-making is highly dependent on a variety of medical specialists, therapists, and the citizens themselves for documentation and to move the processes forward. As caseworkers rely on varying specialists, for example, in order to comply with the requirements for how a case has to be assembled and documented, makes casework highly unpredictable and thus hard to model. As we seek to make sense of data about casework and use AI for decision-support, understanding how collaboration takes place becomes even more critical. Thus, a core challenge for CSCSW-scholars is to empirically describe how collaborative work functions as a basis for the responsible development of AI systems” (2021, p.7).

2.6 AI Decision Making and Decision Support in the Public Sector

While the last subchapter showed that the literature and research has started to take notice of AI systems as possible subject of their studies, the majority of recent academic and private research interest is linked to the wider and more established use of commercial applications of AI (Sun & Medaglia 2018, Ransbotham et al. 2017). Especially the

impacts of AI to the fields of high tech and lucrative industries such as the automotive, financial, retail, education, telecommunication and travel sector have been thoroughly studied (Sun & Medaglia 2018, Chui, 2017, Dirican, 2015). Despite some recent development in the public sector and its utilization of advanced algorithms the number of related research and studies on its impact are still quite scarce. The public sector is not only starting to initiate its strategies on AI but has shown, as was mentioned before, numerous promising and diverse areas of use. The range seems ever growing with algorithms embraced by law enforcement, tax service and the health sector. The health sector can only be described as an extraordinary, exciting place for the use of machine learning systems. While all of these areas have shown some form of benefits as a direct result of their AI based analytics, the literature fails to go beyond measuring outcomes and efficiency alterations (Sun and Medaglia 2018, Mitchell et al. 2016, Goldmsmith & Crawford 2014, Nuance Communications 2016).

2.6.1 ADM

AI as automated decision making, or support tool has become a famous example in the innovation of the public sector. The idea is to conceptualize an algorithm as a recommendation system or simply the AI system depending on the situation, subject of decision and context of application. Therefore, these systems "…can take forms ranging from decision-support systems that make recommendations to human decision makers and/or nudge users of these systems in a certain direction, to fully automated decision-making processes that make decisions on behalf of institutions or organizations without human involvement. In this sense, human decision makers rely to varying degrees on automated decision-support systems when making decisions that either relate to themselves or to others" (Araujo et al. 2020, p.613). The level of human involvement differs depending on the system, leaving some user full autonomy over the decisions and recommendations while other users might be bound to at least consider the algorithmic expressed outcome. In addition to that, the system itself has shown to bring implicit expectations of future actions that might be influenced by the human user or the decision maker. The reason for this, lays within the feedback loop, that feeds the system with every decision made in order to continuously adjust the algorithm.

Other levels of human involvement can be expressed by the lack of involvement in almost completely automated decision-making processes where the results and outcomes of the decisions are communicated but do not involve any human user during the process of decision making itself. This means that "in the worst case, such a system will leave the subject of the decision in the dark both about the data used to in the decision, as well as how to contest the outcome, or even whether the subject of the decision had a choice in

whether it is participating or not in the process in the first place" (Araujo et al. 2020, p.613). What makes these systems so interesting to the research discussion is particularly the reason that public or academic focus has either shifted or is quite ambivalent between the benefits or the fear of misdemeanours.

The assumption of neutrality and objectivity is one of these interesting research objectives. The system itself is "...often created for the purposes that are often far from neutral: to create value and capital; to nudge behaviour and structure preferences in a certain way; and to identify, sort and classify people" (Kitchin 2017, p.18). However, the public sector is a unarguably different setting from such measures and while many private firms developing these algorithms claim that their system "…are carefully and strategically articulated as impartial and objective socio-technical actors in the discourse surrounding their implementation and usage in different aspects of daily life, as happens for example with search engines or other recommender systems" (Araujo et al. 2020, p.613, Gillespie 2014), the public sector might do good in ensuring exactly this.

2.7 Other AI Research

2.7.1 2.6.2 Public AI Policy

The literature showed that a greater amount of empirical research focusing on AI in the public sector during the last 20 years has been intrigued to focus on the trade-offs in relation with the AI integration and investigating the AI's capabilities to transform the workforce (See figure 3). Considering the unique nature of the challenges associated with public sector in contrast to the private sector it is noteworthy to highlight the lack of empirical studies on these concrete impacts that AI systems have and could have on the public sector. The false but sometimes very present assumption that AI is but just a code that follows normal if-then logic used by many other automation methods makes it hard for people to focus on the importance that proper research and its resulting implications on a human technology- world relation might have. The extension of a systems behaviour towards a capability of genuine learning is something that is ground-breaking incomprehensible to many of its users. It represents the principle of an ideal technology that can be applied to almost any public sector context, with constant or constantly changing environmental settings where pre-programmed automation may fail to adapt (Sun and Medaglia 2018, Russell & Norvig, 2016).

Social and ethical issues are increasingly focused on and the important achievement of an appropriate balance between privacy and data gathering is one of the most pressing topics to study in AI adoption. This comes of course with the introduction of regulatory issues and the question of supervision (Begg 2009, Hengstler et al. 2016, Gulson & Webb 2017,

Sun & Medaglia 2018). A more human individual centric research branch could therefore help answer concurring problems related to the former studies and the ongoing automation of decision-making processes within the public sector, but also what these mean to the mediated world of the users and the ethical implications that might arise to all involved stakeholders.

The work of other authors that have spent resources in order to investigate the numerous problems of AI when confronted with terms like goals, drivers, barriers and risks included a wide variety of AI in public sector case studies with the goal to select no single but multiple cases that can guarantee a more thematic as well as a geographic diversity of literature on the topic of automation systems. Kuzuemksi and Miscura have used this work as a fundament for summarizing context specific and important factors that have been derived from the examination of the specific systems (2020). In their case study analysis, they have focussed on the overview regarding the regulatory approaches that have been done in AI across different jurisdictions in the public sector. A key finding suggests that even at a basic level, the use of AI in the public sector could be an easy opportunity of using the technology as an instrument of control and exercising it over the citizens (Kuziemski and Misuraca 2020).

The findings and literature have also indicated that the power factor has been a central interest of most considerations. The assumption is that by adopting systems based on automation, the organization increases its control over the physical place of citizens, their resources and most importantly their information. The role that research, academia and other civil or non-profit organizations have, and therefore put upon themselves is that of a an observer who is responsible for scrutinizing the use of automation systems. And while the studies tend to concentrate on the goal setting, procurement, feasibility and implementation, the aspect of AI policymaking is still strongly affected by the government (Kuziemski and Misuraca 2020).

Bringing us to the very interesting and lively debate on AI policy that as was mentioned before has been a central point of considerations when thinking of AI implications and effects. But an interesting aspect concerning the idea of a self-governance in association with AI systems brings its own serious considerations to the power relations in place and their disregard in this debate. More recent attempts to overview this topic have led to a meta-analysis about the ethical guidelines. The result of this was defined by a its reluctant approach towards enforcement mechanisms, lack of clarity to which norms should be prioritized, and a significant gap between agenda setting and its implementation (Kuziemski and Misuraca 2020, Jobin et al. 2019). Other authors and their works have somehow displayed different outcomes that have a more hopeful tone and support the

potential of AI that includes professional norms together with regulations and governance (Kuziemski and Misuraca 2020, Gasser & Schmitt 2019, Veale and Brass 2019). The notion seems to be a critical observation in order to discern the theories and sentiments of a modern digital world. This can be done by underlying everything to the simple goal of creating public value or by rethinking the established relations with technology and government. This process itself is inherently political but politics do not always care about the single citizens reality (Kuziemski and Misuraca 2020, Mazzucato, 2016; Misuraca, Geppert and Codagnone, 2017). Therefore, when thinking about implementing AI and making it of use especially in the public sector, one important question is what is the aim that it is optimizing for? As Kuziemski and Misuraca write "no proper guidance for the public sector use of automated decision systems can fail to imagine the states of the world it envisions, and the values that it wants to support" (Kuziemski and Misuraca 2020 p.10).

2.7.2 2.6.3 AI Experiences

The emphasis of the current literature on AI in the public sectors has therefore begun to shift towards the considerations that include the directions of the fundamental development of the whole field and examine the existing and established practises from different settings with governments, municipalities, and public agencies (see image 3). But what are the goals and desired outcomes of automation tools? Who benefits from this commission and which or who's benefits should be prioritized? And what can we learn from this when conducting this study on the use of such an automated decision support tool from a post phenomenological perspective? The literature shows that research is having these important conversations about the ends and means. The complemented effect of this discussion is that besides the usual benefits, trade-offs are being acknowledged as well as the starting point and the desired states of our reality that are influenced by the intervention of an AI system integration (Kuziemski and Misuraca 2020).

The wide range of possible use cases for AI and automation technology, and within different setups and contexts, has led research to try to recreate specific scenarios that users might encounter. This means that users or participants in the study were given the task of describing the scenario in textual form as well as drawing their interactions in visual interpretations. The next step included a brainstorming process that would help filter the information that users might want to see when using the AI system. These kinds of studies show a certain degree of observational analysis and anecdote story-saying that will further on be important for this papers' research. The results showed that a more social perspective on certain design features is indeed a central aspect when it comes to the user's realization and interpretation of the technology. The suggestions that were

made included a wish to know more about what happens to other people getting recommendations from the AI systems or who else got these recommendations. These findings indicate that users are likely to show a form of experience sharing with other colleagues or users as a matter of information gathering in order to better understand their own "use". Ehsan et al write in their work that the "ideas converged to what their participants coined as the "4W"—who did what with the AI system, when, and why they did what they did— in order to have adequate socio-organizational context around the AI-mediated decisions" (Ehsan et al. 2021, p.8-13). They noted that an interesting observation related to the four questions share similarity with the design elements of work done in the field of group awareness and groupware work (Ehsan et al. 2021, Gutwin and Greenberg 2002). Only the question that focus on the "why" are rather explanatory. Therefore, when considering how to represent the "why", participants in their study suggested an open-ended textual representation to capture the nuances behind the decisions (Ehsan et al. 2021).

The experience a user might have with an AI system is not yet at the core of AI related research, but the findings suggest that this approach might bring important ideas that extend the previous realization of solely relying on technical or algorithmic transparency as a factor to understand and strengthen the complex process of decision making and automation. And despite convincing arguments for the benefits of AI, some consensus about limitation exist that help users understand that algorithms are not yet able to take into account the entire contextual factors that might be important or a matter for a specific decision. "Not everything that you need to actually make the right decision for the client and the company is found in the data" (Ehsan et al. 2021, p.8-13). The research participants in this exemplary case did point to "the fact that even with an accurate and algorithmically sound recommendation, there are things [they] never expect a machine to know [such as] clients' allegiances or internal projects impacting budget behaviour" (Ehsan et al. 2021, p.8-13). Therefore, it can every so often be that "the context of social dynamics that an algorithm is unable to capture is the key and that the real life is more than numbers, especially when you think of relationships" (Ehsan et al. 202,1 p.8-13).

The previously mentioned social transparency concept can be extended to the context of decision making. This means that the local context of past decisions is made available creating a form of knowledge between different users that is shared among them. With the help of such crew knowledge the process of decision making itself and the use of the AI system is the centre of a more social relation. The aim of this approach however can be reduced to the surge for strengthening or improving decision making capabilities. And while this might indeed happen, what is being observed in this sense by the literature is that a crew knowledge might lead to the single individual users boost of decision

confidence and other actions that are based on a sense of social validation. This means that the user's facilitation towards doing his or her work might also alter the decision making of the AI and making it more resilient of contestability depending on the perspective of the observer. Ehsan et al. (2021) gained this knowledge during different cases across various domains and wrote the following on this finding: "When asked to elaborate on the concept, participants defined it as informal knowledge acquired over time through hands-on experience, knowledge that is not typically gained through formal means, but knowledge that's essential to do the job. Crew knowledge is learned via informal means, mainly through colleague interactions. It can encode idiosyncrasies like client specific quirks. Participants referred to their team as their "crew", with a sense of identity and belonging to a community membership. We can think of crew knowledge as informal or tacit knowledge that is acquired over time and locally-situated in a tight-knit community of practice–an aggregated set of "know-hows" of sorts. While ST features may not explicitly encode a complete set of crew knowledge, they provide in-situ access to the vital context of past decisions that carry elements of crew knowledge" (Ehsan et al 2021, p.8-13).

2.7.3 2.6.4 Recent development of AI on the Labour market

Other findings of AI use in public sector organizations indicated a characterization of the user and his collaboration towards practices that include documentation. The documentation has been observed to be part of discussing the distribution of support and benefits in this specific case concerning job placements. Therefore, a collaborative concept of street level decision making may deeply influence the underlying principles for designing AI systems valuable to the public sector and its services. Valuable in the perceptions made by users includes AI support for management issues. This means that when a specific problem occurs that is related to the case it requires more resources because of complexity. The case has shown that the support of the system has become a starting point for the collective practice in relation to the individual's case when deciding on and assigning a specific job placement program (Ehsan et al. 2021).

The argument often mentioned in the literature is that AI can strengthen their case by means of management. Other cases also related to labour measures express, that with the help of more advice, job seekers as vulnerable citizens could profit from better suited decisions. However, the underlying point that these studies show, is that the caseworker and user of the automation tool heavily relies on the system to help and strengthen his confidence in making decisions which could not only alter their work habits but their underlying assumptions about the use of such technologies for other domains. Many people that were interviewed during this previous research have had a strong feeling about

the value of information but also about the support that the systems gave them when interacting with citizens and clients (Fluegge at al. 2021). As the study wrote "caseworkers' perception of the value of AI was only to some degree determined by the level of discretion; it was seen as more valuable if the AI component could support their decision mandate towards management arguing for collecting medical documentation although it is costly, and citizens arguing for a specific welfare program or internship" (Fluegge at al. 2021, p.12).

The shift of focus from outputs and impacts towards the social and sociotechnical perspective on any human or organizational input has shown that the actual users of such systems are often more comfortable with systems that do not completely automate their decisions or work but are preferring tools that use algorithmic decision-support. The result of this notion is that on the one hand using AI for simple tasks to begin with and on the other hand leaves places and especially cases of uncertainty to the humas responsibility is a likewise considerable approach and opinion in the literature on AI. Fluegge et al. write "…it is important to consider the things that can make simple decisions complex. In our context, is collecting or not collecting medical documentation a decision meant to retrieve information more quickly, to assess the case on enlightened grounds, or to maintain a trustful relationship with the citizen? All of these can be at stake, and something the caseworker reflects upon before deciding. This is a challenge facing the design of AI systems for public services, and perhaps a solution to this could be to remove AI from the moment of decision-making" (2021, p.19).

The relation between bureaucratic organized workspace and the use of algorithms has proved to be a useful lens for studies that analyzes AI in public sector organizations. And although the past research has expressed to some extend the importance of algorithms and automation as players that impact the lives of actual people in different ways, the focus often returned to the matter of who is running these systems and that these algorithms may not have or at least have a different degree of democratic accountability as well as legal demands of equal treatment or transparency. Therefore, the argument that was found quite frequently across the selected current public sector AI research is to avoid the theoretical disconnection and rather focus on analysing or explaining algorithms or AI in public services in terms of actual bureaucracies (Fluegge et al. 2021, Weber 2012).

While the literature is to no surprise heavily concentrated towards research on the impact of automation or AI and private sector, the findings are still relevant to understand socio technical and implications to individuals' behaviours and perceptions of AI and the world. For example, does the accumulation of such technologies reshape labour force participation rates and therefore also the individuals' attachment to the workforce in

typically advanced economies? This means that the notion that often comes with AI analysis, is leveraging the variation in the routinizability of occupations and occupational composition. That's why the discussion on this topic points to significant negative effects of automation on the participation rates of the private sector and the possible emulating public sector. Some more individual examination may confirm that bureaucrats often have similar routinizable activities as previously private employed workers that dropped out of the workforce for the very reason of automation (Grigoli et al. 2020).

A recent discussion paper by Artan Loxha and Matteo Morgandi explores and compares employment seeking people and profiling methods adopted by key public employment services of the Organization of Economic Co-operation and Development (OECD) member countries. The study includes advanced profiling methods but also have more applications outside the scope of this study, including job matching and labour market monitoring that are related to the systems examined in this thesis. The paper writes the following about the case of automation, "the tools…are part of the toolkit to assist customers in finding an appropriate exit from unemployment. While profiling tools can also assist with monitoring local labour market dynamics. The generation of soft and hard data on job seekers from profiling could also enrich labour market statistics, especially at the local level, and can help monitor its changing dynamics over time" (Loxha and Morgandi 2014, p.7). And while this study showed that self-service platforms are becoming mainstream in public sector institutions the concurrent issues and challenges are often only referred to the transparent handling of such automated self-service classifications and the need to be ensured in alignment with a country's data privacy regulations (Loxha and Morgandi 2014).

Other authors have extended the idea of classical AI research and tried to explore the implication that such technologies have to social services related fields like unemployment duration or the subsequent job quality. The result show that occupational risks can significantly reduce the re-employment probability of unemployed workers in addition to those problems deriving from pronounced gender differences, rising ethical consideration to the use of such systems that involve humans as subject to the decision of the AI (Schmidpeter and Rudolf 2018). This problem frames the importance of this studies main goal of identifying the post phenomenological technological human world relations and how they might have shifted with the integration of the now thoroughly explored implications of AI based automated decision-making tools. The review was able to elaborate on an interesting evolution of social, philosophical and ethical aspects of AI in the public sector as well as building a foundation of knowledge that will help interpret the results of this study.

3 Methodology and Theoretical Framework

This chapter begins with a general overview of the research methods used in this study. Meaning a discussion on the qualitative methods and fundamentals of a more philosophical approach on technology studies. Based on this a theoretical framework for this study will be outlined. The theoretical framework of post phenomenology and its defining features, concepts, origins, and evolution towards a method in the field of philosophy of technology is explained, while also mentioning current applications in relatable research. Concluding this a detailed description of the post phenomenological approach as a methodology for data gathering and analysis is provided.

3.1 Introduction

This subchapter starts with distinguishing the primary methodological approaches of conducting research in the field of technology research which are qualitative, quantitative and mixed methods research. Starting off with qualitative research, this approach manages to focus on the examination of human attributions of importance. Another way to put it would be to say that this method studies subjects in their natural setting while attempting to make sense of or interpret certain phenomena in regard of the meaning the human subjects manifest to them (Denzin and Lincoln 2011). The close link between this interpretation and the philosophical orientation towards constructivism, advocacy or participation could reason that qualitative research more often than not involves the act of exposing the interpretations or meaning accredited to actions by the research participants and the textual interpretation of human action descriptions, practices or cultural contributions (Creswell 2009, Willig 2013, Yuen 2021).

As mentioned, the main qualitative research data collection strategies include observations, text data or media data as well as interviews. Classification of common quantitative research study examples and the defining boundaries of a data collection strategy are phenomenology, grounded theory, ethnography, case studies and narratives. Therefore, the typical qualitative research process evolves over open questions (Creswell 2009, Yuen 2021). By definition quantitative research methods in contrast manage to focus on the utilization of predefined and measurable variables such as numerical data. This data helps the researcher to draw possible correlations and causal relationships. The research philosophical emphasis on this approach is typically considered as post-positivist and has closed questions regarding a present hypothesis that is being tested. The data collection strategy of quantitative research focuses on surveys and experimental design. This means that the main concern of these strategies is the identification of variables, their relationships and measurables as well as statistical approaches to guarantee important

factors of quantitative research such as validity and reliability (Creswell, 2009, Yuen 2021).

The last mentionable approach is a mixed method, where the researcher systematically combines the two previously explained techniques of qualitative and quantitative research in order to support a more pragmatic, well-rounded study on the selected field and research questions. Based on these three short introductions, this study focusses on a qualitative approach and while the other two methods are considered as rigorous, the decision was determined by the outlined research questions of this paper. This goes hand in hand with the nature of the research questions and the frequently practiced use of qualitative methods in the post phenomenological literature (Creswell 2009, Yuen 2021).

3.2 Conducting Qualitative Research

As a matter of this paper, the defining characteristics of qualitative research are as follows:

1. “Collecting data in a natural setting or context where participants would experience the issue or topic-at-hand

2. The researcher directly examining and gathering data, rather than rely on instruments developed by others

3. Using multiple sources of data such as interviews, observations, and documents to identify common themes

4. Inductive reasoning and data analysis to make sense of abstract or emerging data by organizing them into themes or categories, and even collaborating with participants to shape ideas

5. Emergent design and questions

6. Interpretative inquiry of multiple perspectives using a specific theoretical lens involving social, political, or historical contexts

7. Holistic account of the complexity of the issue or topic” (Creswell 2009, Yuen 2021, p.33)

This study therefore tried to consider these qualities while conducting the qualitative research and answering this works research questions. The overarching goal of this research is to investigate the experience of public sector users when handling services

with the help of AI technology and the insights about using it. Reciting the research questions, they were formulated as follows:

- Q1: How do users at the Estonia Unemployment Insurance Fund (EUIF) apply artificial intelligence to facilitate their work and offer social services to citizens?
- Q2: How do the users at the EUIF perceive the influence of Artificial Intelligence on their working practices and interactions with citizens or colleagues?
- Q3: What are the ethical considerations of AI in providing social welfare services for employment seeking citizens?

For a more in-depth exploration of these questions this study utilized a post phenomenological approach. Post phenomenology is a certain philosophy of technology with its roots in phenomenology and is dedicated to study and understand human-technology relations. As formerly mentioned, this approach is considered to be part of qualitative research and has been used across various in-practise settings and contexts (Aagaard 2017).

The underlying thought of this choice has been the ability of post phenomenology to explore very different experiences and use of a certain technology while expanding the borders of a technology's affordances and its understanding across multiple and diverse users and contexts of use. This means a study of technology which focusses on examining the implications of an intertwining relationship between a technology e.g., the technological effects on the user and the other way around in addition to the surrounding "world" context it is utilized in (Yuen 2021). As this study is exploring the diverse ethical and social insights of AI technology used in social services provided by a public organization, the examination of a public official experiences is a central element of this. Therefore, this research applies post phenomenology as the theoretical lens for the conducted case study of Estonian public officials using automated decision-making software in the Estonian Unemployment Insurance fund counselling services for unemployed citizens.

3.3 Post phenomenology

This paper already stated that post phenomenology is a novel approach to the social and cultural roles of technology. Meaning that studies in this field have been investigating a technology in terms of their relations between human beings and the technological artifact itself. Henceforth it is done by focussing on the multiple ways a certain technology helps to shape the relations between the human being and the world around him. Since we categorized this approach as part of philosophy of technology, it is worthy to mention that

this does not mean this research and other post phenomenological research tries to apply philosophical theories on the technology itself but rather build on top of a technology or technological development as a starting point and then perform a philosophical analysis (Rosenberg and Verbeek 2015).

3.4 Intentionality

Breaking down the key terminology of post phenomenology, the concept of intentionality has derived from phenomenology and describes the scepticisms towards classical philosophical studies on consciousness in addition to the mental representation of a certain object. As academia tried to take up intentionality into the post phenomenological approach, it has been defined as the fundamental direction of a consciousness structure, meaning that the complete consciousness is permanently directed at something real or imaginary (Aagaard et al. 2018, Adams and Thompson 2016, Yuen 2021).

At the start of intentionality, the literature described a characteristic of mental representation with a particular direction on an object. Expanding on this idea, philosophers have thought of depicting additional existential notions as part of intentionality. This means that in addition to the basic concept, an examination of the intertwining relations and connotative meanings that humans can attach to their world is included (Adams and Thompson, 2016, Yuen 2021). Philosopher Heidegger introduced the phrase of "Dasein" meaning that we as humans are beings in the world, and the similar important researcher in the field of phenomenology and therefore post phenomenology, Ihde has built on top of this and presented an idea of the concept of intentionality which focuses to express the intertwining relations as a reciprocal connection of the human world. And as a matter of poetic beauty this mentioned human world can be described as the "grand theatre of objects arranged in space and time as perceived by individual subjects and can be considered the horizon or background for shared human experience and meaning" (Husserl 1970, Yuen 2021, p.36).

In this studies introduction the dynamic characteristic of our world has already been shortly mentioned and building on this thought the literature argues that the lifeworld that has just been described needs to be considered as such a dynamic universe where we as human beings can live through as well as live with. Therefore, the human world relation can be separated into a "human to world" role that is a natural tendency of how humans relate to the world and a "world to human" role where the world influences the human and how they perceive and conduct themselves. What Ihde did to this understanding of intentionality was to add technology and therefore the question how technology mediates a humans perceptual or likewise actionable relation to the world around him. We can therefore assume that post phenomenology originated with this introduction of Ihde and

that the main emphasis of this approach is in “describing and analysing how different technologies condition the human lifeworld and understanding the complexity of human-technology-world relations” (Yuen 2021, p.36).

3.5 Human technology Relations

Breaking down on the previous subchapter and one of Idhes most significant contributions to this uptake on philosophy of technology, a closer look on the different forms that human technology relations may take can be worthwhile. It has been proven quite useful for researchers to examine these various ways a human being can establish for instance bodily perceptual relationships with technology. This concept has shown a practical value for case studies of user relations to technologies, which implies that the categorization of the different ways of technological mediations help describe concrete relations of users of technologies. Another important contribution has been the demonstration of variability and context specific user experiences itself, by arguing against the oversimplification of a one-for-all examination of user experiences. However, this does not mean that the list of different “human-technology relations” forms this study introduces is or claims to be in anyway exhaustive but rather help to articulate various ways in which users develop these relationships with technology they use (Rosenberger and Verbeek 2015).

3.5.1 Embodiment Relations

The idea behind these relations is as mentioned before to help articulate some of the ways that users develop bodily or perceptual relationships with the technology they use. Starting with the Embodiment relation, Idhe tried to explain the mediation with which a certain technology might transform the users actional and perceptual engagement with the world. This means that the so defined “embodied” technology changes the users experience through its device, while the device itself is also considered by the user’s bodily awareness. This relation is represented by the following diagrams where when the technological mediation that occurs between a user and the world around him can be characterized as follows (Rosenberger and Verbeek 2015),

Human – Technology – World

while the embodied relation to a technology may be characterized in the following sense:

(I – Technology) → World

Rosenberger and Verbeek (2015) also provide a simple and understandable example for this kind of relation, with the eyeglasses. They glasses can be worn but they are not always perceived by the person who wears them as they might perceive the many other things around them. The user rather looks through them into a world that has been more or less transformed. So, in this sense the glasses can be conceived as part of the user's perceptual experience.

Another aspect that comes along with this, is the notion of transparency that a certain human-technology relation may have. This means that the technology itself has some degree to which its device or a particular aspect of that device "fades into the background of a user's awareness as it is used" (Rosenberger and Verbeek 2015, p.14). This means that as a result of a user increasingly getting accustomed to the embodiment of the device while simultaneously certain habits of bodily and or perceptual actions develop in familiar day to day use of the technology, the device itself may take on a degree of transparency (Rosenberger and Verbeek 2015).

Human-technology relations in general and especially the embodiment relation have a structure that allows a certain magnification or reduction. Idhe himself wrote: "Embodiment relations display an essential magnification/reduction structure which has been suggested in the instrumentation examples. Embodiment relations simultaneously magnify or amplify and reduce or place aside what is experienced through them" (1990, p.76). In this context the transformation that is rendered by the mediation of a certain technology towards the user's experience are not neutral. The user does not only perceive and or receive the desired changes of their abilities in mind but will rather also encompass other changes "…with some of them taking on the quality of trade-offs, a decrease of a sense, or area of focus, or layer of context" (Rosenberger and Verbeek 2015, p.16). There is a distinction between the micro perception and the macro perception, where micro perception is focused on the individual and his own bodily sensations while the macro perceptions concerns are the cultural, historical, and anthropological dimensions of the explored experiences. However, in the original concepts of the human technology relations it is argued that the two perspectives are but inextricable and therefore not looked upon separately (Rosenberger and Verbeek 2015).

3.5.2 Hermeneutic Relations

The second relation of the human technology mediation can be identified as the hermeneutic relation. Hermeneutics is a philosophical term that relates to the idea of contemplations between the nature of language interpretation and the translation. In the context of this methodology the "notion of hermeneutic relations refers to technologies which are used through an act of perceiving and interpreting the device's readout"

(Rosenberger and Verbeek 2015, p.17). This means that in contrast to the embodiment relation the user does not experience the world through the device but rather experiences a changed and transformed encounter with the surrounding world though the immediate experience and interpretation of the technology itself. This is shown by the following representation:

$$I \rightarrow (Technology - World)$$

Rosenberger and Verbeek (2015) present a further example to clarify the underlying thought behind this relation. While looking at a wristwatch, the user interprets the hands or the digital display of the device and as a result of this hermeneutic relation, the experienced perception of time has changed and or transformed the access to time itself. Applying the case of hermeneutic relations to technology, involves possible similarities between a user reading the displayed interface of the mediating technology and him reading written language. Rosenberger and Verbeek write, "a key aspect of one's encounter with a written language is the degree to which one "knows" it, that is, the degree to which one understands how to read that particular language. If a person has no experience with that language, words written in it will convey no meaning at all. But if she or he instead is deeply fluent, then the meanings of words jump from the page almost automatically. For the person deeply familiar with a language, she or he does not need to slowly and actively interpret each letter, then put them together to make a word, and then interpret the meaning of that word; the meaning emerges all at once in a perceptual gestalt. This is also true for hermeneutic relations to technology" (Rosenberger and Verbeek 2015, p.17). This means in regard to the wristwatch example the transformed time perception only occurs in the case where the user already knows how to tell the time. And while many standard examples of hermeneutic relations closely relate to certain technologies with visual representation, it can just as well apply to further devices stimulating different senses.

3.5.3 Alterity Relations

The third concept of the human technology relations are the alterity relations. They refer "to devices to which we relate in a manner somewhat similar to how we interact with other human beings. The idea is that some forms of interface are devised specifically to mimic the shape of a person-to-person interaction, and that sometimes we encounter a device as itself a presence with which we must interrelate" (Rosenberger and Verbeek 2015, p.18). In this sense the focus lays on the experience of engaging with others and that an interaction through and with technologies may take over the role of that other part in the interaction. This relation is represented as follows:

$$I \rightarrow Technology – (- World)$$

Technology as an externalized other part of an interaction is in its most common form any kind of computer interface scheme where the user is under pose of any direct questions. However, they can also be observed in cases where the technology acts not in line with its customs expectation but to the contrary of it (Yuen 2021). The example that Rosenberger and Verbeek (2015) give is a display and more specifically an ATM machine that asks the user questions regarding his withdrawal and other functions. And while we interact with these devices the relation does not really claim that the user mistakes the technology for an actual human, but that with the interface the modes take some analogous from. “It could be predicted that as computing advances, and as our abilities to create sophisticated computer programs that simulate human interactive style increase, we will see more and more devices designed with an alterity-style interface. It is already the case that this form of relation can be increasingly seen in automated interactive customer service phone calls, GPS devices that read aloud driving directions, and voice interactive personal assistant smartphone applications. Accordingly, post phenomenologists continue to develop this idea” (Rosenberger and Verbeek 2015, p.18, Irwin 2006; Bottenberg 2015, Wellner 2014).

3.5.4 Background Relations

The last notion is considered the background relation which is primarily focused on technologies that create a specific environmental context to the user. The background relation is commonly established with devices that the user may not be using directly but instead has some form of interaction with it that shapes his or her experiential surroundings. Possible examples would be a refrigerator as well as an air conditioning unit or the heating system which are constantly running in the “background”. The device may not be as present to the user’s perception, but the argument is that it nevertheless becomes part of the experience and immediate environment of the user. “Technologies to which we share background relations stand back in our awareness not simply because we have grown accustomed to their usage, but because they quite literally form the backdrop of our experiences. They shape our experiences, protecting us from the elements or keeping our food safely chilled, but do so in ways that do not require direct interaction” (Rosenberger and Verbeek 2015, p.19). The relation can be expressed as follows:

$$Human \rightarrow (Technology/World)$$

This means that human perceptions and their actions in the world are not only constantly influenced by many technologies and functions that are happening in the background, but

they are also doing this without the immediate awareness of the user's consciousness (Yuen 2021).

3.6 Post phenomenology as a Methodology

After examining the theoretical framework of post phenomenology, it can also be utilized as a methodology itself and as mentioned before a part of academia has started to apply this approach when investigating science and technology. While philosophy is not an empirical field, this approach realizes an empirical orientation as a significant part of doing philosophy. This means that post phenomenological studies use this orientation as a concrete starting point for philosophical reflection rather than a positivist basis of the philosophical knowledge (Achterhuis 2001, Rosenberger and Verbeek 2015).

However, this empirical-philosophical approach does not follow a strict post phenomenological methodology since the studies are diverse and too context-sensitive for that. Appling this methodology rather focusses on understanding the role that technologies play in the relations between humans and the world, and on analysing the implications of these roles. The literature argues that it is the relations which are the basis of the impact technology has on human beings, its culture and society. Therefore, "this focus on human-technology relations implies that post phenomenological studies always include empirical work as a basis for studies conducting philosophical reflections" (Rosenberger & Verbeek 2015, p.31).

An empirical orientation is required in order to understand human-technology relations and the roles the technology embodies in the human beings' experiences and practises. A possible approach on doing this is by developing the study on the basis of literature and empirical work done by other researchers, from self-conducted studies, or from an analysis of a person's experience involved in the technology of interest. And instead of aiming for an accurate description of a certain technology, post phenomenology emphasizes on investigating the character of the various dimensions of the relations between humans and these technologies, and their impact on human practices and experiences (Rosenberger and Verbeek 2015, p.31). Typically, the study will explore how a certain "world" is constituted around the relations between technology and humans in addition to a specific subject that can also derive from it. Based on these findings, "post phenomenological studies typically make a conceptual analysis of the implications of technologies for one or more specific dimensions of human-world relations—which can be epistemological, political, aesthetic, ethical, metaphysical, et cetera. The central question then is how technologies help to shape knowledge, politics, aesthetic judgments, normative ideas, religious experiences, et cetera" (Rosenberger and Verbeek 2015, p.31).

The next few subchapters provide an overview of the post phenomenological approach generating data and its analysis and the case study design.

3.6.1 Case Study Design

Case studies are an essential element of the just introduced post phenomenological perspective of concrete human-technology relations and technologies. They do not only instantiate the concepts and commitments of the underlying framework, but they also act as laboratories within which the basic post phenomenological ideas are questioned and redefined. This means that if post phenomenology provides practical contributions to the contemporary interdisciplinary discussions on design, scientific research, and technology policies, then case studies are a core driver for displaying this value (Rosenberger and Verbeek 2015).

The case study examines instances of individuals using the technology during a certain process, activity or event and the various variations of its use. This means qualitative data is gathered through observation and interviews and the researcher then creates a detailed reconstruction of the technological phenomenon based on the collected data. It could therefore be useful to identify preliminary or recurrent matters of interest. In order to analyse the data, the literature on post phenomenology has retained appropriate concepts to investigate the multivariant ways for example that technology can be of use across various different users and context (Yuen 2021). These concepts will be explained later.

This study proposed the qualitative, post phenomenological approach in order to frame the prior outlined research questions and the explained methodological approach of post phenomenology. Focussing on the specific technology, in this case the automated decision-making software of the Estonian Unemployment Insurance Fund called OTT and how the public servants are utilizing the application, this research considers this to be a suited case study design. The data itself will be collected through workplace observations and interviews with the users, while data analysis will employ the post phenomenological concepts of technological intentionality, multi-stability and human-world relations to examine the public servants experience of using decision making software (Yuen 2021).

3.6.2 Data Collection

The data collection of a post phenomenological approach involves as mentioned before, gathering pre-reflective data from multiple sources as well as anecdotal information that can be generated by either doing self-observation, the observation of others or interviews. However, in accordance with the specific research topic the data collection can also

include surveys or conducting field research in addition to the numerous different analytical approaches that could help understand the gathered information. These approaches, as mentioned before are focussed on employing the concepts of multi-stability, variation and mediation and therefore combine the theory with empirical investigation (Yuen 2021).

Post-phenomenological data collection is manly incorporated by the two main methods of interviews and observations concerning the user and his use of the observed technology. The data is therefore created through a direct examination of live experiences and gathered through multiple sources by applying typical qualitative questions that are exploratory and open-ended. This means that the research is employing an interpretative inquiry research philosophy and lens while examining and gathering the data (Yuen 2021). The common sources of pre-reflective data for the post-phenomenological approach are embodied by field observations and interviews in addition to sources that may also include the gathering of materials from online databases, historical documents, journals, policy reports, and artworks. The collected data can then be reassembled and help create posthuman anecdotes for data analysis (Yuen 2021, Adams and Thompson 2017). The importance of recording the related experiences regarding the use of certain technologies but also towards other humans and nonhumans may help gather the relevant data (Yuen 2021).

Interviewing the user as a human being about their interaction with the technology of interest as part of the research means understanding the user's attendance to objects and the acclimatisation to certain things. This form of object interviewing can be done by the application of a set of heuristics that include gathering so called posthuman anecdotes as well as "listening for the invitational quality of things" (Adams and Thompson 2016, p.46). This means the reason behind these posthuman object interviews is gathering and reassembling anecdotal material that could be of benefit when trying to understand the phenomenon that is being looked upon. In addition, this approach can deliver crucial data for further analysis (Yuen 2021).

Based on the fact that this process embodies gathering observational data as well as creating new interview data in order to establish these anecdotes, a question that derives is, what an Anecdote actually is. Adams and Thompson (2016) define them as little stories that are a part of any ordinary conversation. They write, "the subject of the observation or interview reencounters lived through details of incidents or experiences. An anecdote must reassemble and resemble a possible human experience or observed moment of everyday life. Acknowledging the fictive status of anecdotes is also an admission that all research texts involve a kind of creative reconstruction, selective cuts, and even poetic

invention. To describe the lived-through moments of our lives, or to render an observed event in words, we must necessarily rely on the taken-for-granted fundament and limits of human language to convey what we experience and notice. Too, much of the world as we experience it cannot be described, and so must fall silently between the lines. All such descriptive texts are necessarily attempts to recapture and render intelligible what appeared to us originally and pre-reflectively before language. The anecdote, like all texts, is a fabrication" (Adams and Thompson 2016, p.25).

3.7 Data Collection Methods in this Study

As part of this study, the gathering of pre-reflective materials concerning the use of automated decision-making tools for evaluation and categorization employment seeking people in Estonia, happened in the form of an online background search. The focus lay on the promoted features as well as any information given to the time of use and the instructions of the technology itself. The conducted material was added by additional information from a presentation slide, where the system was explained in its more technical components, overarching functions and goals. Based on the gathered information a clear list with the intended features of the system was generated and could therefore be of use when creating the interview questions and doing the observations. For example, could the list of features help understand how many and which features the user uses or is aware off, while maybe also observing some background or unintended use of certain functions.

The main methods of data collection were however related to the interviews that were conducted with the users of the automated decision-making system. The users are defined as the career consultants of EUIF that also directly interact with citizens and the IT administrators working at the EUIF who are mainly interacting with the back office. The consultants as well as the IT administrators are both direct users of the OTT decision support system that has been chosen in this case study. Both groups of people will therefore be used as the purposeful sample representation of this thesis. Due to the still ongoing pandemic situation in early 2022 observations have not been done. However, the meetings with the users have been done online and latest about one hour each. In these online meetings the pre-interview materials were discussed in addition to the semi structured interview done one-to-one. The open-ended questions were chosen in order to attempt drawing out the users own reflections and recollections concerning the technology and what uptakes or thoughts they may have had on using the system in their workplace. Considering a better understanding of the nature of these question the following example questions are given (see Appendix):

- What was it like when you first worked with the OTT tool?

- Can I be held responsible for the actions which this technology empowers? Would I feel better if I couldn't?

- What knowledge has the use of this technology disclosed to me about others? Is it good to have this knowledge?

An important part of this approach is recording and transcribing the interview in order to later analyse and keep records of the result. This was done with the consent by each interviewee. The researchers' notes were attached to the additional information that was gathered in the preliminary search and the interview of the user using the automated decision-making system.

3.8 Data Analysis

The post phenomenological process of data analysis mainly focusses on reflecting the data collected with the before mentioned post-phenomenologically concepts. This also includes the study of certain breakdown moments or any form of accidents and anomalies in addition to the common dissection of the human-technology-world relations that might help unravel and trace the users' materials, responses, and passages. With this post phenomenological approach, the research establishes a foundation for philosophical technology related empirical research of the automated decision-making system (Yuen 2021, Adams et al. 2018).

The before mentioned concepts of post-phenomenological analysis is characterized as a philosophy of technology framework with the purpose to go beyond the examination of the technology's affordances. It does this by exploring its multi-variation in regard to different contexts of application. Therefore, there may exist no one particular approach to the analysis of the gathered data. However, this study used the workings of the human-technology-world relations associated key concepts, variations and mediations, to achieve the analytical process (Yuen 2021).

- **Variations**: In the beginning of philosophy of technology research, Variations were introduced as brainstorming techniques that would allow the researcher to recognize essential elements of a phenomenon. This means a separation between variant structures and the invariant components as well as the essence of the object by applying various or multiple viewpoints and angles. This initial idea was later then adapted by the post phenomenological concept of multi-stability, that focusses on the technology contexts specific dependences as well as material relations for it.

- **Mediations**: Mediations served the purpose of describing the relations between humans and their world not only by examining the way technology shapes these relations but actively looking upon them in regard to specific features that are being amplified and/or facilitated while simultaneously other features in the contrary might be reduced and/or inhibited to the user's experience (Ihde 1990, p.75).

3.9 Multi-stability

"How should we understand the way that technology at once in part determines our choices and actions, and yet at the same time itself remains open to our manipulation and interpretation? How is technology both something we design and use for our own purposes, and also something that influences, restricts, leads, inclines, or controls us? One of post phenomenology's main responses to this question can be found in the notion of multi-stability" (Rosenberg and Verbeek 2015, p.25). Multi-stability simply means the idea that any technology, in this study automated decision-making tools, can be used to serve various purposes and therefore also be of importance or means in different forms to the user. Another important aspect about this idea, is that multi-stability can elaborate the materiality of the device, meaning that it constrains the possible relations towards a certain use or meaning. Technology is not supposed to mean "simply" anything or as a matter should be used to do anything. However, some of these relations can appear stable and therefore define certain stabilities or variations of the technology in the context of multi-stability (Rosenberger and Verbeek 2015).

The contributions, that the possible multiple stabilities, variations, and combinations of relations make, might give the researcher a view of the different parts of the technology and imply that it is more than just the sum of them. The nature of this holistic concept shows that technology may be revealed in very diverse ways, while serving different purposes across various contexts. A simple example could help illustrate these ways and show how our perception is trained. In the beginning Idhe (1986) developed this notion of multi-stability while working on visual perceptions, where he introduces the experience of someone who is confronted with a simple visual illusion. The interpretation attempts of the visual illusions, manifested that there is more than one way to do so. The Necker cube (Figure 1) represents the most well-known example. It can be interpreted in two ways, either as a three-dimensional box with the topmost side on the upper right, or as another box with the topmost side on the lower left (Rosenberger and Verbeek 2015, Idhe 1986).

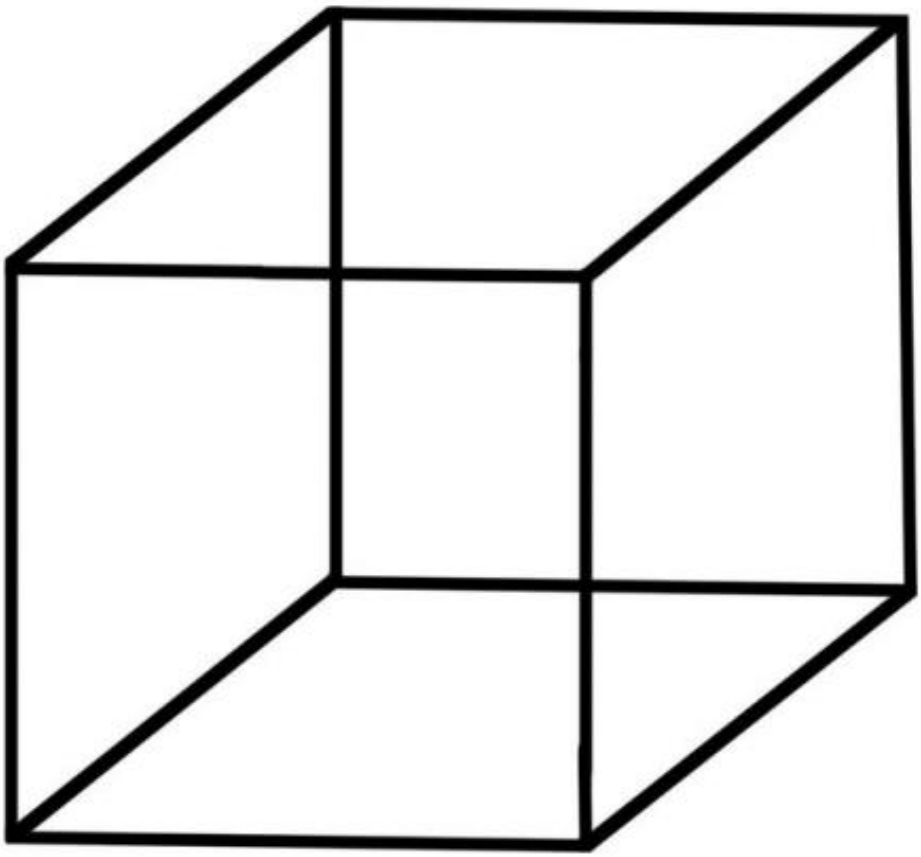


Figure 4. Necker Cube (Rosenberger and Verbeek 2015, p.27)

This Illusion is as mentioned before used to demonstrate the multi-stability of visual perception. This means that each cube shape that the person confronted with sees can be understood as a separate stability. The stabilities themselves therefore each appear separately in terms of their visual representation. Idhe himself wrote, “If a background story can be found to allow a third variation to appear, its stability will coalesce almost immediately” (Idhe 1986, p.96, Rosenberger and Verbeek 2015). The parallel between this example and technology experiences is that it shows how some variations of a certain phenomenon may change from becoming obvious or apparent or on the other hand passive and hidden when the cube is rotated or the viewpoint of the technology changes.

Despite having its origins in the examination of human visual perceptions, the notion of multi-stability can likewise be used when applying it to human-technology relations. The technological artifacts are understood to be represented by a potentially greater number than one possible embodiment or hermeneutic relation and therefore also support the idea of multiple stabilities. Rosenberger and Verbeek sate it as following, “a technology that supports the multiple stable embodiment relations is one which could offer multiple

potential transformations of a user's bodily-perceptual encounter with the world" (2015, p.26).

Assuming the application of this concept to the context of the post phenomenological research approach, "variations of a phenomenon reveal themselves through the examination of a technology's diverse use cases and functions. Oftentimes, studying a technology's multi-stabilities reveal familiar uses or cultural norms, as well as a variety of partial trajectories or alternate views that may help consolidate multi-stable thinking" (Yuen 2021, p.57, Ihde 2002). In other words, the thought of a technology's multiple stabilities serves the researcher to highlight the technology's specific context-depended and materially situated relationality (Rosenberger and Verbeek 2015).

A common example is the smartphone. The primarily purpose and use of the smartphone is the communication with other persons, usually in different places. And while over the recent years, technology has undergone rapid development and innovation, it led to new functions and possibilities resulting in a transformed experience of our phone usage. The idea behind multi-stability as a post phenomenological analysis approach deriving from the assumption that it can be viewed as an empirically testable hypothesis where the various stabilities and patterns of the technology may be observed and perceived from a certain first-person perspective. This additional layer of understanding can be utilized to describe the interactions between the user as a human and other entities as technology or non humas (Whyte 2015, Yuen 2021).

3.10 Analytical approach

Making sense of the conducted and generated qualitative data from the interviews with the help of the right analysis strategy allows more than one interpretation approach. As was mentioned before, this study tried to reassemble anecdotes around the experience the user had. This method combines the human technology-world relation framework as introduced by Idhe with the goal of understanding the findings. The heuristic definition involved the examination of the intentionality in the form of the human-technology relational analysis that served the process of examining the different entanglements and relations (Embodiment, Hermeneutic, alterity, background) of the user and his surrounding world (Yuen 2021). In accordance with the post phenomenological approach the following aligned methods are common research procedures to ensure a basis of empirical significance.

3.10.1 Variational Method

The first method is performing some sort of variational method, where different possibilities or uses of a technology are identified by examining the materiality of the technology, the physicality or bodily utilization, common practises and the contextual background surrounding the overall use of the technology. This method incorporates some of Husserl's concepts of reflection and unpacking the multi-stabilities of the technology across different instances and might therefore help understand the impact of AI technology to the personal impression of users in the public sector from a completely new analytical perspective when compared to the many studies explored during the literature review. (Yuen 2021). The terminology itself can be traced back to the basic method of brainstorming. And even though it is derived from the works of Husserl, it was altered to suit the purpose of post phenomenological research. "For Husserl, variational analysis is a method in which an object of study is observed from multiple perspectives for the purpose of identifying features that remain present from any vantage point" (Rosenberger and Verbeek 2015, p.27).

Multi-stability can also be used with a negative connotation, where the technology is "subjected to variational analysis as part of a critique of another account seen to be somehow totalizing, essentializing, or overgeneralizing" (Rosenberger and Verbeek 2015, p.28). And additionally, the positive notion of multi-stability research can "seek to uncover new information about concrete cases of human-technology relations, or to reinterpret cases in productive ways. In such projects, variational analysis is taken up not simply to demonstrate that a technology is in fact multi-stable (as in the negative usage), but to investigate alternative stabilities to the "dominant" one, and to consider what these alternatives might teach us about the dominant stability" (Rosenberger and Verbeek 2015, p.28). This idea of the examination of how technology mediates certain notions and relations between users and the world can also be described by the empirical turn introduced by Irwin (2016). Therefore, in order to examine the different perceptions, interpretations and trajectories, the method suggest the use of multiple in-depth interviews with human user participants where all the data is analysed and used to compare and point out contrasts between the multi-stabilities or other altercations across the case study.

3.10.2 Variational Cross-examination

The previous segment showed the exposure of stabilities and the technology's multi-stable status as an object of study. The technology may therefore identify with some of that object's stabilities. And as a result of that, the next step is covered by the variational cross examination. The variational cross examination builds upon the stabilities and tries to go further by referring to the critical contrast of the various stabilities that they have

against each other. This additional analytical step helps learn about the particular stabilities through their comparison with one another. Rosenberger therefore recommends that post-phenomenological research should adapt the process of variational cross examination as explicit second methodological step after the variational analysis. He further argues that this method is already implicitly used and at work in many positive and concrete case studies in the field of post-phenomenological research that also refer to the concept of multi-stability (Rosenberger 2020).

The idea behind this approach is to elucidate any of the various stabilities of a "multi-stable" technology in a more critical contrast to others. This also means that the most dominant stability may be the best positioned in regard to its revelations about it. The reasons for this are according to Rosenberger the following, "since the dominant stability typically represents the usage for which a device was made, and the primary way it tends to be taken up in practice, it can be easy for a person to simply and unreflexively believe that this is what this technology "is for," or that this is what it "really means." And it can sometimes be difficult to step out of this kind of thinking. For those who primarily relate to a technology's dominant stability, these relations to it will often be steeped in normalcy" (Rosenberger 2020, p.6).

In regard to the everyday usage of some technologies, this implies that a certain given device may be experienced in the most simple and uncomplicated way, resulting in the perception that the technology is for this dominant purpose and only for that. Rosenberger also defines embodied dimensions for this "normalcy" of the technology, "in this technology's everydayness, in the very way that it is encountered as normal, one's relation to a device in its dominant stability can become set within bodily-perceptual habituation. Through these kinds of habits, the device will enable it to be experienced as immediately meaningful in this dominant context. Add to this the fact that the device itself—while multi-stable and open to other meanings and uses—has likely been designed in a way that optimizes its usage in terms of the dominant stability" (Rosenberger 2020, p.6).

This means that according to the technologies main and dominant stability that it has been designed and produced in its purpose, and consequently suites well to the proper terms. However, not only the device or technology itself has been in a sense optimized for that, but it can also occur that the wider social, geographical and cultural context reflects the dominant stability and therefore reinforces this optimization or support of inclination to use the technology in the particular suiting manner. The cross variational examination for these very reasons tries to see through the common normalcy of the technology. This approach is not always easy whether the research is using a post-phenomenological or another approach. The goal is to extract insights from the context of many specific design

elements and break down the potentially deeply imbedded habits of perception and understanding with the use of a post-phenomenological cross variational analysis (Rosenberger 2020).

The variational cross examination therefore also includes an elaboration on the features that are subject for comparison. This examination can be done in three categories. The categories act as expansion parts to the fundamental post-phenomenological theory and elaborate connections to related fields of thought. The presented categories are by no means exhaustive. Comportments and habits: This category focusses on the human side of the human technology relation with regard to the issues deriving from technological multi-stability. “This refers to the particular ways a user approaches a technology, and the particular ways they understand it, that enable them to engage the device in terms of one specific stability. To take up a technology in terms of one particular stability, rather than other possible stabilities, how must one physically comport oneself with respect to the device? What must one know about it?” (Rosenberger 2020, p.7) This means that when confronted with dominant stabilities, this category is about the observation of humans who fail to recognize their design functions and rather develop relational strategies through which these dominant relations are encountered as normal, based on the users’ internal conceptions or learned pre-perceptive habits.

Networks and co-shaping’s: This category focusses on the contrast between the technology’s various stabilities in terms of their connection between contexts that some objects or humans might associate with each. This can be done by examining the conception of the extended social collective of the human user and the technological artifact in the actor network perspective. This means going beyond the level of an individual user’s experiences and extend the post-phenomenological insights with those of social theory or sociology and anthropological approaches. This can for example be an actor-network theory, which describes technology in terms of its role within a network consisted of other devices or humans. Humans and technology both can be seen as actors and are joined in a shared network that enacts certain programs of action. The result of this thought is that the responsibilities for the actions performed is shared between the different actors (Rosenberger 2020, Latour 1990). In this way a more social perspective on the collective actors is given in which a productive approach towards investigating a technology is to comprehend its multi-stability, which is the foundation of putting the various stabilities in contrast in regard to their social networks and how they may contribute toward them (Rosenberger 2020).

Post-phenomenology is all about how humans, devices and their world are shaped by the technological mediations (Rosenberger and Verbeek 2015). Therefore, it can be assumed

that humans alone are not the only subject when trying to examine them as actors, interpreters, and decision makers, but that their relation to others and their devices is crucial to this process. Verbeek writes, "we need to replace the 'prime mover' status of the human subject with technologically mediated intentions. In our technological culture, humans and technologies do not have separate existences anymore but help to shape each other in myriad ways" (Verbeek 2011, p.16, Rosenberger 2020). As this study is focused on the technology behind decision making, it can be helpful to consider the importance this category puts to cases of moral or political decision making.

The literature shows us that "the context of our decision-making, the constitution of the options we decide between, and the authority of decision makers are all co-shaped by technological mediation. We can integrate our ideas about multi-stability and variational cross-examination into this body of post-phenomenological thought on mediation and co-constitution. As we spell out a technology's various stabilities, we can critically cross-examine these stabilities as cases of technological mediation that each co-constitute humans and the world differently" (Rosenberger 2020, p.8, Verbeek 2016).

Material tailoring: The final category of material tailoring includes the aspect of stabilities that are subject to critical comparison in the concrete way they have been changed or transformed in order to achieve an enhanced service to a specific function. This category again can be expanded or build on top of other fields such as social and anthropological theory in addition to the philosophy of technology. In Rosenberger work on cross examination, he mentions Madeleine Akrich's enunciation of a script theory on the development of technology in conjunction with actor network theory (Rosenberger 2020, p.8). The thought behind this is embodied by the extension of our previously elaborated metaphor which used defined humans or devices as actors, so that it conceives both humans and nonhumans as following a social script. Thus, the term inscription is used to refer to the material modification done to technological devices with the aim at achieving an enhanced circumstance for them following the social script of the social network (Rosenberger 2020, Akrich 1992). Another term that is associated with material tailoring derives from the work of Andrew Feenberg and is called "systematization". It refers to the alterations made to a device so that it is suited or at least better suited to fit in the larger socio-political purpose (Rosenberger 2020, Feenberg 1999).

These two theories can therefore be used to contrast a technology's various stabilities in terms of their ideas from the social script theory and systemization like material specificity. Concluding, the term of material tailoring refers to the concrete ways a device can be changed to better suit a particular stability in this case. Therefore, the result is considering a technology's alteration towards a better service of that usage, which also

includes adopting technology for an alternative stability and how the device may be transformed to better execute this alternate usage (Rosenberger 2020).

3.10.3 Techno ethical Analysis

As Rosenberg and Verbeek and others state, technology is understood as non-neutral and in the post-phenomenological research field, ethical considerations in regard to utilization of technology and its mediation is a common point of interest. The affordances of a technology or its societal impact and constraints can be morally specific or have a value burdened nature (Yuen 2020, Rosenberger and Verbeek 2015). This assumption can also be applied to the technology used in the public sector and might have a significant influence on the way in which public services are delivered and how users are subject to the non-neutrality of advanced an intelligent technologies, which makes this ethical approach worth exploring to the explored case of this study. The aim of an ambitious techno ethical analysis can be to examine the instrumental or means-end problems of positive or negative outcomes from utilizing a technology as a tool. This means uncovering the technology's embedded values, scripts, or political biases in addition to exploring the human experience that is co-constitutively conditioned by the users' habits and use of the device (Yuen 2020, Adams 2019).

The technological mediations that were mentioned before can also be seen in a techno ethical way, were they tend to have an intrinsic ethical duality between the actor's involvement and the alienation of that very actor. The ambivalence is not only fundamentally inherent but is particular in the developing or integrating conditions of a technology. Therefore, this ethical contradiction of technology may be a central point of interest when examining and assessing the techno ethical impacts of resulting opportunities and impediments on the user's perception (Yuen 2020, Rosenberger and Verbeek 2015). This process is called involving-alienating structure and while there are various different factors involved in the effects of the ethical dimension of a specific device or technology, the emphasis is laid on the importance of a technology's design and its uptake following the subsequential ethical insinuations (Yuen 2020, Rosenberger and Verbeek 2015).

The distinctions that can be made when examining the techno ethics of technology are between the instrumental techno ethics, the socio-material techno ethics and the existential techno ethics. Common ethical thinking and questions spread about the instrumental or ethical assessments of a technology, which focus on the cause and effect of ethical impacts that can be associated to the utilization of a technology. This approach assumes however that the technology itself is a neutral tool or artifact, and only the user as a human can be seen as the moral agent that determines the impact the technology has

whether it is on them, others or society in general and whether they are good or bad according to their believes. So, in contrast to the findings of Rosenberger and Verbeek, in this case the technology is believed to be simply a tool and humans as users decide on the use as well as the resulting impacts (Yuen 2020, Adams 2019). Thus, when talking about instrumental ethics and its perspective, the main area of concern includes the "...exploration of human values or rights, particularly topics concerning the consequences of technology use such as privacy, quality of life, relationships, communication, participatory policy development, and general technology uptake or trajectories" (Yuen 2020, p.62). The constructivist view of technology and the reflection on the ethical significance of technology suggests that the issue around the tool and impact perspectives are located with societies limitations where the domain of the technical is merely neutral. The often expressed and related phrase "guns do not kill people, people do" implies that guns are just a mean to the peaceful or violent act depending on the user (Introna 2017).

A possible alternative perspective on the ethics of technology is called disclosive ethics that highlights the socio-political perspective of techno ethics. This approach is based on the social constructive argument that technology is socially constructed and therefore indicates an ethical concern. This signifies that the technology is political by design and as such already implies certain interests while simultaneously excluding others. However, the person responsible in the design process might not even be aware of their decision making on socio-political or ethical impacts (Introna 2017). The assumption lays on the statement that people in fact are often times not aware of this implication but just trying to solve very ordinary everyday solutions to potential technical problems. This means that even without being aware, the designer might still make assumptions or contains certain values and beliefs related to his or her own personal experience when constructing the devices and technology (Introna 2017). Introna gives the following example: "the ATM bank machine assumes a particular person in front of it. It assumes a person that is able to see the screen, read it, remember and enter a personal identification (PIN) code, etc. It is not difficult to imagine a whole section of society that does not conform to this assumption. If you are blind, in a wheelchair, have problems remembering, or are unable to enter a PIN, because of a disability, then your interest in getting access to your account will be excluded by the actual design of the ATM. In this way the ATM embodies a particular understanding of the world (humans) in front of it" (Introna 2017, p.13).

Despite technology artifacts embodying interest already from the start, it nevertheless means that users can not intentionally or unintentionally reinterpret or alter the way technology implies or provides opportunities for them to pursue their own interests. Introna argues that "users often 'read' and use technology in ways unintended by the designers/implementers. However, as these technological affordances become embedded

in larger infrastructures (practices, systems, spaces, organizations, etc.) it becomes increasingly difficult to use the technology in ways other than in the way it was set up to afford possibilities (or not)" (2017, p.13). The intention of including or excluding specific interest into the technology makes it not only political but as mentioned above insinuates an ethical perspective. The focus on the techno ethic approach of disclosive ethics lays therefore on the way in which the interests are being built into the technology and the related practises. This argumentation denotes that as part of the design process the ethical reflection should be an inherent part. On the matter of information technologies important assumptions and biases are commonly obscured and subsumed in the algorithmic black box and therefore difficult for the researcher to scrutinize them (Introna 2017).

The third and last perspective is called existential ethics which can be expressed as the closest related perspective to the concepts of post-phenomenology (Yuen 2020). The implication lays on the relationship between technology and society. So instead of solely focussing on the human as a user or on the design process the idea is that technology and society co-constitute each other from the beginning. They are not only each other's condition of possible being, but the technology goes beyond being the artifact alone and instead contributes to the technological attitude or disposition that the artifact makes. This means that the artifact appears as meaningful and necessary from the starting point to the finish. Once the technology exists in from of an artifact its disposition makes it meaningful but also discloses the world of the user beyond the mere presence of the technological artifact.

As part of the techno ethical examination the aim of this perspective is an ontological disclosure of the implied ethics. In a sense this approach opens up and tries to reveal the conditions of possibilities that make "…particular technologies show up as meaningful and necessary (and others not). It seeks to interrogate these constitutive conditions (beliefs, assumptions, attitudes, moods, practices, discourses, etc.) so as to problematize and question the fundamental constitutive sources of our ongoing being-with technology" (Introna 2017, p.18). Furthermore, the fundamental belief of the existential ethics revolves around the belief that the humans and the technology are both equal parts in the conditioning of one's experiences and relations to the world around them. Deriving from this comes a very common ethical issue concerning how technology supports and augments certain habits or relations or weakens other aspects of the user's perceptual frameworks in terms of understanding and human capacities (Yuen 2020).

3.11 Data Analysis methods used in this study

The conducted and created data from the before defined data gathering methods were reassembled into the explained human -anecdotes, which were then subject to a mixed

analysis method between the variational examination, the cross variational examination, and a techno ethical examination. The reassembled anecdotes are based on the data collected from the pre reflective research, the observations and the interviews serve exclusively as analytical material for this study. The anecdotes are presented by several thematical extracts which are ensured to keep the interview partners anonymity. The ethical use of the data is an important aspect of this work and is limited for the research purposes of this study only. The anonymity of the participants is ensured by changing any data that could be traced back to them and providing pseudonyms in order to keep their personas confidential. The created transcripts are also kept confidential. Therefore, With the help of these methods an attempt for a more profound and deeper understanding of the data in relation to the research questions was performed. The analysis methods made use of the theoretical framework that defined the concepts of multi-stabilities as well as the mediation of the human technology world relations.

4 Study Research and Background

4.1 Introduction

To examine the empirical implications from a post-phenomenology perspective on AI systems in public sector services, I will use the case study of the decision-support system OTsustusTugi (OTT) used since October 2020 at the Estonian Unemployment Insurance Fund. This specific case was chosen in order to represent an already fully integrated and used system within a public sector institution. The system itself has been, as just mentioned, live for roughly over a year and since then the project has proven that modern machine-learning models could be successfully implemented and operated in and with public sector work processes as well as existing IT solutions like the famous Estonian data registry X-Road. The case also represents a claimed successful project that helped to improve the organization's service and efficiency in addition to making the work more pleasant and save money (Nortal 2021).

The fact that this tool has been more thoroughly used compared to many other relatable, new and emerging technologies, guarantees a certain proficiency within the actual use that might help gather more profound and meaningful results. And while every case has its limitations, this specific choice has also been made in order to focus on a participating country within my Pioneer master studies, allowing me to apply previously acquired knowledge on the country, culture and digital literacy. The Estonian Unemployment fund is not only important for providing the services to Estonian citizens but an important integrator of emerging AI technology. Besides the examined OTT tool other systems such as the MALLE also work based on the newest AI technology allowing the Institution new potential on labour market measures and analysis.

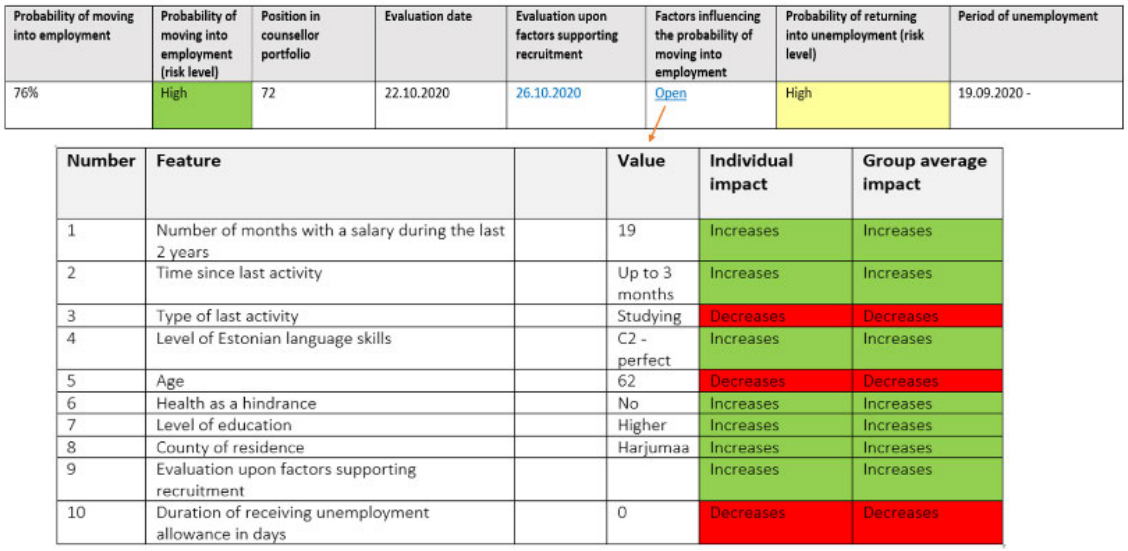
The labour market and with-it long term unemployed citizens, as the tools main data subject, provide a very fragile and vulnerable area of application. This fact contributed to my choice and interest into this specific case. Besides the data subject, Estonia is known for its digital infrastructure, which also implies a certain degree of digital expertise not only within the private but also the public sector, establishing a research worthy constellation on interested stakeholders ranging from the top-level experts and designer to the actual career counselling users and the subjected citizens. Therefore, this case study may contribute to the discussion on ongoing digitalisation projects not only in Estonia and its public sector but additionally to associated countries that have adapted the fundamental structures of the Estonian digital state like the X-road system and are trying to follow its path towards providing new and innovative ways of public and especially social services.

The popularity can be seen in the fact that the OTT decision-support tool project has also won a recent award from Estonia's government digital services competition for best use of data. Framing the project to be an "excellent example of how e-Estonia's infrastructure (i.e., the X-Road, electronic identity, and registries) creates an opportunity to use the data so that it generates added value in everyday decision-making, be it on a personal, organizational, or even country level" (Nortal 2021). The website claims that the decision-support tool also helps the Unemployment Insurance Fund better understand the labour market's ongoing processes and develop better labour policies. The EUIF handles about 70000 new unemployed people annually, providing over 30 services.

The underlying aim of this project is based on the complex and critical developments on the labour market and should help the allegedly 350 consultants of the Estonian Unemployment Insurance Fund inspect and analyse a far greater number of unemployed and job seeking individuals and their information. The data that is analysed is used to create appropriate action plans to each and every citizen with the goal of helping them successfully return to the labour market. This process includes collecting and analysing a great amount of data about unemployed citizens but also data related to the economy and state policy measures. The system therefore applies artificial intelligence technology that can be best described as a random forest machine learning model. The model itself has therefore been trained and tested with the help of unemployment data of the last years. The system itself uses numerous different attributes and indicators that are said to allow the system to evaluate each citizen and claims to calculate an approximation of their likelihood to find a new job (Nortal 2021).

The description of the system further states that the different attributes are equally about the employment seeking citizen and include for example their education, previous job experience, right to benefits, health restrictions, as they are about the relation to the labour market and the economic situation which for example can be expressed as the number and type of available positions in different regions and the number of newly unemployed people. The system has an additional prognosis capability that allows the EUIF to calculate the chances of people re-entering unemployed within the next year. The claim is that EUIF users and consultants can offer suitable labour market services based upon the factors found in an alleged interactive dashboard that helps display the results of the machine learning process (Nortal 2021). The exact data used by the tool is divided between the job seeker and includes Socio-demographic data, Previous unemployment data, ICT literacy, belonging to risk groups, unemployment benefits, labour market history, labour market services received, earnings after unemployment spell and the labour market data that consists of the overall inflow into unemployment, outflow into

employment, and number of vacancies in the area corresponding to job-seekers job wishes (See figure 5).



Probability of moving into employment	Probability of moving into employment (risk level)	Position in counsellor portfolio	Evaluation date	Evaluation upon factors supporting recruitment	Factors influencing the probability of moving into employment	Probability of returning into unemployment (risk level)	Period of unemployment
76%	High	72	22.10.2020	26.10.2020	Open	High	19.09.2020 -

Number	Feature		Value	Individual impact	Group average impact
1	Number of months with a salary during the last 2 years		19	Increases	Increases
2	Time since last activity		Up to 3 months	Increases	Increases
3	Type of last activity		Studying	Decreases	Decreases
4	Level of Estonian language skills		C2 - perfect	Increases	Increases
5	Age		62	Decreases	Decreases
6	Health as a hindrance		No	Increases	Increases
7	Level of education		Higher	Increases	Increases
8	County of residence		Harjumaa	Increases	Increases
9	Evaluation upon factors supporting recruitment			Increases	Increases
10	Duration of receiving unemployment allowance in days		0	Decreases	Decreases

Figure 5. Example Data Output of OTT (Võrk, 2021)

The underlying methods of the OTT risk model relies on five years of data that combines approximately 250 thousand spells allowing the random forest model to calculate the exact risks of finding a job and returning to unemployment. The risk model is disintegrated of effects in order to avoid the creation of a black box algorithm and is newly trained about once a quarter year. The product description offered by the developers further addresses that the OTT tool is not only aimed at the service of unemployed people but that the consultant is supported by showing where to put their effort and priorities. These proposed beneficial functionalities towards the counsellors include new support channels such as online, phone or real-life consulting, the ability for the counsellor to decide upon the frequency of contact, highlighting the most important factors affecting the client's job seeking journey as well as getting an overview of their portfolio and risk level of their clients. Therefore, the counsellors are influenced in their planning and scheduling of their individual workload as well as categorizing the risk of the citizens (See figure 6).

This signifies that when unemployed citizens are what they call as "at low risk", the system implies that this citizen will likely find a new occupation on their own and therefore be independent from the services offered by the EUIF and their consultants or at least only require little assistance. This indicates to the contrary that individuals with a calculated high-risk level of long-term unemployment are categorised and considered to need more attention and assistance. The tool is not bound to street level bureaucratic use of the consultants but is supposed to also help managers gain a better overview which

allows them to manage and relocate accordingly and allows for a workload distribution that benefits and supports each consultant and the citizens as his or her client. Finally, the project description incentivises that the overall assimilation of the tool will assist the EUIF and help understand ongoing labour market developments, processes and policies (Nortal 2021).

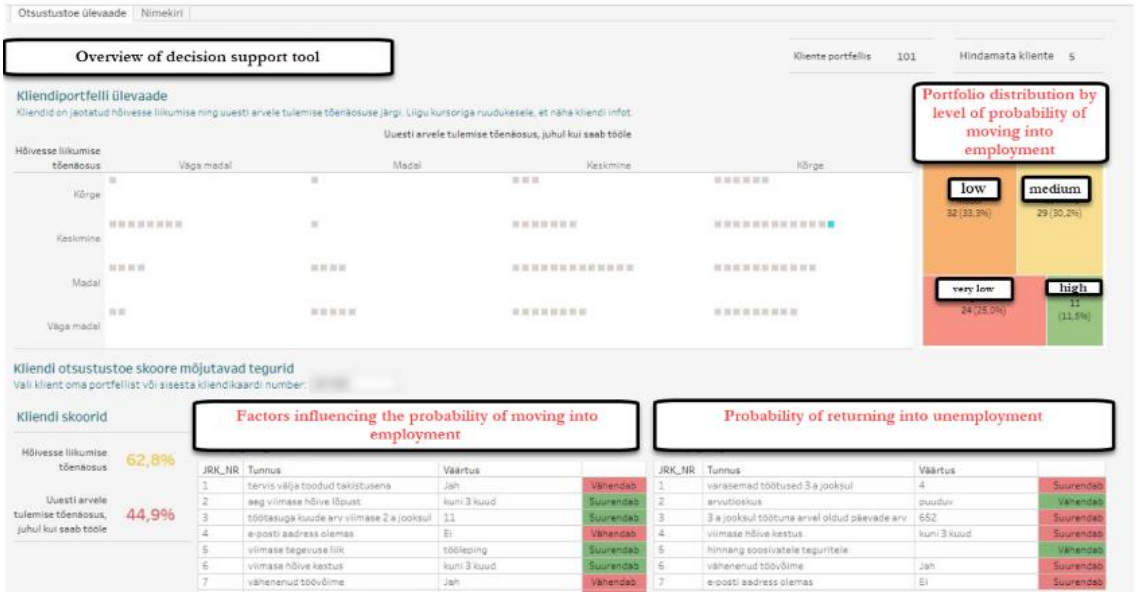


Figure 6. Example data output of OTT 2 (Võrk, 2021)

But how is this decision support system impacting not only the lives of the citizens but also the experiences of the consultants and the experience among users themselves? The previous introduction of the tool has promised significant benefits not only to the citizens as clients but also to the consultants of the EUIF as users of the system. How do the consultants consume the information, how do they self-identify themselves as confident decisions makers and evidently as good consultants while they are interacting with the tool's technology and its design?

4.2 Study Research Design

Coming back to the beginning of this thesis, three research questions were elaborated. This research is aimed at exploring a user's experience with AI and their insights on using this form of technology deriving from their personal perceptions of how their work facilitation and other practises have been affected. With this in mind, the case of the Estonia Unemployment Insurance was chosen with the goal of examining the consultants' experiences and discernments on using the OTT decision support tool.

To investigate these proceedings, I started by contacting the EUIF, private individual through AI research contacts and the software provider of the OTT decision support tool,

Nortal. This initial contact included a brief explanation of my study intentions and the background of my research. The communication has primarily been via email and was extended to online meetings for both sides to elaborate more on their behalf. This process included me explaining my studies in more detail and providing a certain motivation for the significance of this research. The first meeting was held between me and the project manager from the EUIF department for development and responsible for another AI system called MALLE that assesses the impact of labour market services. As a result, from this call, the project manager and I agreed up focussing on the OTT decision support tool and the consultants rather than the MALLE assessment tool because it is mainly used by chief information technology specialists. That means caseworkers or in this case consultants have no use of the system and have in fact no access to the system. This information was provided by another employee of the EUIF.

The next step included the EUIF project manager that was contacted reaching out to colleagues that work as consultants and have direct access to the OTT system as well as direct contact with citizens. In addition to this ongoing request, a private contact to a former employee of EUIF has also allowed me reach out to old colleagues of her. The consultants would then on a voluntary basis offer to be subject for one-on-one semi structured interviews. Knowing that interviewees were participating on a voluntary basis they still remained with the right to withdraw at any moment, which means that at least three to five consultants needed to be contacted in order to provide the required quality on conducting this post phenomenological analysis. Following up certain difficulties regarding the collaboration with the EUIF employees and their participation, the sample also includes the former employee. Further, the sampling implied reaching out to contacts and snowballing through these, resulting in three conducted interviews. Beginning this thesis in February 2022, the interviews were scheduled around March and April 2022 with the consultants working with the OTT support system at the EUIF or indirectly participating in the work of the EUIF and its IT landscape.

As mentioned before the agreement with the project manager and later on with the direct contact to the EUIF employees implied that participation was completely voluntary, and dependent on the actual experiences and current use of the employee with the OTT support system. Since the process of consulting a citizen, itself is a very discreet matter, observations of each interview using the system were unfortunately not included also because the situation around the covid crisis made it harder to access any real-life meetings, leaving online meetings as a practically non-existing alternative due to confidentiality and data protection reasons. Due to reasons of other ongoing research and rebuilding projects the timely availability of each employee was very limited. This meant also limiting the interview to under half an hour. This represented itself as a quite difficult

task. However, the 25-to-30-minute online semi-structured interviews were conducted with the consultants using an interview prompt consisting of a predefined set of open-ended questions. The questions itself were build and chosen around the work of past phenomenological and philosophy of technology work (Yuen 2020, Sarcasas 2020).

As a matter of data collection and quality insurance, all interviews were audio recoded with the given permission by the interviewee. The recordings were then used in order to create written transcripts. Confidentiality and anonymity are given by removing any information that could lead to the identification of the interviewee and replacing their real names with pseudonyms. A short list of the participants using pseudonyms follows:

- Mrs. Rummo is a former IT system developer and has worked at the EUIF for over two years with her leaving the institution in 2020, She can be described as a middled aged women coming from an information systems background and was mainly focused on the data and technology infrastructure at the EUIF with a moderate knowledge about past and current career counselling or management affairs in regard to the OTT tool.

- Mrs. Dena is an experienced career consultant that has worked at the EUIF for about six and a half years, allowing her to express her past a current experience working with and without AI technology. She can be described as middle to old aged and has a social work background

- Mrs. Bahka is a less experienced project worker at the EUIF that has previously worked as a career consultant also at the EUIF for around two and a half years. She can be described as young to middle aged and has a strong financial and management background.

- Mr. Yu is an IT system administrator responsible for ensuring the functionality of the OTT and other systems to the counsellors. He has nor real interaction with the citizens nor with the day-to-day tasks of the consultants but regularly evaluates comments and inputs from fellow employees at the EUIF and is in contact with outside partners from private of public organizations.

The results of the conducted interviews are presented in the following Findings chapter. While trying to analyse the gathered information, the interviews were closely examined and used to create common themes for the post human anecdotes. The anecdotes are the foundation for applying the analytical methods previously introduced, allowing a deeper understanding of the post phenomenological implications of the use case.

5 Findings

The following chapter tries to elaborate on the findings using the collected data from the pre reflective research that contains the information given by the EUIF website, the email exchange with employees and users of the OTT system, the literature on the AI systems and practises at the EUIF as well as the conducted semi structured interviews as the main source of information related to the personal perceptions and experiences with the case related OTT AI system. The data was, in accordance with the formerly introduced methodological concepts, structured into relatable themes that were subject to the post phenomenological methodological analysis. This allows the following chapter to elaborate on the various sperate anecdotes which were extracted and formed during the qualitative separation process. Furthermore, each Subchapter then represents one of the created anecdotes, describes the main points of interest and lays the findings into the post-phenomenological analytical perspective in order to create interpretations that might help answer this thesis's research questions. This is done by using common themed interview extracts to highlight the creation of a common based subject anecdote. The chapter therefore shows the way in which a fundamental notion of reassurance emerged through the integration of the automated decision support tool technology and its handling showed the very context-dependent human technology relation and the inherent multi-stability. Decision support and resource allocation tools mediated not only relations in distinct circumstances but facilitated or restricted the user's behaviour and actions appropriate to their individual affordances. The found variation has had direct results on the reported experiences in accordance with their task of counselling and their encouragement in this task.

The OTT tool was designed to support and relief the consultants and through this enable them to provide better services to the job seeking and unemployed citizens of Estonia. Additional intended functions included the creation of portfolios that would allow mangers to assess and locate resources according to the needs of the consultants and the number of inquiries. The software itself is activated for a majority of the time and helps provide the consultant with day-to-day information and suggesting the frequency of each upcoming appointment as well as predetermines some line of communication for special cases or at least suggests them for cases when the user needs assistance or relies on the results and past experiences of and with the system. In this way, the tool was used in response to an active situation as framed by the users when they decided to press and view the suggestion of the tool. But conversely the tool is also intended to provide the monitored data on the single cases and a regular created portfolio of the specific workload each consultant has, in order to support the scheduling and management of the local capacity.

The just explained commitment also means regular input in the form of data or interactions from users that would help the system adjust and take the right measurements. The OTT system relates the user to a massive transfer of personal data connected to the labour market but also including other fields that have been obtained by numeral processes including directly requesting citizen information in order to support the decisions that will lead to their unemployment service or consultation through the consultants. While I was not able to conduct any information about the citizens comfort of being monitored this exact feature did not account the consultants' experiences or their confidence. Throughout the data analysis process, the gathered information was reassembled into the posthuman anecdotes, which were then the foundation for employing the post phenomenological methods that would allow different variations to be observed during the OTT case. The observed variations therefore help explore possible mediations of the human-technology-world relations and thematic analysis.

5.1 Mediating a Parallel World

By associating the quantitative and numerical probabilities with the specific behaviour a job seeking citizen is likely to fulfil, the career counsellors of the EUIF can issue his or her advice and support and is therefore directly extending his or her consultation action through the OTT support tool (Consultant – OTT → Citizen). The moment one of the EUIF consultants accesses the systems information and sees the interface, he or she also sees the unemployed citizen in terms of numbers and probabilities, ultimately presenting their clients as a digital overlay provided by the OTT system (Consultant → OTT – Citizen). Furthermore, associating the numerical values to the real and specific present but also future skills and behaviours of the unemployed citizens, lets the career consultants effortlessly tally the citizens employment behaviour and the contraventions to their employment status compared to other citizens and numbers. While this might have a great impact on the citizens human-world relation, the AI technology is kind of a constant reminder of their possibility to enter the labour market and their performances in certain aspects of their life. However, since they are but the data subject and not directly using the OTT tool, many of the citizens are not aware of the quantification of theses parameters related to them.

"We talk about the jobs and where he can go…, and also about the applications and how they are going. The data is not talked about, sometimes I ask questions that help me understand…. when the client is not good with computers, I can ask and help find courses to do. I use the data in the interaction to build better relationships."

Mrs. Dena gives a very clear explanation on the function of the OTT system and how the role of the data is fundamental to the interaction between the consultant and his or her

clients during every day works processes. In this way the OTT system is not only a decision support tool but also serves as a tracker of labour market accountability with the goal of taking this information to induce a positive long-time effect on their behaviour. The system spotlights the most important aspects and parameters according to the design of the system and therefore drawing attention to these very aspects. To put it in post phenomenological terms the technology is amplifying these very parameters in the citizens life but also due to the regular awareness in the case of the consultant, he or she might be prone to missing out on other important factors that could have been overlooked.

"I look at the information and think sometimes when the clients tell me something, can I share this into the file and is this the data I see. Yes, it can frame the way I think what I want to say. But I take that information for granted."

Extracts not only from Mrs Dena but also from Mrs Bahka another experienced career consultant showed similar viewpoints to create this mediating perspective of the systems data and its functions. Thus, the system and the integrated numeric system might not only be bound to its primary functions but actually acts as a broadcasting system to the users who were able to identify the importance of certain parameters making him or her manage his or her own behaviour or interpretations related to labour market issues. The OTT system leaves little room for self-interpretation through the user; however, this does not mean that the resulting actions are solely based on the OTT possibilities. One example can be a career counsellor deciding to single handily compare the data himself with related cases he has worked on or otherwise experienced as a form of meditative counselling method as the following extract will elaborate on.

"Yes, I think the experience… the data can be normalized…I think it is not good when saying it is the reality, not one-to-one. It can play with our head. We also want approval of the information and help create better services. So, we all think of the numbers and what they mean. They make things possible that maybe do not… do seem before. There is … anxiety, I think there there is a impulse a modern impulse to control much of our experiences. I also want to know all about me."

In conclusion, the OTT support systems numerical values are but the simplified reflective measure of the citizens behaviour that are used to help the counsellor gain important insights about the job seekers individual labour market performance in the past and therefore the basis for offering advice on comparable citizens and cases. This though leads us to the assumption that the OTT tool reinforces the ongoing work practises at the EUIF and strengthens the consultant's confidence in his actions and decisions during the consultation. This also means that the numeric parameters and possibility-based criteria of the tool also unconsciously encourage the career counsellors to be more subjected or

responsive to cases of the represented labour market performances and behaviours. The case shows that the OTT could influence the realization of the counsellors understanding of the citizens needs and devote his decision and or attention towards particular citizens. And without realising users may become sensitive to every apparent deviation of what the system thinks is a normal pattern of things. The apparency of out of the shelf "citizens" will make us anxious and as Mrs. Bahka shared during the data collection, the system is in some form extending personal responsibilities and how it might help in their personal tasks.

Interviewee: "Hm… It can be that we think this system is a machine and and this machine maybe just used for confidence in our work."

Interviewer: "So it might evade your feeling of responsibility?"

Interviewee: "It think you.. yes it can be. The algorithm and the structured work and process helps.. in general for this kind of convinced feeling the worker has."

Interviewer: "But would you say that its just the distribution of agency and it becomes hard to hold anyone accountable? It becomes very intriguing to say the machine said so."

Interviewee: "Hm.. yes maybe. But this also happens…ehm sometimes in most bureaucratic systems anyway"

While the system itself is meant to relief the case counsellors work and as a consequence improve his ability to provide the best possible service, making information available that was in a sense farther away before integrating these systems. The technologies functions and information that is used by the counsellor may not have been within his normal reach bodily but has had found some new process to come to him or her and in that sense the system not only shortens the distance of available information but also shortens the distance of the users own lived experiences and what might have been important to him or her. This means that just like every ICT technology, the use of more and more digital solutions like the OTT system, will alter the meaning of distance to the user. However, the findings also suggest that along with this process comes a new meaning of "importance" as something that the user actively engaged and decided on before might now require less attention from his side.

5.2 Multi-stability

"And eh much more data than we need. It can be a problem. then the question is arising is what are you doing with the data you are not using why you are getting it for example.

Another kind of.. another kind of ethical questions about data exchanging and data using and using."

As the extract from Mrs. Rummo the former IT developer shows, one crucial thought behind the EUIF IT systems seems to be the way in which data can and is used. The multi-stability of the OTT decision support tool is most unequivocal in the way information is showcased in a structured and implicit manner (also see fig. 5). Therefore, this subchapter will henceforth focus on the informational or "data" multi-stability of the OTT. As was already mentioned during the exploration of the automated decision-making research during the literature review, such systems tend to do some form of classification and ultimately also consist of an indicative and analytical function that allows for a standardized representation of the citizen and his or her related information. The underlying computer code or data is often a reduction of a much deeper, more complex and definitely more chaotic human world and narrative. Automated decision making or support tools have come a long way since the initial applications of the identified uses, up to the handling and categorization of data related to the unemployed and job seeking citizens of Estonia.

"Yeah, they actually have unemployment data and every time the institution has the data about the person in Estonia for example in a working or on social interest and they are having any kind of benefits that kind of thing and eh… and eh… This allows the system to work like it does."

It becomes evident from this case and the extract material that the OTT tool was customized and tailored to do exactly this and therefore present some level or at least an increased level of granularity to the information and needs of the user. However, this diminution and standardization of information also allows it to become multi-stable and reusable by other means that might not be in the centre of awareness of its users and its data subjects. The OTT is accessed by the career counsellors but also by administrative and management personal for several reasons, in very varying circumstances and functions. For example, will the career counsellor or consultant use and have a different access to the OTT than an IT person or a local manager conducting resource reports. This also means that each user might work in slightly different sections of one and the same systems of the OTT tool. Core information might still be shared and made available by all though. The following email exchange is one example of the clear difference between the users access to the system and its data.

"Also, I'd like to specify that I am a chief specialist, and I don't have direct contact with clients. My work consists mostly of career rehabilitation service analysis and sometimes design or planning and organizing. People I contact outside of my organization are

mostly representatives of service providers or government agencies. Ordinary caseworkers use the system in a completely different way and don't have the same access".

Processed data or information has many advantages, one of them being its affordance to be put considerably more effortlessly to a great number of reasons and functions. This implies that the collected information or fragments about a job seeking citizen can be dragged out of the citizens unemployment context, and be used for purposes ranging from research, profiling, observation in the easiest way. On the contrary, when handling unstructured or unprocessed information, then the data that allows the underlying system or machine to easily read and understand but it is not usable to the human user. This means that only the system can handle the decision support notifications, the automated analysis of labour reports and the computational confirmation of related measures to the specific OTT user for a given job seeking citizen and the given evaluation or categorisation. The constant process of retraining the systems algorithm means that there is also an increased need to use or reuse the handled information of the OTT for learning purposes which will only extend in future scenarios.

"I think that in the future the system can be extended to other areas…It has to be of course adjusted and examined but the overall direction is a... of the public sector here in Estonia… is going to towards more automation."

The result of this analysis is that there are at least four distinct primary functions, perspectives or just stabilities of the data used by the OTT support tool that can be identified as following:

- Individual job seeking citizens counselling (job suggestions, follow ups, communication of benefits, making appointments)
- Resource allocation (financial, human and knowledge)
- Labour market analysis and measures (Policies)
- Machine Learning and Innovation
- Scientific Research (AI, Public Services, Management)

Other peripheral uses could not be gathered, but this does not mean that they are not there. But if we think of the data as the building blocks of something like a metaphorical park bench, while the communicated uses are maybe all intended, they therefore represent a person sitting on the bench. However, the intentional use of this data for personal or just

not thought of purposes (that do not need to do out of misdemeanour) can represent a person “misusing” this park bench as somewhere to sleep or lay down. The interview gave no reason to assume that there is such behaviour or any indication of will towards changing the data structure so as to prevent any alternative use. Nevertheless, the reason this thesis introduced the famous example of the Necker cube was to draw a possible analogy between the case specific multi-stability of information and the optic illusion that was previously mentioned. The particular factor that this system does not only handle the data but interprets and might mediate its own “world” makes different stabilities related to the mediating role of this data quite interesting. The one perspective of the cube aimed at describing the job seekers individual data, needs to be clearly recognizable for the observer in the given interpretation of information provided by the OTT system. However, while this is one image the other perspective of the cube must be visible at the same time, meaning that the OTT system must also provide the information for the other stabilities that were previously identified at the same time. And while the Necker cube has the observer choose between two possible images of the illusion, the consideration of more stabilities tries to fit into this cubic image producing a far more complicated representation or drawing. The only way of achieving this, is by deliberately choosing to let go of a certain level of detail in each of the included stabilities. What this analogy is trying to say is that when we look at the highly detailed illustration of one possible stability of the OTT systems information, it is hard to discern the other stabilities into the same image and this also goes the other way around. Leaving the system to abstractly interpret each function to his or her own will without any human interaction.

It becomes obvious that in order to produce a representation of the data where all stabilities are recognizable, it needs to be deprived of some of the specific details. This abstract representation implies an indigent, unusual or reduced outcome. Therefore, with the help of this analogy it becomes clear that there is a tension among the need for multi-stable information that has been or can be structured and standardized and the users basic need for human eloquence and richness in human interactions as well as his or her world interpretation which implies or allows a certain degree or uncertainty and relies on personal impressions.

Other findings suggest a possible overlap of interests between the institutional priority of providing a “better” service to unemployed citizens and labour market management on the one hand that highly relies on the data and demand specific information and the users human experience and knowledge acquisition through human world interactions and on the other hand, by implementing a system that relies on a massive amount of data and algorithms, the post phenomenological image assures to include all the needed details related to the categorization of the job seeking citizen. However, this type of information

becomes useless for someone who does not comply to the algorithms categories and therefore from providing the information that the users have started to heavily rely on, while counselling the unemployed, has the image of human interaction and external expertise become lost or needs to be put back together.

Nevertheless, from this post phenomenological position on multi-stability of the labour information of vulnerable citizens and its mediation through the OTT support tool, the primary obligation of the Estonian unemployment insurance fund is to provide the best possible service and support to the citizens of Estonia, other functions being subordinate, this thesis might argue that the main responsibility of the OTT tool and the surrounding system implementation should be to guarantee the best conceivable service and support during the citizen encounters with the EUIF and its employees. This might be of special importance when considering improving unemployment or labour market information exchange between possible employers, the agency and the patient, particularly due to the fact the OTT system and the data relates directly to how the career consultants use the system. The findings of this thesis constitute that from this perspective, the information gathered by the EUIF and especially the OTT system is already to some degree developed to allow the users to focus on the main stability of citizen consultation and job offering but has also shown great importance to a second stability related to the support and relief of workers at the EUIF. To

I have become used to look at the... the data. It is good help to prepare before the.. the talk with clients. I think that it is a big help for us… and the citizen is more happy with us. The citizen, I think, ehh benefits from the information… for example can I look at his file and do some work before meeting him… ehm This makes the talk not so long. It can be that we only have a call and help him, and he does not come to a personal meeting. Without preparing I think I feel more anxious. But the data also makes me feel distracted or no focus.

5.2.1 Reinforcing "Public Values"

"I find it really motivating to see that peoples ehm to come to our service centre or call and that I can help them with the use of the tool ehm this motivates me and shows me how important this technology can be and how it can affect many lives for the better."

The above extract from the interview with the long acting and experienced consultant is a typical illustration of the findings concerning the perception of their work. The career counsellors and the EUIF in general are motivated to help and provide the best possible service to each induvial citizen, but as one anecdote showed, motivating the citizen to establish some form of everyday habit that will increase their likelihood of re-entering

the labour market or even lowering the chance of re-entering unemployment within the next years, the underlying vocabulary and design of the processes and the OTT system are aimed to mediate specific and therefore maybe biased sets of public values. The caseworker sometimes has the chance to perceive and understand the unemployed person on their own, but often times the use of specific language within the whole workflow reveals a hermeneutic relation in which the OTT system user understands or assesses the citizens behaviour only through the lens of the OTT system or other software in use.

Without being the data subject, the career counsellor might subconsciously notice the underlying incentive towards fulfilling the behaviourist public values that are mediated and feel like they are meaningful and worthwhile. The assumption of this finding is that users and their related surroundings, whether this means the unemployed citizens themselves or colleagues and friends, are influenced by the routine use of observable and quantifiable values in the citizens that align with the AI technology's designers' beliefs regarding the important factors of unemployment and may be more inclined and spread among those while other traits of unobserved people fail to be looked at. As a consequence, the constant present of easy and available information is subverting other more complex processes and used as additional important indicators of their future employment status. In a way the multi-stabile image of the data that is used distorts the human intuitional function while sharpening the stability of humans as data and information mediators.

The ethical implications that emerge around these issues highlight an ongoing alignment of the career counsellors' values with those of the OTT support tools design and thereby its designers. The literature review showed that there are several concerns to this type of ethical implications of AI design. The post phenomenological view and socio-material ethics perspective of the previously introduced script theory can be used to show how the consultant might adapt their personal behaviour and use of technology in accordance with the values and context of the OTT system. The idea is that the AI technology, as a result of the findings of this thesis, might lead to a scripted workflow that has a severe influence on the counsellors' interactions with the clients but also with his or her colleagues.

While this thought is at the moment rather abstract, the findings show that future development and technology integration is very likely and that suggests a strongly increased potential for technological devices and systems that might invite certain actions towards the counsellors and other employees at the EUIF. The theory behind scripted values and actions raises great questions concerning the social and political power relations in place and how those might influence the user's interaction and his perceptions. The also introduced network actor theory may help analyse the underlying

relations of the EUIF as a workplace and can be used to suggest that responsibilities are, as mentioned during the previous subchapters, tending to be shifted from the roles of human actors towards the artificial actors within the network. This interpretation of the theory also confirms possible assumptions that can be made regarding the ICT dependent structure of the EUIF. It becomes quite easy to say that it is not his or her fault, the system, machine or the underlying processes bound to this network, pushed him or her behaving a certain way. Ultimately this kind of power relation is not really considered to be conducive. To the contrary, if we think about the power relations of the technology, the literature already showed strong and well-established role of the private sector withing AI integration in the public sector, and this use case only confirms these results. The technology itself is provided by private stakeholder, as are the workshops and trainings of the personal to large amounts. This could mean integrating private values and work process subconsciously within the network of the EUIF institution.

"During a workshop we were told by an expert how to optimize the system and how.. how to explain it."

As part of this research, it became clear that not only the OTT tool, but many AI tools and systems are designed by private parties with their own specific biases. The biases are not always clear, but the findings showed little influence of the career counsellors towards the system design which indicates a strong power relation in regard to the authority stakeholders of this case. The user might imply his own reflection and might also be imaged as an authority figure towards the clients and enforcing their values through the technological lens of the OTT system during their consultations, job suggestions and other interactions.

5.2.2 Automated Decision Making and Hermeneutics

"I have not quite trusted the system in the beginning… and ehh have not taken too much information, respectively which were already available to me before. That is, most of the time, or rather the first weeks, I did my usual work and at the same time. Ehh… I worked on the cases with the OTT and mostly still used my own estimation to work on them. Relatively many details were better or more extensively better solved and recognized by the system."

The above extract shows and elaborates on the characteristic process that the career consultants encountered when introduced with the AI system. Therefore, this thesis tried to understand the issues concerned to the AI capabilities of categorization and prediction of certain actions and therefore behaviours of vulnerable citizens and how they would change the perception and experience of the user. The data that is obtained in order to do

so is gathered by widely scanning and saving huge amounts of personal information that is then processed with the aim of finding emerging patterns only visible to the AI but the not the human user. The “machine” in this sense is not only able to find patterns but use those to match with the citizens specific traits or properties in a way that would never be possible or in the realm of normal human capabilities. As the Nortal website advertises, the OTT would to be able to predict the citizens future employment status just by comparing their personal information and categorizing it with other similar cases already saved in the national register or gathered directly through the counsellors, ultimately feeding this technology. What does this imply on the abilities of the human user? The assumption is, that the system is basically able to see something in the personal data representation of a job seeking citizen while the human consultant might fail to notice something or not even see anything at all (Nortal 2021).

While the literature review and theoretical framework have already shown the importance of the ethical considerations that arise of the just explained tendency that comes with the AI technology the findings elaborated further on this. When thinking about the example of how glasses are the intermediaries between the human using it and his surrounding world, this metaphor of technology world mediation can be extended to our case of the OTT decision support tool. The underlying truth of the glasses is that they are intermediaries between information in the form of light that comes from the persons surroundings and makes him or her see these. The OTT is but an intermediary between the personal information of the unemployed person as well as various other information from the “world” and the career consultant. This is where the concept of hermeneutic applies to the OTT tool. It turns out to be that the technological device, that mediates the user’s perception regarding what the consultant “sees” in his client, is not so different from the way the glasses mediate what he or she might actually see in the visual sense.

Since the OTT tool has only been live for about two years, the career consultants until that point in time had looked on the citizens and labour market information, unguided and unaided. In post phenomenological terms this means that there was a more direct relation between the subject and the world described by the following diagram.

User – World

With the implementation and integration of the decision support tool, the AI technology of the OTT fulfils the partly role of the technological mediation. This might occur only when comparing the tool to the previously congenital methods, since the workflow and process inherited before the integration might have already included other forms of technological devices or in that sense software that could have fulfilled the functions of a technology mediation. Coming back to this papers case study the function of the

technological mediation of the OTT leads the hermeneutic aspects of post phenomenological theory to the relation of the user's perception. What the career consultant is confronted with when having to meet the clients, whether this happens digital or in person, direct or indirect, the data and categorizations of the system are processed and represented as a resulting "image' of the client to the user. While this might not be understood as a more precise description of the real world for the user, the enacted information of the world is expanded by the AI technology and therefore expands the profile of the citizens. This means that the users perception merges with that of the OTT tool creating new assumptions and behaviours.

These findings are interesting considering that the AI technology is not only limited to situations where the devices are the object of human interpretation, but the OTT tool actually extends the analysing capability to the AI technology in addition to the human interpretation of the unemployed citizens information. The underlying algorithm of the OTT support tool executes its own interpretative operations that allows the system to single handily categorize and create its own decision and suggestions to the user. While this task seems really impressive, it is something the consultant has previously done and continue to do on his or her own whether he or she is aware of it or not and therefore matches the behaviour and actions of the human user to a great degree.

The result of complex calculations performed by the algorithm is therefore an interpretation that has been achieved in a highly similar path to those done by the career consultants and his or her internal functions. Humans constantly perform tasks where they have to obtain information and single out cases where we seek to have proficient comparable personal experiences, perform estimations and then link them together to come up with our own perceptions. Decisions are based on this process of interpretation and while the designers of the OTT decision support tool claim its AI to be sophisticated and making all this on its own parallel to the human user, it alters the previously shown diagram as follows.

User – (AI – World) – World

In our case the career consultants' (User) experience their interaction with the citizen through the AI based OTT tool (AI) that has already done its own interpretation of the citizens (World) and created a new and different form of how the world is mediated to the human user influenced by the calculations and interpretations done within the algorithms of the AI technology.

5.2.3 Background Relation and Transparency

As mentioned in the theoretical framework of post phenomenology as a methodology, technologies can be described in a special sense of transparency. A sense that lets the technology become part of our surroundings. That it is a means of experience but nothing that we directly experience during the act of operation. As Verbeek put it, "…transparency concerns our awareness of the mediating role technologies play rather than our direct experience of the technologies themselves" (2012, p.394). In this sense the actual use of the OTT tool can be specified as the definite interaction of the user with the OTT.

"I think interactions with the customers are still good ehm although not the same as before, the system makes it easier for us to how do you say ehm … to kind of analyse the customer and their abilities on the work market so …. Ehm so to say we do have less work to do regarding this… thanks to the system but because we still interact with the customer face to face there still is a personal component that ehm."

The selected case implies that this will most likely mean a career counsellor using the technology and observing the citizens information, but it can of course also be the management or IT personnel analysing and examining specific data or processes related to the EUIF. The context of the first mentioned example would be defined as the context of the counsellor-citizen relationship and therefore the main emphasis according to Verbeek would lay on the awareness of the technology mediation and its role instead of focussing on any form of social-political aspect. What this means in our case, is that the background relation or transparency of the OTT tool implies that on the one hand the career counsellor and on the other the job seeking citizen might not be aware of the resulting impact that the OTT tool has on their counsellor-citizen relationship. What the findings showed was that the OTT itself was no part or considered subject of their interaction in form of conversation or other exchanges, but that the system itself and the use is not transparent. The career counsellor has to find the right information which implies direct interaction with the system through clicking or entering input. However, this relation is part of the general computer related situation and might not so be specifically technology related to the OTT.

While the application of the OTT system might not be transparent, the formerly mentioned context of the human world relation in from of the citizen-consultant relationship is accompanied by a high degree of transparency. The findings did not only show that the OTT is hardly ever mentioned or included into the actual interactions between the consultants and the citizens, but also showed that the users do have some form of awareness when considering the integration of OTT to their everyday workflow

and process as public service provider. But what does this mean? The computer or the phone are similar examples of technologies that are just as an important part of the interaction but likewise do not get actively mentioned or referred to during the consultation. And the information that is processed and presented by the tool may come up, but it will still uphold some form of black box around it, meaning that references are not always very specific nor part of the active discussion or written exchange.

This might go back to the consultant's assessment of the situation and while the information may support his counselling, showing the citizen his own data in other forms than job suggestions, might not lead to the wished-for outcome of the counselling interaction. Concluding the aspect of background relations and transparency, it can be said that while the software itself is little transparent to the user, it is very transparent to the citizen and the handled data might occasionally be opaque to the interaction, the overall impact remains transparent to both sides.

5.2.4 Thinking outside the "Black Box"

In addition to the citizen as a client and direct representation of the world within the introduced human technology world relations, the world doesn't end just there. There is much more to the world of the career consultant as a user of this technology. The findings also showed an interesting perspective on the user's perception on his environment. Data is just the means of a digital life; it usually seems to be something that is immaterial, only existing in the servers and clouds. And while the interviewees showed that the use of large amount of personal data was perceived as something worthwhile, the user fails to immediately acknowledge the fact that the OTT, the data, the servers running the clouds certainly have a very real and material footprint. This discussion of a false digital perception that is created through an intangible software lets the users believe that his or her actions are but a major part of saving important natural resources.

"The advantages, there are only advantages that we have... the data and we are data based and ehhhh ... we have ... if the... if the ja ... all the eeh services are offered to citizens based on the data we have about the customers and previously in old times… we have the solution where the customers or citizen should ehhh they should ask data from each institution on the paper and call the other institution and ask services and so the citizens should walk with the data from one institution to the other institution and now it is different the system that is data is changing.. is changing thing ... if a a customer citizen is come to us and wants to have a benefit or kind of service and we need the data for that, then we just asking from x road solution that the one register is asking from another register the data and how we we and have the data in right place. You get to offer the services. And that is a good thing and that that is the data is protected it is in the systems

it is not some kind of in the papers and walking around ... that is less papers and and walking around the papers."

But the "reality" of course can as easily appear to be the exact opposite. While the findings showed no indication among the interview partners that would suggest some form of mediated perception about the consequences the use of a large-scale automated system might have to environmental aspects, the intention of the system simply doesn't seem to be the mediation of a "world" like the material reality of mines, where for example rare earths are extracted which are needed to build important parts of the technological infrastructure as well as the inhumane work conditions that might be linked to this.

To be fair, this thought is very far removed from the comfort of the user's personal experience in his or her personal little environment, but the recent developments of climate change might alter this experience in the future. With regards to the necessary resources to perform the required tasks of the EUIF, the services appear to be constituting a seemingly digital realm for the users where dependencies of material resources are never mentioned. Whether this is something that happens with a degree of intentionality or is just one of the many variational experiences that seem to fall out of focus while designing or using the system, the excitement of new and thrilling technology definitely alters our conception of the physically necessities of this new world. Maybe it is important for the user to think more about the labour and extraction in material costs, that go in to the "digital world" providing the EUIF the services and support from technology that is of value to their services.

5.3 Digital Agenda running on Sensible Information

"Disadvantage eh… is a… the data protection is what we have some problem with. The… with the data exchange… with the eh sometimes we are getting too much."

Thinking about the personal perception and experience of the users, many seem to realize the impact that large amount of data might have to their lifeworld and how their services might be delivered, resulting in second thoughts like the one formulated by Mrs Rummo. When talking about data privacy, it becomes clear that with the use of large public registers a considered number of private portions of the citizens life, of their experiences (how long they have been unemployed or not) are increasingly made available to the users of systems like the OTT support tool. An interesting finding regarding this issue showed that the user might wonder, if some of those aspects of his or her own life might not better be left private and the autonomy of the data subject lets them feel that there is some kind of obligation towards them as data observers. The obligation to be able to say," I don't need that kind of data or interpretation (When is the citizen likely going to be unemployed

again) and maybe it should be better to learn and look away" is an interesting thought. Individuals might start considering if this could also have a misdemeanant effect on their personal interaction with the client. Or are the clients just a means to an end for the system pushing the Estonian public employment agenda? While this thesis did not focus on the goals and the overall agenda set by the EUIF, lowering nationwide numbers of unemployment can certainly be assumed to be a key priority. This position could have some severe ethical consequences to the interpretation of the mediated reality. For the OTT tool the data subjects are but raw material for a project, this means when mediating the perception of the surrounding world as something that is represented more and more by data, the human-technology world relations might become something disposed of a will. Which implies that "things in the form of data representations" are maybe only standing reserve to the EUIF processes and could in an exaggerated way objectify the clients and therefore denote them to no moral standing.

"Yes, sometimes I think there is too much artificial things, and we have built a lot of things around us. We can … can lose sight of this…these things and on the data and the artificial processes."

6 Discussion

6.1 Contributions

The purpose of this thesis was to understand the implications and mediated experiences of AI use in the specific context of the public sector providing citizens labour market services. Despite the limitations, at the end the beating heart of scientific thinking is not made up of certainties but rather thinking of the constant motion and power that is precisely the capacity to always question everything and begin over again. This thesis tried to do exactly that, approach this topic with the spirit to subvert the present state and search for a new or simply extending explanation of our world, only to then look out to the future and further research intriguing new questions and subverting it all over again. Nevertheless, the following key aspects can be named that helped contribute towards this goal:

- Redirecting the focus of current AI research to a more user centric approach in the public sector.
- Elaborating on the experience and possible use of post phenomenological methodology as an empirical study design in the field of modern technology.
- The identification of possible relationships between the main factors of the technology and the mediation of the user's world. At the heart of this relational interpretation of techno social theory is not a world we describe in terms of objects, things or entities in isolation but focussing on the continuous act of interaction upon each other including digital services.
- An external perspective on the very relations that manifest AI technology in the course of a human interaction, and which properties of this technology becomes actual, including ethical implications.

According to the pointed-out aspects, this thesis tried to create or partly create theoretical knowledge based on the user's perception and experiences with AI technologies to facilitate public labour services and describe the influence of this very technology on their own mediation of the world. The accumulated information of this process and possible future research could have severe influence not only of the philosophical discussion of technology but on the design of AI systems as well as ethical implications that could be considered in public AI policy developments.

The overall research goal was to answer the defined research questions. Since the underlying literature review has showed that the overall topic of AI in the public sector is

most recent and of great interest, it has not yet been thoroughly academically explored in a post-phenomenological way, nor has there been a lot of studies concerning socio technical implications to the public sector. That is why this thesis attempted be of some importance while also aiming at creating motion to this field of study. Henceforth, this work has clearly showed its contribution to the field of public sector innovations in the form of AI systems. The findings have indicated the many ways that advanced and intelligent support systems may influence and mediate our own personal impressions of ongoing real-life processes and interactions with citizens, co-workers and the technology itself. This work has also contributed to the understanding in regard to values and relationships that have been subject to this change. Furthermore, it is safe to say that based on the findings of this thesis there are several ethical implications that arise around this very aspect of the technology in use and related to Artificial intelligence, creating an import bridge between public AI studies and the socio technical research of technology.

6.2 Future Research

This thesis, as was just mentioned, showed the relation between the AI technology used in the EUIF and the users' perceptions and experiences of their reality. Future studies should include the exploration and measurement of technologically mediated perceptions and their relation to other possible variable that could be observed during this research. The repetition of similar approaches could also help validate and extend the results of this study. Furthermore, future studies in this field could be extended to several dimensions. First of all, examining how the OTT tool and its implications on the perception of the users and citizens could evolve in the future. This might include a more representative sample that might not only include more consultants and other users of the OTT but also the citizens as data subjects. But beyond the temporal dimension of future research, extending this field of research to other use cases within other domains and country specific contexts could have a great contribution towards the field of public sector AI research but also philosophy of technology.

Comparative studies might also include other areas of services that apply to greater areas of impact like facial recognition in public places. However, the continued exploration of how AI technology is utilized and affects the involved parties could help expand knowledge related to the understanding of human technology world relations. While automated decision making and support systems are represented only by the OTT tool in this study, there are many other systems and applications related to this with numerous amounts of areas of use. Not only does it make sense to study the future developments of as system that is constantly learning, but with the thought of Moore's laws it becomes obvious that the availability of more and more computational power for less resources

might also mean that related technology will spread on many dimensions including an organizational, geographical and cultural level. This ongoing digitalization and growth of AI technology will consequently bring forth several areas of possible future research that focuses on the philosophical implications of a world that might be on its best way towards loosing access to the real world.

6.3 Limitations

A common first limitation of this work is related to the nature of structured literature reviews. While using multiple online databases and digital libraries this paper cannot guarantee the entirety of the identified literature which was also never aimed to be exhaustive. The proceedings included initial and simple searches as well as structured and evaluated searching methods whilst selecting the source papers for the state of AI research in the public sector, but this still leaves the possibility of not capturing all the relevant studies and previous work, that has been done in the relevant field of this study. This restriction is reinforced by the lack of resources and the short period of time in which this study was conducted.

Another common constraint of this process is the author's bias. Although this thesis has tried to be done with an open and unbiased view on all the papers and insights, it is natural that the selection of relevant papers and the extraction of related data could have been influenced by personal and biased judgement. Driven by a genuine curiosity in understanding the effects of technology to the nature of reality, meant choosing a suiting approach to do it. This approach is heavily focussed on the human technology world relations that might help examine the consultants' experiences with the OTT support system, the underlying social and cultural decisions in addition to the ethical and technological implications to the field of public services and especially social services. This attempt therefore includes several research steps that are prone to limitations.

Semi structured interviews that are followed by a complex set of analytical measures trying to reflect upon the gathered materials, including the reassembling of individual anecdotes, meant that the research had to compromise on a small sample of participants limited withing the EUIF. Not only is the sample small but it is bound to the OTT support tool and therefore fails to represent various other systems that might show some similar features or are potentially completely different. The scope of this study has been carefully chosen in accordance with the available resources and therefore fails to acknowledge experiences, views and perceptions beyond those of the consultants like citizens and other stakeholders.

Further limitations arise around the possible occurrences that simply cannot be explained or generalized since the results are solely reflecting the consultant's perspective only. This means that the other strongly influenced people of this tool are out of any possible determination. Their opinions and views remain unknown as well as their degree of acknowledgement of the systems and their interaction with it. Other problems such as security and data privacy concerns by citizens are not gathered and can therefore not be represented by the outcomes of this study. The results are but only representing the experiences and insights of the professional users from the EUIF only.

7 Conclusion

This thesis tried to understand and elaborate the use of artificial intelligence in the public sector and what impacts and affordances this technology might have on the user and his or her relation to the world. This was done by comprehensively exploring the current state of research concerning public sector AI studies. The systematic literature review provided the basic oversight and knowledge to help structure this thesis's motivation and problem definition of a missing or at least underrepresented individual and more in-depth examination of the mediating role of AI. In order to achieve relevance to the explored field of study and appropriately addressing the framed problems, an extensive and comprehensive description of the underlying theoretical framework and methodology was provided. This approach included the choice of the post phenomenological methodology which allowed this paper to apply the empirical methods from a philosophical perspective and illustrate the variant prospects in which the AI technology influences scripted or limited extensions and affordances to the users.

While the chosen case study examined the OTT automated decision support tool as one example of an applied representative AI technology used in the public sector, the contributions and insights could also relate to similar areas of application and therefore many related instances in analogous environments. The following critical analysis through post phenomenological concepts contributed to the understanding of ethical implications related to the integration of the AI tool into the EUIF work practices and has shown that while many critical aspects are still overlooked the role of the user has not only slightly changed but his individual and collaborative experience of the workplace, his interactions and therefore in this sense his life world was influenced by this newly introduced technology.

Therefore, in this study the actual hands-on experiences of the users of the OTT tool were explored and included exposing the many ways in which the AI system and its data can be facilitated to support the work that is done by the EUIF, the influence this technology might have on their interactions, perceptions and the ethical implications that might arise. All this was done and addressed by gathering empirical data and information through conducting preliminary research and semi structured user interviews. The gathered data was further analysed using the theoretical methods of post phenomenological research approaches. Based on the findings of this thesis, prejudices concerning how reality is structured and made as the result of our individual experiences has helped understand the users of the OTT system at the EUIF and how they might apply and reflect on the AI technology and its possible mediating role on emergent factors and implications of an "artificial new lens". This experience has shown to be limited and while this study might

hind at a new lens through which we can “see” the world that is mediated by AI, there is no certainty when trying to comprehend it, but certainty that nourishes the search for future knowledge on this and other lenses constituting our Lifeworld.

Human beings are a part of nature and but one case among uncountable natural phenomena subject to the underlying laws of this very nature. Technology is made of the same matter and motion, achieving the case of artificial intelligence and the mediation of its own sensation of existence can amaze and motivate to understand how this process itself works and is viewed. The findings suggest that this artificial interpreted world is not yet as recognizable nor the exclusively correct perspective on the human life world. However, exploring this topic showed that AI technology used in this context has led to new facets of the citizens life that are immensely illuminated by the algorithm, making visible what might have been invisible to normal humans. The interaction between the artificial world of the data and the real world of the user allows the EUIF and the systems designer to manipulate the employee’s agency to his or her will. The AI might influence what the user sees and therefore ultimately what he or she focuses on. Which eventually means what he or she decides. The discovery of this function comes as no surprise, what could be explored during this research was, that the implications of a relationship between what we as humans see and the “world” are notable. This means that we are often times not really observing what we see but projecting our image of life and the world based on what we know. However, what we know has been without doubt influenced by the artificial technology including the unconsciously developing process of scrutinizing the world of the public and its citizen based on integrated misconceptions and biases provided by an alternative lens.

The system itself is often not perceived to have this big of an impact on the agency when compared to many other technologies. But what can be seen is that the user has been increasingly shaped by Estonia’s digital structure and reliance on data. The public sector institutions and governments are therefore the foundation for affecting the agency and behaviour of many citizens and employees though the integration of large amounts of data and AI technology. Affordances that arise are manifold. Examples of these are questions like; Is this particular system really measuring and interpreting what the citizen wants it to measure? Does the algorithm distort the users’ values or at least simplifies them to artificial metrics? And while there might be more questions, the public sector plays a crucial role in this sense that it might not always be clear what the real goal of a complex and intertwined state institution is. The findings showed that although the tool is praised to be of great public value, some users see the actual value over proportionally on the side of management and organizational responsibilities. Other problems that have been identified can be traced back to the actual transparency of this artificial world. Not

every user is or can be aware that he is changed by a system, that has an underlying structure of data and algorithms where many incentives and logics are hidden if not to the user than definitely to the data subject.

The users of the OTT and therefore the career counsellors are compelled by the thought that their work has been simplified, giving them, some feeling of knowing exactly what he or she is doing and why he or she is doing it in regard to their work at the EUIF. The management on the other hand would know exactly why this decision succeeded, framing the "automated" rules of probability to justify any technological transformation within their institution. Questions naturally arise around this very provoking interpretation of the thesis' final results. To what extend are these considerations obsessed with the importance of output? Of course, it is in the political interest of Estonia to lower the numbers of unemployment, and if we align with an industrialized and product-oriented world than these affordances are of no importance. However, on the individual scale users and this study can only assume citizens as well might question what is truly achieved and if it is a thing of true value. Like many other activities the value might not be in the output but in the process itself. Furthermore, the promise of offering the best service to the citizens, is defined as one of the major aims of the OTT decision tools designers. This study has shown that in the light of this promise the users were given real constraints concerning the lives of public citizens but, in a sense, only artificial objectives where they accepted the AI's design and henceforth its experience.

Moreover, on the thought of AI and its role as a means of agency itself, what is special to this use case is that there is not just this artificial world or environment of data related to the citizens and the user but that the AI system itself tells and defines the abilities and to an abstract extend predefines future obstacles citizens and people must confront like re-entering unemployment. The creation of a fundamentally alternative subject, or agent that is described by his or her scores and probabilities also means a system deciding on the very abilities of this agent and basically telling it what to do. The algorithm will match this agent with the perfect job, training, or other measure he or she might need, which might seem like a good thing. And without doubting that, this condition can likewise be interpreted from a different view whereas humans may perhaps experience a hard and unforgiving world. The AI technology and the mediated reality then somehow may possibly contribute to the user's partial relief from the negative perceived real world and might incentivise the users to seek refuge in the artificial world of pre-determined decisions. A mediated world where problems are clear and calculable, leaving them to expect this clarity to extend to the real world. Finally, if we merge this with the many other ongoing technological alterations and means of AI and how it is able to track and judge the behaviour of so many, the possibilities to modifications seem to be limitless.

References

Aagaard, J. (2017). Introducing postphenomenological research: A brief and selective sketch of phenomenological research methods. *International Journal of Qualitative Studies in Education*, *30*(6), 519-533.

Aagaard, J. (2018). Magnetic and multistable: Reinterpreting the affordances of educational technology. *International Journal of Educational Technology in Higher Education*, *15*(1), 1-10.

Aagaard, J., & Matthiesen, N. (2016). Methods of materiality: Participant observation and qualitative research in psychology. *Qualitative Research in Psychology*, *13*(1), 33-46.

Achterhuis, H. J. (2001). *American philosophy of technology: The empirical turn*. Indiana University Press.

Achterhuis, H. J. (2001). *American philosophy of technology: The empirical turn*. Indiana University Press.

Adams, C., & Thompson, T. L. (2016). *Researching a posthuman world: Interviews with digital objects*. Springer.

Adams, C., & Thompson, T. L. (2016). *Researching a posthuman world: Interviews with digital objects*. London: Palgrave Macmillan UK.

Adams, C., Aydin, C., Blond, L., Funk, M., Ihde, D., Petersén, M., ... & Turville, J. (2018). *Postphenomenological methodologies: New ways in mediating techno-human relationships*. Rowman & Littlefield.

Akrich, M. (1992). The description of technical objects. In Shaping technology/building society: Studies in sociotechnical change, edited by WE Bijker and J. Law, 205-224. Cambridge, MA: MIT Press.

Alkhatib, A., & Bernstein, M. (2019, May). Street-level algorithms: A theory at the gaps between policy and decisions. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (pp. 1-13).

Ammitzbøll Flügge, A., Hildebrandt, T., & Møller, N. H. (2021). Street-Level Algorithms and AI in Bureaucratic Decision-Making: A Caseworker Perspective. *Proceedings of the ACM on Human-Computer Interaction*, *5*(CSCW1), 1-23.

Araujo, T., Helberger, N., Kruikemeier, S., & de Vreese, C. H. (2020). In AI we trust? Perceptions about automated decision-making by artificial intelligence. *AI & SOCIETY*, *35*(3), 611-623.

Begg, R. (2008). Artificial intelligence techniques in medicine and health care. In *Intelligent Information Technologies: Concepts, Methodologies, Tools, and Applications* (pp. 1750-1757). IGI Global.

Bottenberg, F. (2015). Searching for alterity: what can we learn from interviewing humanoid robots?. *Postphenomenological investigations: Essays on human-technology relations*, 175-189.

Brayne, S., & Christin, A. (2021). Technologies of crime prediction: The reception of algorithms in policing and criminal courts. *Social Problems*, *68*(3), 608-624.

Buhr, D., Frankenberger, R., & Ludewig, A. (2016). On the Way to Welfare 4.0–Digitalisation in Estonia. *Friedrich-Ebert-Stiftung*.

Carl Gutwin and Saul Greenberg. 2002. A descriptive framework of workspace awareness for real-time groupware. Computer supported cooperative work 11, 3-4 (2002), 411–446

Chandler, D., Levitt, S. D., & List, J. A. (2011). Predicting and preventing shootings among at-risk youth. *American Economic Review*, *101*(3), 288-92.

Christin, A. (2020). The ethnographer and the algorithm: beyond the black box. *Theory and Society*, *49*(5), 897-918.

Chui, M. (2017). Artificial intelligence the next digital frontier. *McKinsey and Company Global Institute*, *47*(3.6).

Creswell, J. W., & Creswell, J. D. (2017). *Research design: Qualitative, quantitative, and mixed methods approaches*. Sage publications.

Dave, K. (2019). Systemic algorithmic harms. *Data and Society*, 1-2.

Dawes, R. M., Faust, D., & Meehl, P. E. (1989). Clinical versus actuarial judgment. *Science*, *243*(4899), 1668-1674.

de Sousa, W. G., de Melo, E. R. P., Bermejo, P. H. D. S., Farias, R. A. S., & Gomes, A. O. (2019). How and where is artificial intelligence in the public sector going? A literature review and research agenda. *Government Information Quarterly*, *36*(4), 101392.

Denzin, N. K., & Lincoln, Y. S. (Eds.). (2011). *The Sage handbook of qualitative research*. sage.

Desouza, K. (2018). Delivering artificial intelligence in government: Challenges and opportunities.

Desouza, K. C., Dawson, G. S., & Chenok, D. (2020). Designing, developing, and deploying artificial intelligence systems: Lessons from and for the public sector. *Business Horizons*, *63*(2), 205-213.

Dijkstra, J. J., Liebrand, W. B., & Timminga, E. (1998). Persuasiveness of expert systems. *Behaviour & Information Technology*, *17*(3), 155-163.

Dirican, C. (2015). The impacts of robotics, artificial intelligence on business and economics. *Procedia-Social and Behavioral Sciences*, *195*, 564-573.

Drechsler, W. & Kostakis, V. (2020). Is the Smart City a Good City? Published in Berkman Klein Center Collection, accessed on February 10th 2022 (https://medium.com/berkman-klein-center/is-the-smart-city-a-good-city-233a42bdcd46)

Dzindolet, M. T., Peterson, S. A., Pomranky, R. A., Pierce, L. G., & Beck, H. P. (2003). The role of trust in automation reliance. *International journal of human-computer studies*, *58*(6), 697-718.

Eggers, W. D., Schatsky, D., & Viechnicki, P. (2017). AI-augmented government. Using cognitive technologies to redesign public sector work. Deloitte Center for Government Insights, 1-24.

Ehsan, U., Liao, Q. V., Muller, M., Riedl, M. O., & Weisz, J. D. (2021, May). Expanding explainability: Towards social transparency in ai systems. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (pp. 1-19).

Elish, M. C., & Boyd, D. (2018). Situating methods in the magic of Big Data and AI. *Communication monographs*, *85*(1), 57-80.

Este, R. A. (2013). *Raising the alarm: an examination of innovation and philosophical denial*. Brill.

Eubanks, V. (2018). Automating inequality: how high-tech tools profile, police, and punish the poor. *St. Martin's Press*.

Feenberg, A. (2012). *Questioning technology*. Routledge.

Gasser, U., & Schmitt, C. (2020). The role of professional norms in the governance of artificial intelligence. In *The oxford handbook of ethics of AI* (p. 141). Oxford: Oxford University Press.

Gillespie, T. (2014). The relevance of algorithms. *Media technologies: Essays on communication, materiality, and society*, *167*(2014), 167.

Godin, B. (2008). Innovation: the History of a Category. Project on the Intellectual History of Innovation, Working Paper No. 1

Goldsmith, S., & Crawford, S. (2014). *The responsive city: Engaging communities through data-smart governance*. John Wiley & Sons.

Green, B., & Viljoen, S. (2020, January). Algorithmic realism: expanding the boundaries of algorithmic thought. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency* (pp. 19-31).

Green, B., & Viljoen, S. (2020, January). Algorithmic realism: expanding the boundaries of algorithmic thought. In *Proceedings of the 2020 conference on fairness, accountability, and transparency* (pp. 19-31).

Grigoli, F., Koczan, Z., & Topalova, P. (2020). Automation and labor force participation in advanced economies: Macro and micro evidence. *European Economic Review*, *126*, 103443.

Gulson, K. N., & Webb, P. T. (2017). Mapping an emergent field of 'computational education policy': Policy rationalities, prediction and data in the age of Artificial Intelligence. *Research in Education*, *98*(1), 14-26.

Hengstler, M., Enkel, E., & Duelli, S. (2016). Applied artificial intelligence and trust—The case of autonomous vehicles and medical assistance devices. *Technological Forecasting and Social Change*, *105*, 105-120.

Human Rights Watch. 2020. Automated Hardship - How the Tech-Driven Overhaul of the UK's Social Security System Worsens Poverty. September 2020

Husserl, E. (1970). *The crisis of European sciences and transcendental phenomenology: An introduction to phenomenological philosophy*. Northwestern University Press.

Ihde, D. (1986). Experimental phenomenology, second printing.

Ihde, D. (1990). Technology and the lifeworld: From garden to earth.

Ihde, D. Philosophy of Technology: An Introduction, Paragon House (1993).

Introna, L. (2017). Phenomenological approaches to ethics and information technology. *The Stanford Encyclopedia of Philosophy (Fall 2017 Edition)*

Irwin, S. (2005). Technological other/quasi other: Reflection on lived experience. *Human Studies*, *28*(4), 453-467.

Janssen, M., & Kuk, G. (2016). The challenges and limits of big data algorithms in technocratic governance. *Government Information Quarterly*, *33*(3), 371-377.

Janssen, M., & Van Der Voort, H. (2016). Adaptive governance: Towards a stable, accountable and responsive government.

Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, *1*(9), 389-399.

Johns, F. (2016). Data mining as global governance. The Oxford Handbook on the Law and Regulation of Technology

Katell, M., Young, M., Dailey, D., Herman, B., Guetler, V., Tam, A., ... & Krafft, P. M. (2020, January). Toward situated interventions for algorithmic equity: lessons from the field. In *Proceedings of the 2020 conference on fairness, accountability, and transparency* (pp. 45-55).

Kitchin, R. (2017). Thinking critically about and researching algorithms. *Information, communication & society*, *20*(1), 14-29.

Kulynych, B., Overdorf, R., Troncoso, C., & Gürses, S. (2020, January). POTs: protective optimization technologies. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency* (pp. 177-188).

Kuziemski, M., & Misuraca, G. (2020). AI governance in the public sector: Three tales from the frontiers of automated decision-making in democratic settings. *Telecommunications policy*, *44*(6), 101976.

L.M. Sacasas (2020). The Analog City and the Digital City, How online life breaks the old political order. Essay

Lee, M. K., Kusbit, D., Kahng, A., Kim, J. T., Yuan, X., Chan, A., ... & Procaccia, A. D. (2019). WeBuildAI: Participatory framework for algorithmic governance. *Proceedings of the ACM on Human-Computer Interaction*, *3*(CSCW), 1-35.

Leslie, D. (2019). Understanding artificial intelligence ethics and safety: A guide for the responsible design and implementation of AI systems in the public sector. *Available at SSRN 3403301*.

Levin, I., & Mamlok, D. (2021). Culture and Society in the Digital Age. *Information*, *12*(2), 68.

Logg, J. M., Minson, J. A., & Moore, D. A. (2019). Algorithm appreciation: People prefer algorithmic to human judgment. *Organizational Behavior and Human Decision Processes*, *151*, 90-103.

Loxha, A., & Morgandi, M. (2014). Profiling the Unemployed: a review of OECD experiences and implications for emerging economies.

Makarius, E. E., Mukherjee, D., Fox, J. D., & Fox, A. K. (2020). Rising with the machines: A sociotechnical framework for bringing artificial intelligence into the organization. *Journal of Business Research*, *120*, 262-273.

Makarius, E. E., Mukherjee, D., Fox, J. D., & Fox, A. K. (2020). Rising with the machines: A sociotechnical framework for bringing artificial intelligence into the organization. *Journal of Business Research*, *120*, 262-273.

Mazzucato, M. (2015). "6. Innovation, the State and Patient Capital". *The political quarterly*, *86*, 98-118.

McQuillan D., AI realism and structural alternatives, data justice Lab, June 7th 2019.

McQuillan, D. (2018). Rethinking AI through the politics of 1968. *Opendemocracy*.

McQuillan, D. (2019, November). AI Realism and structural alternatives. In *Data Justice Lab Workshop, Cardiff University. http://danmcquillan. io/ai_realism. html*.

Misuraca, G., Broster, D., & Centeno, C. (2012). Digital Europe 2030: Designing scenarios for ICT in future governance and policy making. *Government Information Quarterly*, *29*, S121-S131.

Mitchell, C., Meredith, P., Richardson, M., Greengross, P., & Smith, G. B. (2016). Reducing the number and impact of outbreaks of nosocomial viral gastroenteritis: time-series analysis of a multidimensional quality improvement initiative. *BMJ quality & safety*, *25*(6), 466-474.

Mohamed, S., Png, M. T., & Isaac, W. (2020). Decolonial AI: Decolonial theory as sociotechnical foresight in artificial intelligence. *Philosophy & Technology*, *33*(4), 659-684.

Molnar, P. (2019). Technology on the margins: AI and global migration management from a human rights perspective. *Cambridge International Law Journal*, *8*(2), 305-330.

Nortal (2021). OTT – an AI-powered success story in the public sector. Nortal HQ, accessed on 20.3.2022 https://nortal.com/blog/ott-an-ai-powered-success-story-in-the-public-sector/

Nuance Communications (2016). Australian taxation office continues transformation of the digital customer experience by launching online virtual assistant with Nuance. Accessed through website on the 13th of January 2022 https://www.nuance.com/about-us/newsroom/pressreleases/ato-launches-nuance-nina-virtual-assistant.html.

Pasquale, F. (2019). The second wave of algorithmic accountability. *Law and Political Economy*, *25*.

Petersen, A. C., Christensen, L. R., Harper, R., & Hildebrandt, T. (2021). " We Would Never Write That Down" Classifications of Unemployed and Data Challenges for AI. *Proceedings of the ACM on Human-Computer Interaction*, *5*(CSCW1), 1-26.

Petersen, A., Christensen, L. R., & Hildebrandt, T. T. (2020). The role of discretion in the age of automation. *Computer Supported Cooperative Work (CSCW)*, *29*(3), 303-333.

Pfotenhauer, S., & Jasanoff, S. (2017). Panacea or diagnosis? Imaginaries of innovation and the 'MIT model'in three political cultures. *Social studies of science*, *47*(6), 783-810.

Ransbotham, S., D. Kiron, P. Gerbert, and M. Reeves. "Reshaping business with artificial intelligence [reseach report]." *MIT Sloan Management Review* (2017).

Rosenberger, R. (2020). On variational cross-examination: a method for postphenomenological multistability. *AI & SOCIETY*, 1-14.

Rosenberger, R., & Verbeek, P. P. (2015). *Postphenomenological investigations: essays on human-technology relations*. Lexington Books.

Russell, S., & Norvig, P. (2005). AI a modern approach. *Learning*, *2*(3), 4.

Russell, S., & Norvig, P. (2020). Artificial intelligence-a modern approach 4th ed.

Šabanović, S. (2010). Robots in society, society in robots. *International Journal of Social Robotics*, *2*(4), 439-450.

Šabanović, S. (2010). Robots in society, society in robots. *International Journal of Social Robotics*, *2*(4), 439-450.

Sánchez-Monedero, J., Dencik, L., & Edwards, L. (2020, January). What does it mean to 'solve' the problem of discrimination in hiring? Social, technical and legal perspectives from the UK on automated hiring systems. In *Proceedings of the 2020 conference on fairness, accountability, and transparency* (pp. 458-468).

Sarcasas (2021). The Questions Concerning Technology, The Convivial Society: Vol. 2, No. 11 accessed on March 26th (https://theconvivialsociety.substack.com/p/the-questions-concerning-technology?s=r#details)

Saxena, D., Badillo-Urquiola, K., Wisniewski, P. J., & Guha, S. (2020, April). A human-centered review of algorithms used within the us child welfare system. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (pp. 1-15).

Schmidpeter, B., & Winter-Ebmer, R. (2018). How do automation and offshorability influence unemployment duration and subsequent job quality?. *Available at SSRN 3235595*.

Selbst, A. D., Boyd, D., Friedler, S. A., Venkatasubramanian, S., & Vertesi, J. (2019, January). Fairness and abstraction in sociotechnical systems. In *Proceedings of the conference on fairness, accountability, and transparency* (pp. 59-68).

Selbst, A. D., Boyd, D., Friedler, S. A., Venkatasubramanian, S., & Vertesi, J. (2019, January). Fairness and abstraction in sociotechnical systems. In *Proceedings of the conference on fairness, accountability, and transparency* (pp. 59-68).

Shneiderman, B. (2020). Human-centered artificial intelligence: Reliable, safe & trustworthy. *International Journal of Human–Computer Interaction*, *36*(6), 495-504.

Shneiderman, B. (2020). Human-centered artificial intelligence: Reliable, safe & trustworthy. *International Journal of Human–Computer Interaction*, *36*(6), 495-504.

Smith, A. (2018). Public attitudes toward computer algorithms. Pew Research Center

Star, S. L., & Strauss, A. (1999). Layers of silence, arenas of voice: The ecology of visible and invisible work. *Computer supported cooperative work (CSCW)*, *8*(1), 9-30.

Suchman, L. (1995). Making work visible. *Communications of the ACM*, *38*(9), 56-64.

Suchman, L. A. (1987). *Plans and situated actions: The problem of human-machine communication*. Cambridge university press.

Suchman, L., Blomberg, J., Orr, J. E., & Trigg, R. (2017). Reconstructing technologies as social practice. In *The Anthropology of Organisations* (pp. 431-447). Routledge.

Sun, T. Q., & Medaglia, R. (2019). Mapping the challenges of Artificial Intelligence in the public sector: Evidence from public healthcare. *Government Information Quarterly*, *36*(2), 368-383.

Suresh, H., & Guttag, J. (2021). A framework for understanding sources of harm throughout the machine learning life cycle. In *Equity and Access in Algorithms, Mechanisms, and Optimization* (pp. 1-9).

Suresh, H., & Guttag, J. V. (2019). A framework for understanding unintended consequences of machine learning. *arXiv preprint arXiv:1901.10002*, *2*.

Taylor, L. (2021). Public actors without public values: legitimacy, domination and the regulation of the technology sector. *Philosophy & technology*, 1-26.

The Economist (2016). China invents the digital totalitarian state. Retrieved from https://www.economist.com/news/briefing/21711902-worrying-implications-its-social-credit-project-china-invents-digital-totalitarian (accessed on 22.11.2021)

Vaughan, J. W., & Wallach, H. (2020). A human-centered agenda for intelligible machine learning. *Machines We Trust: Getting Along with Artificial Intelligence*.

Vaughan, J. W., & Wallach, H. (2020). A human-centered agenda for intelligible machine learning. *Machines We Trust: Getting Along with Artificial Intelligence*.

Veale, M., & Brass, I. (2019). Administration by algorithm? Public management meets public sector machine learning. *Public management meets public sector machine learning*.

Verbeek, P. P. (2005). Artifacts and attachment: A post-script philosophy of mediation. In *Inside the politics of technology* (pp. 125-146). Amsterdam University Press.

Verbeek, P. P. (2009). Cultivating humanity: Towards a non-humanist ethics of technology. In *New waves in philosophy of technology* (pp. 241-263). Palgrave Macmillan, London.

Verbeek, P. P. (2011). Moralizing technology. In *Moralizing Technology*. University of Chicago press.

von Schomberg, L., & Blok, V. (2021). Technology in the age of innovation: Responsible innovation as a new subdomain within the philosophy of technology. *Philosophy & Technology*, *34*(2), 309-323.

Võrk, A. (2021). Data & automated support tools Guest Lecture by Andres Võrk, Johan Skytte Institute of Political Studies, University of Tartu. E-Governance and E-Democracy at Tallinn University of Technology

Wang, D., Yang, Q., Abdul, A., & Lim, B. Y. (2019, May). Designing theory-driven user-centric explainable AI. In *Proceedings of the 2019 CHI conference on human factors in computing systems* (pp. 1-15).

Weber, M., Fivelsdal, E., & Østerberg, D. (1971). Makt og byråkrati-Essays om politikk og klasse, samfundsforskning og verdier (Power and bureaucracy-Essays about politics and class, social research and values). *Gyldendal, Norsk Forlag*.

Wellner, G. (2014). Multi-attention and the horcrux logic: Justifications for talking on the cell phone while driving. *Techné: Research in Philosophy and Technology*.

Whyte, K. P. (2015). What is multistability? A theory of the keystone concept of postphenomenological research. *Technoscience and postphenomenology: The Manhattan papers*, 69-81.

Wihlborg, E., Larsson, H., & Hedström, K. (2016, January). " The Computer Says No!"--A Case Study on Automated Decision-Making in Public Authorities. In *2016 49th Hawaii International Conference on System Sciences (HICSS)* (pp. 2903-2912). IEEE.

Wolf, C. T., & Blomberg, J. L. (2019, February). Intelligent Systems in Everyday Work Practices: Integrations and Sociotechnical Calibrations. In *International Conference on Intelligent Human Systems Integration* (pp. 546-550). Springer, Cham.

Wolf, C., & Blomberg, J. (2019). Evaluating the promise of human-algorithm collaborations in everyday work practices. *Proceedings of the ACM on Human-Computer Interaction*, *3*(CSCW), 1-23.

Yang, Q., Steinfeld, A., & Zimmerman, J. (2019, May). Unremarkable ai: Fitting

intelligent decision support into critical, clinical decision-making processes. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (pp. 1-11).

Yang, Q., Zimmerman, J., Steinfeld, A., Carey, L., & Antaki, J. F. (2016, May). Investigating the heart pump implant decision process: opportunities for decision support tools to help. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (pp. 4477-4488).

Yang, Q., Zimmerman, J., Steinfeld, A., Carey, L., & Antaki, J. F. (2016, May). Investigating the heart pump implant decision process: opportunities for decision support tools to help. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (pp. 4477-4488).

Young, M. M., Bullock, J. B., & Lecy, J. D. (2019). Artificial discretion as a tool of governance: a framework for understanding the impact of artificial intelligence on public administration. *Perspectives on Public Management and Governance*, *2*(4), 301-313.

Yuen, C. L. (2021). Exploring Teachers' Experiences of Using ClassDojo: A Postphenomenological Study.

Appendices

A Interview Prompts

Part 1: What is the caseworker/counsellors public context?

- What kind of work do you facilitate?
- How many citizens do you simultaneously work with?
- How long have you been working at the EUIF? Using OTT?
- How is your workplace structured?

Part 2: In what ways, and to what extent, does the caseworker/counsellor adopt OTT to facilitate services to citizens?

- Do you recall how you found out about OTT? What interested you about it?
- What were your first impressions about using OTT?
- What considerations or factors influenced your decision about when and how to use OTT as part of your work?
- How did you learn to use OTT initially?
- How do you use OTT? Can you elaborate on your primary and peripheral uses of OTT?
- What is your rationale or philosophy behind using OTT? (e.g., external or internal motivators, reflections)

Part 3: How does the caseworker/counsellor perceive OTT and its impact on employment services?

- How do you think using OTT affects the flow or pace of your work?
- How do you think OTT affects your interactions or communication with citizens and co-workers? (e.g., empathy, suitability, consistency, feedback)
- What habits will the use of this technology instill?
- How do you think OTT affects co worker interactions?
- Have you changed or adapted your working practices or use of OTT in any way because of your experience with it? Any particular events that stand out?
- What practices will the use of this technology displace?
- What practices will the use of this technology cultivate?
- What will the use of this technology encourage you to notice?
- What will the use of this technology encourage me to ignore?
- How do you think OTT supports or changes the type of activities of citizens on the labor market?
- How do you think OTT affects behavioral or management practices at your workplace?
- How does this technology empower you? At whose expense?
- Can you be held responsible for the actions which this technology empowers? Would you feel better if you couldn't?
- How do you think OTT affects your overall well-being? How about your work experiences? How does OTT impact your confidence in working or using technology? What about your feelings of competency in working with or using technology?
- What would the world be like if everyone used this technology exactly as you use it?
- What risks will my use of this technology entail for others? Have they consented?
- Can the consequences of your use of this technology be undone? Can I live with those consequences?
- What limits does the use of this technology impose upon you?