



Explaining It Your Way - Findings from a Co-Creative Design Workshop on Designing XAI Applications with AI End-Users from the Public Sector

Katharina Weitz
University of Augsburg
Augsburg, Germany
katharina.weitz@informatik.uni-augsburg.de

Ruben Schlagowski
University of Augsburg
Augsburg, Germany
ruben.schlagowski@uni-a.de

Elisabeth André
University of Augsburg
Augsburg, Germany
elisabeth.andre@uni-a.de

Maris Männiste
University of Tartu
Tartu, Estonia
maris.manniste@ut.ee

Ceenu George*
TU Berlin
Berlin, Germany
ceenu.george@tu-berlin.de

ABSTRACT

Human-Centered AI prioritizes end-users' needs like transparency and usability. This is vital for applications that affect people's everyday lives, such as social assessment tasks in the public sector. This paper discusses our pioneering effort to involve public sector AI users in XAI application design through a co-creative workshop with unemployment consultants from Estonia. The workshop's objectives were identifying user needs and creating novel XAI interfaces for the used AI system. As a result of our user-centered design approach, consultants were able to develop AI interface prototypes that would support them in creating success stories for their clients by getting detailed feedback and suggestions. We present a discussion on the value of co-creative design methods with end-users working in the public sector to improve AI application design and provide a summary of recommendations for practitioners and researchers working on AI systems in the public sector.

CCS CONCEPTS

• **Human-centered computing** → **Empirical studies in HCI**; **User centered design**; **User studies**.

KEYWORDS

Human-Centered AI, Explainable AI, User-Centered Design, Co-Creation, Focus Group, Social Assessment, Unemployment Insurance

*The paper originated during the author's guest professorship at the University of Augsburg



This work is licensed under a Creative Commons Attribution International 4.0 License.

CHI '24, May 11–16, 2024, Honolulu, HI, USA
© 2024 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-0330-0/24/05
<https://doi.org/10.1145/3613904.3642563>

ACM Reference Format:

Katharina Weitz, Ruben Schlagowski, Elisabeth André, Maris Männiste, and Ceenu George. 2024. Explaining It Your Way - Findings from a Co-Creative Design Workshop on Designing XAI Applications with AI End-Users from the Public Sector. In *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '24)*, May 11–16, 2024, Honolulu, HI, USA. ACM, New York, NY, USA, 14 pages. <https://doi.org/10.1145/3613904.3642563>

1 INTRODUCTION

Artificial Intelligence (AI) has emerged as a transformative force with the potential to change numerous aspects of modern society. The public sector and government are among the domains where AI holds exceptional promise. Governments worldwide tackle complex challenges ranging from service delivery optimization to data-driven decision-making [15]. Integrating AI technologies presents a remarkable opportunity to enhance operational efficiency, responsiveness, and citizen engagement. While AI can yield remarkable outcomes, its inherently complex and opaque nature can create barriers to understanding, accountability, and public trust [1]. Explainable Artificial Intelligence (XAI) seeks to alleviate these concerns by making AI systems interpretable, allowing developers and end-users to understand AI decisions [1]. This interpretability is not just a technical endeavor but embodies a broader societal need for accessible, understandable, and verifiable governance systems. However, the successful implementation of XAI in the public sector extends beyond technological competence; it necessitates the active involvement of stakeholders other than solely AI developers [19]. The collaboration between technologists, policymakers, domain experts, and citizens is paramount to ensure that AI-driven initiatives align with societal values, regulatory frameworks, and the unique needs of the public. Involving such stakeholders in co-creative XAI design processes can enrich the development and deployment of AI applications and foster a sense of ownership, transparency, and accountability for AI systems and their decisions [18].

In this paper, we present the results of a co-creative design workshop with AI-experienced unemployment consultants ($N = 5$) to shed light on problems and possible designs for XAI in the context of social service provision in Estonia. In Estonia, AI applications¹

¹In Estonia, over 50 AI-based tools are already in use (see: <https://e-estonia.com/ai-to-help-serve-the-estonian-unemployed/>)

are already used in the public sector [29]. With its groundbreaking e-government strategy, the Estonian digitalization approach leads innovation globally [29]. By exploring XAI's potential in social service provision, we aim to illuminate the significance of co-creative design approaches that address civil servants. Starting from the idea of Human-Centered AI (HCAI), that AI should be oriented towards the needs of humans [43], we used a classic of empirical HCI research: the user-centered design process. We combine this with the question-driven XAI design process of Liao et al. [30] to develop paper prototypes of XAI interfaces. We highlight the tangible benefits of including end-users in developing highly specialized and data-sensitive XAI designs in public sector settings.

In a co-creative design workshop, we investigated with users the *Otsustustugi* (OTT)² system. OTT is a random forest-based AI tool that predicts the probability of unemployed persons finding a job. Study participants identified current problems and prototyped possible design solutions to improve this system. For our co-creative workshop, we addressed the following research questions:

- **RQ1:** How and in which specific situations/contexts do **challenges and problems** arise while using the OTT system?
- **RQ2:** How can we **improve user acceptance** for the OTT system with a focus on explainability, fairness, control, and ethics?
- **RQ3:** How could **prototypes** look like that tackle the identified issues?

As AI-based social service applications are challenging to access for research, this paper presents insights and findings from a rare collaborative effort with professional AI end-users from the public domain, unemployment consultants. They support their clients, namely unemployed people, in finding a job. We illustrate how HCI practitioners can effectively use a co-creative approach to develop prototypes of XAI interfaces. Furthermore, our research highlights the pivotal role of XAI as an augmentative tool that can help unemployment consultants to evaluate and select the most appropriate measures to help their clients. With the presented co-creative design workshop on cooperation between human expertise and AI in labor market services, we provide recommendations for future AI applications and their development for the public sector.

2 RELATED WORK

2.1 AI in the Public Sector

The usage of AI in the public sector has various opportunities. Wirtz et al. [48] identified ten application purposes and their value for the public sector. For example, virtual agents (e.g., chatbots) can enhance citizen services, offering faster responses and personalized experiences [21, 48]. In addition, AI can automate routine tasks, optimize resource allocation, and streamline processes, leading to significant cost savings and improved service delivery [11]. Here, one use case for AI in the public sector is its usage as a profiling tool. Profiling helps social services assess individuals' needs more efficiently and supports the job-finding process [5]. For example, profiling can identify unemployed people with a higher risk of long-term unemployment and support them with more costly and

intensive services [10]. Including AI in the profiling process raises issues about fairness [14], transparency [6], responsibility, and control [21]. For example, people belonging to a minority (e.g., foreigners) are more likely to be misclassified as high-risk job seekers [11].

While the interest in AI as part of e-government is growing in Europe [38], a user-centered design of the systems that support civil servants is still missing. An AI application already used in the public sector is needed to investigate user needs in a real-world setting. Estonia is one of the leading countries in using the AI-profiling tool OTT in the public sector to predict the probability of an unemployed person finding a job and the probability of becoming unemployed again [41].

As Estonia is one of the first countries to use AI in the public sector, it can serve as a role model for other countries. The findings from our workshop provide first indications about users' needs when using AI in profiling and can support researchers in other countries in developing their own AI systems for public services.

2.2 Explainable AI

Humans have a natural need to acquire and provide explanations. Thus, children already ask “*why?*” and try to gain explanations for phenomena they observe [27]. As AI systems become more and more prevalent in our lives, the explanation of why something was done is also addressed to AI systems. The research area of XAI deals with AI explanations. The goal of XAI is to help users “to understand, appropriately trust, and effectively manage [...] artificially intelligent partners” [19, p. 44]. This goal gained renewed momentum with the increasingly widespread use of deep neural networks, which, as black boxes, left it unclear how they come to their classification decisions. At the beginning of XAI for deep neural networks, research focused on possible algorithms to light the black box. *Feature Relevance* methods focus on letting users of the XAI system know which features in the input data are crucial for a decision [3] (e.g., Layerwise Relevance Propagation (LRP) [4], SHAP [33], or LIME [42]). Another set of methods, inspired by how humans explain things, employs the idea of *Counterfactual Reasoning* [37]. Counterfactual Reasoning methods tackle the question “What if...?” by showing an alternate reality and the AI's decision in that scenario [36]. A distinction is drawn between local and global explanations [32]. *Local explanations* focus on providing insights into the decision-making process of an AI model for a specific instance or prediction, offering a fine-grained understanding of the model's behavior for individual cases. On the other hand, global explanations aim to convey a broader overview of the model's functionality, highlighting patterns and trends across the entire dataset. For some years now, another relevant aspect has been coming into focus: the design of XAI systems. But how should we design XAI systems? On the one hand, there is the question of algorithmic feasibility (i.e., *explainable model*) [19], especially for black box models; on the other hand, it is unclear how explanations of AI systems are best communicated to humans (i.e., *explanation interface*) [19]. What such an explanation interface of AI systems should look like is a much-discussed topic and occupies researchers on XAI. One could argue that the findings in human-human explanations could be transferred to the area of XAI. The question arises:

²OTT means “decision support” in Estonian and is also a common boy's name in Estonia

“Should AI communicate explanations as humans do?” Shneiderman [45] discourages such transfers, as interactions with machines are different and more limited than with humans. Nevertheless, the findings from psychological research are an essential starting point for the design of XAI [37]. For example, Miller [37] stated that users do not expect all possible explanations for an event. These insights can serve as a valuable basis for investigating the effect of AI explanations in user studies.

In our co-creative design workshop, we investigate which kinds of explanations unemployment consultants who work with AI regularly find helpful and how they should be implemented.

2.3 Explainable AI in the Public Sector

Despite considerable research on deploying AI within the public sector, work investigating the potential of XAI in the public sector is rare. The current research landscape predominantly explores the application of existing XAI techniques, such as LIME [42] or SHAP [33], within public sector scenarios. These include the operations of German tax authorities [35], forecasting for municipal wastewater treatment facilities in Greater Cincinnati [34], and the utilization of Linked Open Government Data by the Scottish Government [25]. Adopting XAI in these settings is motivated by a dual objective: firstly, to enhance the transparency of AI systems for professionals within the public sector, and secondly, to validate resource utilization and decision-making processes to the public. In exploring the application of XAI within the public sector, de Bruijn et al. [9] highlight the inherent challenges in deploying XAI in representative case studies, such as decisions on immigration, especially considering that its decisions may not always align with public consensus.

In contrast to merely applying pre-existing XAI methods to AI systems in the public domain, we advocate for a co-creative design strategy that emphasizes the active involvement of stakeholders and encourages fresh ideas. To illustrate the methodology, the example of Estonia is particularly adequate as a country with advanced digitalization processes.

3 THEORETICAL BACKGROUND

3.1 User-Centered Design Process

The user-centered design process is a classic empirical approach from HCI, first proposed by Norman and Draper [40]. It is defined in the ISO standard 9241.210:2019: “Ergonomics of human-system interaction - Human-centred design for interactive systems” and comprises four phases:

- **Understand the Context of Use:** The first phase focuses on defining the users and their tasks in a particular context (social and technical). Questions such as “Who are the main users of the system?”, “What tasks are solved with the help of the tool?”, and “What do users’ work processes look like?” should be answered here.
- **Specify Requirements:** When it is clear who the user group is and the context of use has been identified, the user group’s requirements are considered in more detail. For this purpose, so-called personas (i.e., prototypical users) are defined, and scenarios of use for the tool are sketched. Here, a distinction is made between the tasks the system takes on and those

the user takes on. In addition, usability requirements can be addressed (e.g. “How important is user satisfaction?”, “How important is the system’s flexibility?”).

- **Develop Design Solutions:** After the requirements and possible problems have been defined, this step collects ideas for possible solutions. Here, different design teams develop different solutions. Methodologically, various options can be used here, e.g., storyboards that outline the interaction between system and user or the design of paper or software prototypes.
- **Evaluate Design Solutions:** Different evaluation approaches can be distinguished here: In expert evaluation, as the name suggests, experts give feedback on the design solutions developed. These can be software experts who identify the technical feasibility or possible weaknesses or domain experts who know the application context and workflows well and assess which problems could occur when using the design solution. In addition to expert evaluation, however, controlled experiments can also be conducted with the design prototype. Another option is a participatory evaluation where real users evaluate the prototype in a real-world setting.

Elements of the user-centered design can be found in the participatory design approach, which actively includes users in an iterative, participatory process when developing or reworking a product [46]. Zhang and Zurlo [49] highlight the importance of participant engagement during a participatory design process. They describe that the engagement of participants has an emotional (e.g., interest, boredom, stress), cognitive (e.g., awareness, effort), and behavioral (i.e., physical actions) component that should be considered during the design process. Co-creative approaches foster innovation and user-focused research by including users in the development process [18]. The value of co-creative approaches are, for example, the shifting of power dynamics (e.g., from policymakers to users of systems) and the enhancement of outcomes (e.g., increased quality, novel products, innovations) [47].

3.2 Co-Creation Methodologies for the Design of XAI Systems

As mentioned at the beginning, insights from the social sciences provide valuable anchor points for the design and study of XAI. While Miller [37] provides a comprehensive insight into the overlap between social sciences and XAI research, the work of Hoffman et al. [23] describes how psychological constructs such as trust or curiosity can serve to investigate the impact of AI explanations on users using different methods (the simplest being the use of questionnaires). Another aspect of the investigation is satisfaction with the explanations provided. Gunning and Aha [19] highlight that psychological constructs such as trust and user satisfaction are relevant for measuring the effectiveness of XAI.

However, we need to address the question of which explanations users would like, which involves incorporating users’ perspectives. This means when focusing on XAI in a human-centered way, more is needed than investigating explanations’ impact on users. To understand the XAI’s usability, benefits, and downsides, end-users must become part of the design process. For this, a broad research community agrees that stakeholders (e.g., their needs,

Probability moving to employment	Level of probability moving to employment	Position in advisor portfolio	Date of evaluation	Advisor feedback	Factors affecting the probability of moving into employment	Level of probability of reinstatement	Period of unemployment
22 %	Low	28	01.02.2020	Add feedback	See (opens 10 main factors)	Medium	01.01.2020-...

Figure 1: An example of an OTT output (translated version) that is part of the EMPIS software that unemployment consultants use daily. The example shows the entry for one client, who is predicted from OTT, to have a low probability of finding a job again.

mental models, experiences), as well as the purposes of different AI application scenarios (e.g., healthcare, military, sales, finance), have to be taken into account when creating XAI [12, 16, 17, 22, 37]. Different approaches are developed to co-creatively design explanations in a human-centered way. Schoonderwoerd et al. [44] present the DoReMi-practice for human-centered design. Their approach consists of three components (i.e., domain analysis, requirements elicitation & assessment, and multi-modal interaction design & evaluation). Although their approach involves users as an active element in their process, Schoonderwoerd et al. [44] leave the user interface design to the researchers while users evaluate it. The work of Liao et al. [30] presents a question-driven process to design XAI. It consists of four steps to match users' questions towards an AI system with the respective XAI methods. In their four-step approach, users contribute in the first step (i.e., question elicitation) and the last (i.e., iterative design and evaluation). Similar to Schoonderwoerd et al. [44], users did not design the XAI prototype by themselves.

Co-creative approaches provide an opportunity to address the challenges of XAI by involving diverse users and other stakeholders in a collaborative and iterative process. The rarity of co-creative efforts in XAI design underscores an untapped potential for enhancing an AI system's interpretability and trustworthiness. By using co-creative strategies, the development of XAI solutions can benefit from a rich interplay of insights, expertise, and user feedback, resulting in more effective and user-friendly XAI systems.

Therefore, we present a co-creative design approach to investigate XAI design from the end-users' view (in our case, unemployment consultants). For this purpose, we use the user-centered design process [40] combined with the question-driven process to design XAI from Liao et al. [30]. The involvement of AI end-users from the public sector in designing an XAI interface is the main focus of our paper and will be described in the following.

4 METHODOLOGY

4.1 Apparatus: Estonia's OTT System

Since 2000, Estonia has been a leader in establishing an e-government strategy in the public sector [2]. In Estonia, labor market services are provided to the unemployed and job seekers to find work, promote career development, foster the professional development of workers, and attract skilled labor for employers. The Estonian Unemployment Insurance Fund (EUIF)³ was established in 2001 to

administer unemployment insurance benefits. EUIF's primary objective is to handle unemployment-related social insurance and provide services that help unemployed people find new jobs. Therefore, EUIF's clients are job seekers, employees, and employers.

OTT is a data-driven tool included in the Employment Information System (EMPIS) used by EUIF since 2020. OTT predicts the probability that an unemployed person will find a job within 180 days and identifies the factors that influence this [41] (see Figure 1). In total, 45 factors (e.g., level of education, language skills, region, driver's license, unemployment spells in the last three years) are deemed significant to predict this probability [28].

The model calculates the forecast for transition into employment for each newly registered person unemployed for 35 days. To this end, OTT summarises a person's situation using a random forest-based machine learning model. It uses 60 attributes and indicators to assess each unemployed person who turns to EUIF [41]. Based on the last five years of unemployment register data, it predicts the probability of getting into work during the year⁴. In addition, it calculates the probability of becoming unemployed again and identifies the circumstances that influence this. In this way, OTT is designed to support unemployment consultants by providing actions to meet their clients' individual needs and increase the efficiency of the EUIF.

Two types of stakeholders use OTT at EUIF: employment consultants and case managers. The former use OTT to decide on the distribution of working time of case managers who serve the clients of the social welfare system. The latter use OTT specifically during individual consultations to decide on measures that might be helpful for a client (e.g., language training and driving license). However, while the two stakeholders have different roles, the use cases are relatively similar.

4.2 Co-Creative Design Workshop

The following details the co-creative workshop conducted with unemployment consultants ($N = 5$). The primary objective of this workshop was to design XAI prototypes for the AI-based OTT software. Guided by a user-centric design approach combined with a question-driven XAI design process, the workshop facilitated the collaborative development of potential XAI interfaces specific to OTT.

⁴Description of OTT provided by NORTAL, a company that was responsible for the development of OTT (see: <https://nortal.com/insights/ott-an-ai-powered-success-story-in-the-public-sector>)

³Estonian name: Eesti Töötukassa (see: <https://www.tootukassa.ee/en>)

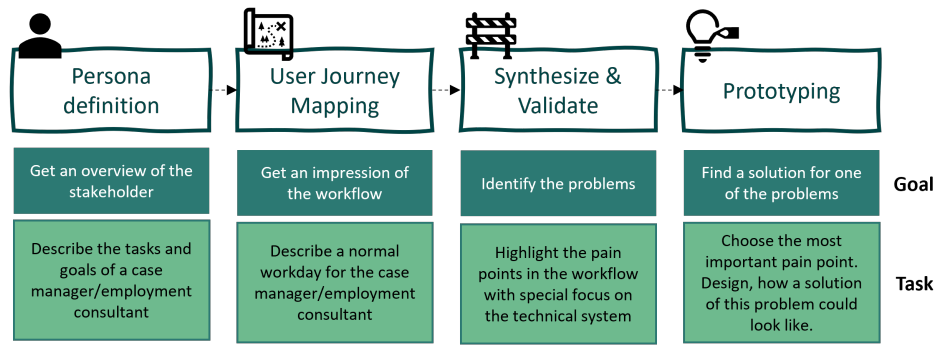


Figure 2: Four steps of the co-creative design workshop we conducted with civil servants of the labor market services in Estonia. The goal was to develop XAI interfaces for the AI-based software OTT.

4.2.1 Participants. To get a complete picture of the stakeholders of OTT, we invited both stakeholder groups to our co-creative design workshop. Five female stakeholders (two employment consultants and three case managers) between 32 and 49 years old participated in our workshop. All participants had some years of experience in their working positions (between 3 and 5 years) and were from the same EUIF department. All participants spoke English. They were recruited by the University of Tartu, who have already collaborated in other research projects on digitalization and AI. The data protection officer of the University of Tartu approved the workshop. Before the workshop started, participants were informed about the goals and duration of the workshop and their GDPR rights. All participants were reimbursed for the day with 400 Euros each. For better readability, our workshop participants will be referred to as *unemployment consultants*, including case managers.

4.2.2 Procedure. The one-day co-creative workshop took place at the University of Tartu, Estonia and lasted from 9:00 to 15:45, including a lunch break and two coffee breaks (one in the morning and one in the afternoon). A team of five researchers were present the whole day to conduct the workshop. One of them was leading the co-creative design process, two actively participated and guided the two sub-groups during the design process, one observed the workshop and took notes, and the last was supporting with translation.

The workshop started with welcoming all participants, introducing the research team and the research focus, and a short round of getting to know each other. Then, the co-creative approach began with four steps and a closing session, including the focus group interview (see Figure 2 for an overview). Unless otherwise stated, all steps took place in small groups (two groups: 1x three persons, 1x two persons).

Persona Definition. Participants were separated into small groups due to their working profiles (employment consultants and case managers). After a short introduction to the design of personas, they developed two personas that represent their jobs to be done during the day, their motivation, and the pain points during the workday.

User Journey Mapping. Based on the defined persona, participants described the day of the persona in more detail in the next

step. This step answered the question: “What does the typical persona workflow look like?” For this, a concrete goal of the persona was defined (e.g., to find a suitable job for the client), and all tasks during the day to reach the goal were described. In addition, tools that are used for the tasks were also collected (see Figure 4).

Synthesize & Validate. After defining a typical workflow, the focus was on the problems during this workflow. Here, pain points related to the technical systems to be used in general, OTT in particular, and problems related to the client were investigated in more detail (see Figure 5).

Prototyping. While the morning was used to get an overview of the context of the OTT use from a persona perspective, the afternoon was used to develop prototypes for one identified problem (see Figure 6). Each group selected one pain point they wanted to design a prototype solution for. After developing a first version of the paper prototype (see Figure 7), the two groups presented these to each other and received feedback on their prototype. In a second iteration, the feedback was incorporated into the design of a second prototype.

At the end of the co-creative workshop, all participants and the researchers had a focus group discussion about OTT and their impressions of the workshop.

4.3 Data Collection & Data Analysis

In this section, we outline the methodology employed for data collection and analysis in our study, which aimed to investigate and enhance AI-supported social service provision through a co-creative design workshop and a focus group interview involving unemployment consultants. Our research design aimed to gain in-depth insights into these AI users’ perceptions, experiences, and perspectives in social service delivery, thereby contributing to a comprehensive understanding of the dynamics and challenges inherent in providing these services.

4.3.1 Workshop. By involving unemployment consultants in the creative process, our methodology aimed to empower these AI end-users to contribute to developing AI systems that align more closely with their needs and values. Data collection during the co-creative workshop encompassed various artifacts, including sketches, paper prototypes, and written notes. During recruitment and planning,

	Mary	Melissa
Demographics	<ul style="list-style-type: none"> • 35 years • Employment consultant • Bachelors' degree • Good communicator & multitasker 	<ul style="list-style-type: none"> • Between 30 to 35 years • Case manager • 3 years of experience in the job • Creating relationships are important
Goals	<ul style="list-style-type: none"> • Be good at the job • Find strengths and weaknesses of clients • Support clients with services 	<ul style="list-style-type: none"> • Create success stories • Work efficient with colleagues • Human interaction & valuable relationship with clients

Figure 3: The participants created two personas: Mary, an employment consultant and Melissa, a case manager.

it was agreed with the local researcher to avoid audio recording equipment in the workshop but instead take written notes. Experimenters presumed that this would help foster communication and trust, as it would leave room for techniques such as active listening, small talk, and establishing rapport. Therefore, one observer and two moderators took notes for every workshop step (see Figure 2). We utilized an observer sheet (see Supplementary Material) consisting of open-ended questions and tables to count and write down information that outlined specific focus topics, such as explainability, trust, fairness, control, and ethics. The sheet provided space for additional observations at each phase of the workshop. We ensured consistency by distributing the same observation sheets to the observer and moderators of both participant groups, allowing them to take notes during the workshop within the defined parameters. This meticulous documentation aimed to enhance the comprehensiveness and reliability of our observational data. By allowing one observer and both moderators of the small groups to document their observations, three distinct perspectives were considered to ensure the acquisition of results and reduce the impact of individual biases.

The observer and moderators' written field notes during the workshop formed the basis for the subsequent analysis. First, an independent evaluator, who was not involved in the main study, digitized and aggregated the written notes from the observer sheets. Thematic analysis [7] was employed to identify overarching themes, divergent opinions, and patterns that emerged during the group interactions. Since the questions in the sheet were related to our three research questions, namely to identify challenges & problems, improve user acceptance, and prototype recommendations, they served as pre-defined codes during thematic analysis. Analysis was done in two iterations: After the first iteration, a discussion among all study experimenters was conducted to discuss initial results and additional emerging sub-themes. This was followed by a second iteration. For example, based on the notes, for RQ1 (challenges & problems), we found four sub-themes: (1) client-related issues, (2) issues with the OTT system and (3) with the interface, and (4) issues with missing data.

In addition to our thematic analysis reports in the result section, we state participants' quotes we noted while observing the two small groups during the workshop. When quoting these handwritten statements, we refer to the respective participant group (e.g., P in group 1).

Regarding XAI, which was the focus of our workshop, we used steps one and two from the question-driven XAI design process [30]. For step one, question elicitation, we collected questions during the workshop the participants would like to ask OTT. In step two, the question analysis, we categorized the questions using the XAI categories from Liao et al. [30]. We found several questions regarding the XAI category "Why". Participants, for example, asked the OTT: "Why did I get this prediction?" From this question, we defined the requirement to "explain the reasons for the prediction" that the XAI interface design should address. Following this process, we identified the user requirements for the XAI prototype that formed the basis for the prototype session.

4.3.2 Focus Group Interview. Complementing the co-creative workshop, a focus group discussion was conducted to facilitate dynamic interactions among participants and generate collective insights. Focus groups are particularly valuable for capturing group dynamics, reflection, exploring consensus, and uncovering divergent viewpoints. The focus group was audio-recorded after obtaining the participants' oral consent at the session's start. While transcribing and analyzing the audio recordings, experimenters noticed that the majority of themes were re-emerging from the workshop itself. This was not a surprise, given the nature of the methodology and placement at the end of the day. All quotes are provided in the supplementary material, but only quotes for newly emerged themes are mentioned in the paper's results.

5 RESULTS

In the following, we will report the results of each step in our co-creative workshop with a special focus on our research questions.

5.1 Persona Definition & User Journey Mapping

Initially, both groups, employment consultants and case managers, had to define personas. The resulting female personas were called *Melissa* and *Mary* (see Figure 3). Participants outlined a daily schedule that was similar for both personas:

- **Start of work:** 8 a.m. in the office
- **Checking schedule:** Getting an overview of the clients for the day, especially their problems and topics/plans for the meetings with the clients.
- **Appointment with client:** Talk about the client’s actual situation, plan the next steps and write an action plan by describing the goals for the next meeting and the steps to achieve them. These meetings are with 6 to 8 clients a day.
- **E-Mails & phone calls:** After each meeting, an E-Mail or letter with the updated action plan and the next steps is sent to the client. Also, contact with other relevant stakeholders (e.g., employers, colleagues) via phone/E-Mail/personal meetings is done.
- **Preparation:** Create the day plan for the next day, checking mails etc.
- **End of work:** 4 p.m.

The AI-based OTT system is used during and after meetings with clients. Since it is integrated into the EMPIS interface, the tool where notes can be made and the action plan is written in, its information is easily accessible by the personas. Besides EMPIS with OTT, the personas use digital tools daily (e.g., Microsoft Teams to communicate with colleagues, the analytical tool TARU, and E-Mail services) (see listed tools in Figure 4).

5.2 Synthesize & Validate

After defining a typical day and the workflow of the personas during the day, we focused on the problems, especially with OTT in this workflow. The results answered our RQ 1: *How and in which specific situations/contexts do challenges and problems arise while using the OTT system?* The participants stated the following sub-themes:

- **Client-related:** Trustful interaction with the client, motivation of clients, bad experiences of clients with previous unemployment, language barriers
- **OTT system:** Difficult to understand evaluation criteria, missing empathy & trust, the system is too rational, no space for personal feedback/notes, and missing edit feature (see Figure 5)
- **Interface:** Too much scrolling, CV location is not easy to reach, jumping between different windows
- **Data:** Missing information from employer register

In a discussion round, we talked about the pain points regarding OTT in more detail (see Figure 8). The participants stated issues related to general topics regarding AI (e.g., missing trust & interpretability) as well as specific problems of the OTT system (e.g., no option to add information about the client). In working out the problems with OTT, the participants repeatedly focused on their clients. Participants highlighted already during the persona creation and in the user journey mapping session that it is important to them to maintain a good relationship with their clients. They stated that this is essential to provide the best possible service.

Therefore, they hesitated to use OTT’s predictions because they felt the results were not easily understandable. They worried that relying on OTT’s predictions may harm their client relationship and erode their trust.

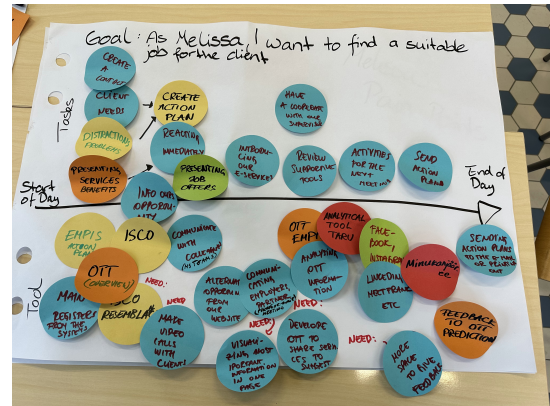
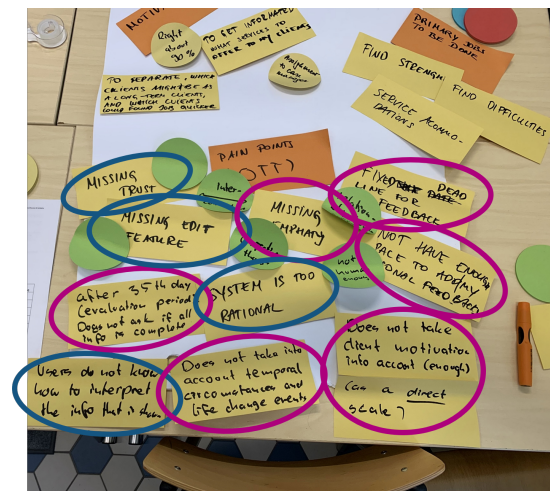


Figure 4: User Journey Map for the persona Melissa. The necessary tasks and tools are displayed to reach the goal of supporting clients in finding a job. Tasks and tools are ranged due to their occurrence during a typical working day for Melissa.



- General problems with AI-applications
- Specific problems with OTT

Figure 5: Problems that participants mentioned for the persona Melissa. The stated pain points can be separated into general problems with AI found in the literature and specific problems regarding the OTT system (i.e., functions and interface of the software).

In summary, participants highlighted several problems while using OTT, including client-related issues, challenges with the OTT system itself, interface issues, and data-related challenges. Participants emphasized the importance of maintaining strong client

relationships and expressed concerns about using OTT predictions due to perceived difficulties in understanding and potential harm to client trust.

To answer *RQ 2: How can we improve user acceptance for the OTT system?*, we focused on the topics *explainability, trust, fairness, control, and ethics*. The key takeaways we found in discussion with the participants were:

- **Explainability:** The presented features from OTT are not very clear, highlighted in statements like “Well, I would not use the predictions of OTT in a meeting with a client. When I cannot explain the reasons for OTT’s predictions, this won’t make a good impression on my client.” (P in group 1) (see a detailed analysis of this in Table 1 and the following text)
- **Trust:** General trust in the system was given. The main reason for this was that the prediction accuracy was high. In addition, participants stated that they trust the developers of OTT: “I think OTT works fine but is limited. I trust the developers of the OTT system.” (P in group 1)
- **Fairness:** Participants had no strong opinion on this topic. They stated that maybe OTT focuses too much on negative features, which decreases the possibility of getting employed again and therefore impacts the prediction, but this was an assumption of the participants due to the lack of transparency of OTT. One participant summarized it as: “I don’t think we have a problem with fairness. OTT is one tool we use, and we decide how to use it. The bigger problem is the interpretation and practical usage of OTT’s output.” (P in group 2)
- **Control:** Participants stated having a weak feeling of control regarding the *outcome of OTT* because they did not know how OTT came to a decision. In contrast, they had a strong feeling of control regarding the *impact of OTT* because they decided which information was given to the client. In addition, the OTT decision is not final, meaning that OTT-based decisions are just suggestions for the unemployment consultants.
- **Ethics:** Since digitalization is a common topic in Estonia, participants stated they are used to AI-based tools. Therefore, they had no general ethical problems using AI-based tools, and OTT in particular. Participants stated that unbiased software is essential to them. Here, they trust the developers of the OTT system that the tool is a fair one. Nevertheless, participants stated to be interested in education courses on this topic to get more insights into the topic of ethics and responsible AI since they have not participated in such courses before.

Since our focus was designing user-centered XAI interfaces, we investigated the feedback regarding explanations in more detail. For this, we structured the feedback from the participants based on step 1 (i.e., question elicitation) of the question-driven XAI design process from Liao et al. [30] by clustering questions to identify types of explanations participants wanted into categories (i.e., step 2 - question analysis). For this, we used three XAI categories presented in Liao et al. [30]. Based on these XAI categories and the questions uttered by participants, we defined user requirements that served as design goals for the prototyping session (see Table 1).

To summarize the insights regarding RQ2, we found that participants have an overall impression that OTT is working well - based on two aspects: (1) their impression that the decisions of OTT correspond most of the time with their own and (2) their trust in the developers. Nevertheless, participants stated the issue that they do not understand the inner workings of OTT and what the selected features of OTT’s random forest classifier mean in detail. As a result, unemployment consultants rarely use the information from OTT in communication with their clients.



Figure 6: The workshop participants address one pain point in the prototype session to develop a first idea for a solution.



Figure 7: Creation of the XAI interface as a paper prototype. Above the paper for the prototypes are notes from the Synthesize & Validate session that the participants want to address.

5.3 Prototype

In the final phase, which took place in the afternoon at the co-creative workshop, we addressed *RQ 3: How could prototypes look like that tackle the identified issues?* Participants worked again in

Table 1: User questions were clustered in XAI categories. We derived user requirements regarding an XAI interface from this.

XAI Category	Questions	Requirement
Why?	“Why did I get this prediction?” “Why is my client getting a low probability to find a job again?” “Why is my client getting a high probability to find a job again?” “Why are the top ten features relevant?” “Why does the feature affect finding a job?”	Explain reasons for the prediction
How to be that?	“How could my client improve the prediction?” “In which area does the client have the greatest potential for improvement?”	Provide suggestions
How (global)?	“How does OTT work?” “How should I interpret the output?” “How can I understand it better?”	Offer inside courses for employees

two groups to select the most important one of the collected pain points regarding the OTT system. Both groups took the pain point “Users do not know how to interpret the information shown by OTT”. For this pain point, two paper prototypes (see Figures 9 and 10) were developed. The paper prototypes are based on the current OTT interface (see Figure 1) and supplemented with a pop-up menu. The paper prototypes addressed the identified user requirements from the Synthesize & Validate session (see Table 1).

Interestingly, both user groups independently rely on textual explanations to *explain the reasons for the prediction*. When asked why they did not want graphics, all participants said they already got many graphics from EMPIS and could work faster with text. One participant summarized this in saying: “You know, I’m shown so many graphics already. I’m glad when I can just read some text, and the information is in there” (P in group 1). In addition, participants stated that text could easily be transferred to other documents (e.g., action plan). Participants provide the following examples of textual explanations they would wish for:

- **OTT feature ‘Working Experience’:**
 - Client is 22 years old. Statistically, having eight months of working experience indicated the tendency to find a new job within six months. (**Addressed requirement: Explain reasons for the prediction**)
 - The Client is 19 years old and in the red zone because of age and no working experience. It is harder to find a job without work experience. (**Addressed requirement: Explain reasons for the prediction**)
- **OTT feature ‘Place of Residence’:**
 - Client has no driver’s license and lives in a rural area. Statistically, having no driver’s license lowers the chances of finding a job in rural areas. (**Addressed requirement: Explain reasons for the prediction**)
 - Client has no driver’s license and lives in a rural area. Since fewer workplaces are within reach of the client, it is harder to find a job. (**Addressed requirement: Explain reasons for the prediction**)

To avoid losing the overview, participants stated that the explanations of the OTT features should appear in a pop-up menu. The option of a pop-up menu to show and hide explanations was

very important to the participants, who only wanted an explanation when needed. In addition to the explanations, the participants would like OTT to provide *suitable suggestions* for the clients to add to the action plan. Finally, it was essential to the participants that the selected explanations and suggestions could be added directly to the action plan for a client with one click instead of Copy & Paste actions to save time.

To summarize the findings from RQ3, participants wanted additional textual details/explanations, but they wanted to control when to view this rather than it being displayed all the time. They seek for OTT suggestions to be actionable and directly transferred to an action plan - the next step in unemployment coaching. Finally, they want to be able to give feedback to the OTT system with their knowledge of the situation to improve future suggestions.



Figure 8: Summarizing the pain points of the OTT system. The deep dive discussion was the basis for the next step of the co-creative approach: to develop a first paper prototype to address one of the problems stated.

5.4 Focus Group

In the focus group discussion at the end of the workshop, participants re-iterated the value of such workshops to increase the understanding and acceptability of AI-based software like OTT: “I know now better how OTT supports me” (P5). “I think I will use it

more” (P5). “For me in personal, I was more informed about why I need to use OTT in my workday and what is the purpose of it” (P3). Participants pointed out that initially, they felt a barrier to providing recommendations for an AI system. “There was a moment where it stuck: then it was explained more, and we were encouraged to go further” (P2). They highlighted that, despite their familiarity with the OTT output, it can be difficult to precisely describe what to expect from an XAI interface in this context. “How to develop OTT concretely: I haven’t thought about this; it was a refreshing experience” (P5). “For me, it was hard to give an example. I know what I want to see there but to give an example of what and how OTT should explain it to us was hard” (P3). “We use it every day as a work tool, but we do not analyze it, e.g., what is necessary to improve” (P1). Finally, they also mentioned the need to provide feedback on the results that OTT provides: “OTT can be improved by our information” (P4).

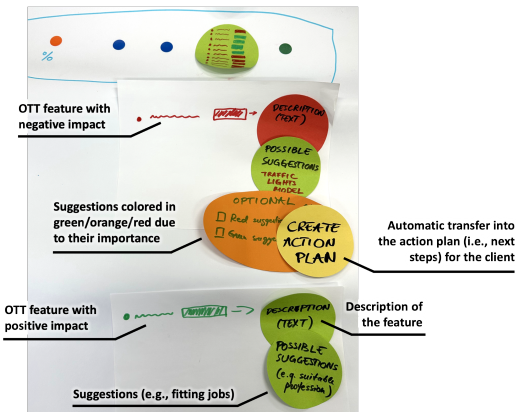


Figure 9: Paper prototype of the unemployment consultants. The dots at the top of the image illustrate the current OTT interface (see Figure 1 for the original interface). A detailed description can be displayed when clicking on the feature. Each feature that has a negative or positive impact on the AI prediction is clickable. In addition, suggestions to improve negative features or to further support positive ones should be highlighted.

6 DISCUSSION OF RESULTS

In the following section, we summarize and discuss findings regarding user requirements that may be valuable for designing XAI systems for similar use cases. We provide references to the related research questions by denoting them with [RQ1/2/3]. RQ1 addressed *challenges & problems*, RQ2 dealt with the *improvement of user acceptance*, and RQ3 focussed on the *XAI paper prototype*.

Seamless Workflow Integration is Key. A critical insight gained from the workshop is the necessity for XAI to seamlessly align with existing workflows in the domain of use [RQ1], in our case, social service provision. Discussions with unemployment consultants highlighted that the successful implementation of XAI hinges on a deep understanding of their institution’s intricate workflows and processes. Furthermore, they stated that providing feedback to

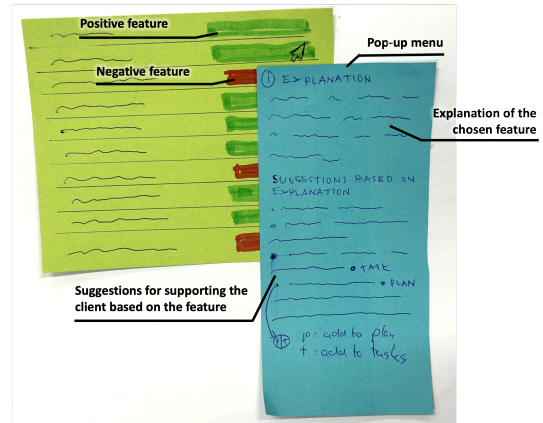


Figure 10: Paper prototype of the case managers. By clicking on a positive or negative feature that influenced OTT’s prediction, a pop-up menu opens that explains the chosen feature and respective suggestions to support the client.

the OTT system, especially on its predictions, and getting detailed explanations are crucial [RQ2]. This influenced topics common in HCAI research like *explainability* and *fairness* and the relationship between participants and their clients. One participant summarized this in the statement: “Well, I would not use the predictions of OTT in a meeting with a client. When I cannot explain the reasons for OTT’s predictions, this won’t make a good impression on my client.” (P in group 1) The currently missing explainability in OTT underscores the need for a comprehensive understanding of (1) the AI system by the end-users and (2) the context of use for XAI designers and developers. Our results underscore the necessity for including user context [31] and adopting a human-in-the-loop approach [39] during the development of AI systems. To enable consultants to use AI tools to their full potential, they should have the possibility to report back to the system when it did not do well or when essential information is missing (e.g., incomplete explanations).

AI is Appreciated as a Tool. Participants revealed a favorable disposition toward integrating AI technologies in their daily workflow. The digital landscape of Estonia and the e-government modernization can be seen as drivers for this acceptance of new technologies [24]. Participants stated in the focus group discussion that the workshop made them appreciate OTT as a tool and what they need to use it more effectively, for example, to include it in their client meetings. In addition, the workshop was perceived as a reflection of their daily routine. They stated that engaging with the XAI tool’s details helped them become more aware of their work’s central goals (e.g., helping clients with concrete actions). As such, our qualitative data indicates that the workshop increased the adaptability of the XAI system and further narrowed down the core user journey that designers need to focus on, namely delivering concrete actions.

Furthermore, the discussions unveiled an optimistic stance regarding issues of fairness related to OTT. One of the participants said: “I don’t think we have a problem with fairness. OTT is one tool we use, and we decide how to use it. The bigger problem is the interpretation and practical usage of OTT’s output.” (P in group 2).

Participants also highlighted the reason for this by saying: *“I think OTT works fine but is limited. I trust the developers of the OTT system.”* (P in group 1). The findings show that the participants value transparency and control and wish to understand the system’s reasoning processes. Based on the participants’ feedback that they have the power to accept or reject system decisions, concerns about fairness diminish [RQ2]. This indicates a preference for relying on their judgments of fairness rather than depending on the inherent fairness of the system. Controlling the narrative, in the form of providing feedback to the system and thus, improving it, was also one of the concluding remarks at the end of the workshop day: *“OTT can be improved by our information”* (P4). Our results confirm the framework provided by Eiband et al. [13] that differentiates between knowledge types when successfully interacting with intelligent systems: active (e.g., feed-backing knowledge to improve OTT) and passive (e.g., gaining knowledge through detailed explanations). These combined results strengthen our finding that trust remains high in such a system as long as various control features are implemented [RQ2 & RQ3].

Users Want to Understand the AI. Statements like *“Well, I would not use the predictions of OTT in a meeting with a client. When I cannot explain the reasons for OTT’s predictions, this won’t make a good impression on my client.”* (P in group 1) hint at a familiar problem users had while interacting with the AI system: OTT’s decisions are perceived as intransparent and incomprehensible, making it challenging to integrate the AI’s output into the client’s action plan, where concrete next steps are planned to increase the client’s chances on the job market. The textual explanations in the paper prototype addressed precisely this need [RQ3]. All participants agreed that efficient work with OTT was essential for their daily work. They uniformly expressed the need for additional support through explanations and suggestions to productively use the system’s output. When investigating the requirements for XAI design, we found that participants focus on three XAI categories (see Table 1), supporting the statement of Miller [37], that users do not want all possible explanations that an AI system could provide. Participants raised questions for local (e.g., *“Why did I get this prediction?”*) and global explanations (e.g., *“How does OTT work?”*) (see Table 1). Local explanations addressed the users’ needs for interpretability on a case-by-case basis, fostering trust and comprehension of individual predictions. Simultaneously, global explanations offer users a more holistic grasp of the AI system, promoting a comprehensive understanding of its overall behavior. Given these insights, we recommend differentiating between local and global explanations, whereby the latter is provided before using the system as part of training material, and the first is provided in situ - when consultants are conversing with employment seekers.

In addition, we found that users wished for explanations in combination with a suggestion (e.g., the client has no driver’s license, which lowers the chances of finding a job: the suggestion is to get a driver’s license). This extends findings from research in the domain of human-robot interaction, highlighting the need for combining textual explanations with concrete suggestions [20].

Text as Modality of Choice for Explanations. One prominent observation from the co-creative workshop was the clear preference for textual explanations over graphical ones [RQ3]. According to

participant feedback, they seek concise language, expecting to help reduce cognitive load and optimize work processes. At the same time, they sometimes found graphical information, already prominently used in the existing software solution (EMPIS), hard to interpret. Participants further stated that textual explanations are easier to integrate into the client’s action plan. For example, one participant said, *“You know, I’m shown so many graphics already. I’m glad when I can just read some text, and the information is in there.”* (P in group 1) Another participant stated that text is more accessible to clients and can be included in a conversation. This result underscores the importance of tailoring explanation formats to the users’ needs and preferences in the specific context of use. Our results indicate that the participants try to understand the prediction of the OTT system while, at the same time, combining this understanding with their long-term knowledge about the unemployment system to advise their clients in the best possible way. As stated by participants, this leads to an additional cognitive load. We assume that the need for textual explanation derives from the need to reduce this additional cognitive load. A reduced cognitive load would enable them to focus on the client conversation, a priority they repeatedly stated, instead of splitting their focus across multiple cognitive tasks. As OTT has only been in use since 2020, this cognitive load could be reduced over time and with the routine of using this system [8]. Therefore, more detailed explanations might be necessary in the training phase. They should be replaced with shorter texts or graphics when users are more familiar with using OTT and have developed an increased trust in the system.

We note that the consensus regarding the explanation format was found during the conversation-heavy phases of the workshop, such as the Synthesize & Validate and the focus group session. However, as stated in the previous paragraph, *“AI is appreciated as a Tool”*, participants also pointed out the need to actively give feedback/provide their knowledge to the OTT system. Reading detailed explanations and offering feedback to the system during client consultations simultaneously impose an additional cognitive load on the consultants. In an upcoming quantitative study, we intend to explore interaction patterns that balance concurrent tasks, such as reading explanations and providing feedback, and manage cognitive load effectively.

7 METHODOLOGICAL CONSIDERATIONS & LIMITATIONS

Co-Creative Approaches are Suitable for XAI Design. We used the user-centered design approach [40] and the question-driven design process from Liao et al. [30] to develop XAI interfaces for social service provision. Both processes are characterized by the fact that they enable a structured approach to the creation of XAI designs. In addition, participants highlighted that they help to reflect on current work processes and how they could be improved using XAI. The combined approach allowed us to identify unemployment consultants’ problems and questions for the OTT system. The participants highlighted that the approach helped them to reflect on current work processes and how they could be improved using XAI. By providing us with concrete requirements and example explanations, the participants provided valuable insights for future research.

For example, we can now assess the feasibility of integrating the desired explanations into the OTT system through its developers.

From a methodological perspective, it is desirable to reflect on alternative approaches that we had considered due to the difficulty in recruiting and planning a workshop with government employees. A less resource-heavy method would be to use heuristics, as derived from the XAI categories, using the question-driven approach. Although such an approach could have revealed a sub-set of the findings we presented in this paper, we strongly argue for a collaborative method for evaluating XAI systems in public sector settings. Our co-creative approach allowed us to adequately consider the specific circumstances of the unemployment consultants (e.g., strong identification with the clients and focus on explanations that reduce cognitive load) and specific cultural aspects (e.g., openness towards AI systems). In addition, including the stakeholders in the XAI design process increases the acceptability of the whole system, as participants stated in the focus group session in statements like “I know now better how OTT supports me” (P5).

Deep Dives are Worth the Time. We had apparent time constraints since we had planned a one-day workshop. Therefore, we focused on one specific pain point: the lack of interpretability of OTT’s output. However, participants expressed their satisfaction regarding this approach in the focus group discussion at the end of our workshop, as they appreciated the focus and depth we invested in this specific pain point, e.g., “For me in personal, I was more informed about why I need to use OTT in my workday and what is the purpose of it” (P3). Notably, this decision was also crucial since the context of use was deemed highly complex by all parties involved.

Communication is Challenging. With the help of the personas and the user journey mappings, the participants could draw a precise and comprehensible impression of their daily work. However, it became apparent that they often lacked the terminology to address the problem when formulating pain points. Here, we decided to provide concrete help with keywords from XAI literature (i.e., explainability, trust, fairness, control, & ethics) to stimulate a discussion. This increased engagement, as participants could anchor their previous thoughts on specific terms. In addition, the experimenter observed that participants found it challenging to transfer knowledge from other UI designs (e.g., smartphone interfaces or desktop software) to their particular use case. Participants felt that their OTT tool was too complex to be represented, as other everyday software systems that they use, which heavily rely on graphical representations.

More Helpers Help More. We had five participants and five researchers in our co-creation design workshop. Even though this 1:1 ratio may seem excessive, it proved necessary. Public service processes are complex and involve dealing with sensitive data. Participants confirmed that this is due to the emotional aspect that an employment-seeking status forebrings. They stated in the Synthesize & Validate session of the workshop that some clients feel frustrated or ashamed when they cannot find a job. In addition, personal data such as grades, school leaving certificates, and employer references may be discussed in the meeting with the unemployment consultant. Although we knew this during the planning process, this assumption solidified on the workshop day. It was crucial for

one researcher to be embedded in each workshop group to take away some of the cognitive load of understanding the research method on top of explaining public service processes and data handling. Considering recruitment is challenging, and workshops are difficult to repeat due to the availability of participants, we recommend an increased amount of researchers for public sector settings.

8 RECOMMENDATIONS FOR HCI PRACTITIONERS IN THE PUBLIC SECTOR

Pay Attention to Client Needs, not only User Needs. Our study found that participants were very engaged with their role as unemployment consultants and strongly identified with their goal of supporting their clients in their job search. We often observed that participants prioritized the needs of their clients over personal needs such as usability or user experience [RQ1]. As such, we strongly recommend focusing on public servants’ goals and understanding the needs of their clients while analyzing the context of use. This could be achieved by including multiple stakeholders in the design process, e.g., clients for social service provision. This is especially desirable from an ethical point of view since clients are substantially affected by the software tools’ recommendations. Work like Eubanks [14] highlights fairness issues for marginalized groups (e.g., lower income). It is essential to sensitize unemployment consultants to such AI biases. Since our participants stated that they would like to have in-house courses for the OTT system, this could be a way to raise awareness of this topic.

Include Cultural and Ethical Aspects. Estonia is known as a leader in developing e-government solutions. However, it might not necessarily be representative of international social welfare systems. For example, Estonia has a small population with less than average - compared to Europe - unemployment rate (5.6% compared to 6.2%).⁵ As such, scalability and having to make “quick” decisions, where the human cannot be involved - due to lack of resources - is not an imminent issue. Hence, we note that it is essential to consider such structural variables of specific use cases and countries (e.g., the structure of the welfare system and the inclusion of AI in social services). In particular, attributes regarding AI and its use in the public sector are driven by society [26]. One reason for this is the welfare history of each country. Kaun et al. [26] highlights the importance of combining social norms and structural variables (e.g., welfare regimes) and individual factors (e.g., impressions of civil servants or citizens). Although we did not focus on structural variables, we want to acknowledge and point out that they influence the results. We even noticed this on a small scale (within Europe) between the different researchers that were part of this study and originating from Germany and Estonia. We propose to evaluate AI, including structural variables typical for the country where the AI system is used.

Get in Touch with the Former System Developers. In discussions with the participants about the AI-based OTT system, it became clear that many design decisions of OTT needed to be made more

⁵Estonia’s unemployment rate in 2022 is 5.6% (see: <https://www.stat.ee/en/find-statistics/statistics-theme/work-life/labour-market/unemployment-rate>) compared to 6.2% in the EU (see: https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Unemployment_statistics_and_beyond)

explicit to them. While participants trusted OTT's developers, it is independent of understanding how OTT works and the implementation details. For studies in the public sector where civil servants are not automatically connected to the developers of a system, as software is mostly not built in-house but contracted, it would be helpful for the researcher to engage with the developers in advance about the background of design decisions (e.g., technical limitations, special requirements for data protection). This may provide valuable information both for researchers and participants during the co-creative design workshop and thereby develop even more concrete ideas for the design of AI in the public sector that consider these constraints.

9 CONCLUSION

AI-based decisions should be understandable for end-users, especially when AI is used in the public sector, where AI decisions may significantly impact individual lives. However, designing user interfaces that explain AI decisions in this highly sensitive context is challenging. By conducting a co-creative design workshop with unemployment consultants in Estonia, who use AI for social service provision regularly, we made a first endeavor to bring design philosophies from user-centered design and human-centered AI to the public realm. This paper reported lessons we learned and observations we made while mapping highly intricate user journeys, understanding user requirements for XAI systems, and supervising unemployment consultants while prototyping XAI design solutions. Our findings show that it is possible to synthesize clear UI preferences and develop concrete design solutions despite working with non-experts in AI. Further, we emphasize a clear desire for interpretable and explainable AI by AI end-users who prioritize helping their clients over simple quality-of-work-life UI features. We want to encourage HCI practitioners to apply co-creative design methodologies to intricate AI-driven use cases and promote the idea of human-centered AI, especially in ethically critical scenarios and use contexts.

ACKNOWLEDGMENTS

The authors gratefully acknowledge partial funding of this work by the German VolkswagenStiftung under project AI-FORA (grant number 98 563).

REFERENCES

- [1] Amina Adadi and Mohammed Berrada. 2018. Peeking inside the black-box: a survey on explainable artificial intelligence (XAI). *IEEE access* 6 (2018), 52138–52160. <https://doi.org/10.1109/ACCESS.2018.2870052>
- [2] Gary Anthes. 2015. Estonia: a model for e-government. *Commun. ACM* 58, 6 (2015), 18–20. <https://doi.org/10.1145/2754951>
- [3] Alejandro Barredo Arrieta, Natalia Diaz-Rodríguez, Javier Del Ser, Adrien Bénézet, Siham Tabik, Alberto Barbado, Salvador García, Sergio Gil-López, Daniel Molina, Richard Benjamins, et al. 2020. Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information fusion* 58 (2020), 82–115. <https://doi.org/10.1016/j.inffus.2019.12.012>
- [4] Sebastian Bach, Alexander Binder, Grégoire Montavon, Frederick Klauschen, Klaus-Robert Müller, and Wojciech Samek. 2015. On pixel-wise explanations for non-linear classifier decisions by layer-wise relevance propagation. *PLoS one* 10, 7 (2015), e0130140. <https://doi.org/10.1371/journal.pone.0130140>
- [5] Sally-Anne Barnes, Sally Wright, Pat Irving, and Isabelle Deganis. 2015. Identification of latest trends and current developments in methods to profile jobseekers in European Public Employment Services. (2015).
- [6] Manuela Battaglini and Steen Rasmussen. 2019. Transparency, automated decision-making processes and personal profiling. *Journal of Data Protection & Privacy* 2, 4 (2019), 331–349.
- [7] Victoria Clarke, Virginia Braun, and Nikki Hayfield. 2015. *Thematic analysis*. Vol. 3. 222–248.
- [8] Michelle Patrick Cook. 2006. Visual representations in science education: The influence of prior knowledge and cognitive load theory on instructional design principles. *Science education* 90, 6 (2006), 1073–1091. <https://doi.org/10.1002/sce.20164>
- [9] Hans de Bruijn, Martijn Warnier, and Marijn Janssen. 2022. The perils and pitfalls of explainable AI: Strategies for explaining algorithmic decision-making. *Gov. Inf. Q.* 39, 2 (2022), 101666. <https://doi.org/10.1016/j.giq.2021.101666>
- [10] Sam Desiere, Kristine Langenbucher, and Ludo Struyven. 2019. Statistical profiling in public employment services: An international comparison. (2019). <https://doi.org/10.1787/1815199X>
- [11] Sam Desiere and Ludo Struyven. 2021. Using artificial intelligence to classify jobseekers: the accuracy-equity trade-off. *Journal of Social Policy* 50, 2 (2021), 367–385. <https://doi.org/10.1017/S0047279420000203>
- [12] Finale Doshi-Velez and Been Kim. 2017. Towards a rigorous science of interpretable machine learning. *arXiv preprint arXiv:1702.08608* (2017). <https://arxiv.org/abs/1702.08608>
- [13] Malin Eiband, Daniel Buschek, and Heinrich Hussmann. 2021. How to support users in understanding intelligent systems? Structuring the discussion. In *26th International Conference on Intelligent User Interfaces*. 120–132. <https://doi.org/10.1145/3397481.3450694>
- [14] Virginia Eubanks. 2018. *Automating inequality: How high-tech tools profile, police, and punish the poor*. St. Martin's Press.
- [15] Adalbert Evers. 2005. Mixed welfare systems and hybrid organizations: Changes in the governance and provision of social services. *Intl Journal of Public Administration* 28, 9-10 (2005), 737–748. <https://doi.org/10.1081/PAD-200067318>
- [16] Julie Gerlings, Millie Søndergaard Jensen, and Arisa Shollo. 2021. Explainable AI, but explainable to whom? *arXiv preprint arXiv:2106.05568* (2021). <https://arxiv.org/abs/2106.05568>
- [17] Leilani H. Gilpin, Cecilia Testart, Nathaniel Fruchter, and Julius Adebayo. 2019. Explaining Explanations to Society. *arXiv preprint arXiv:1901.06560* (2019). <http://arxiv.org/abs/1901.06560>
- [18] Niklas Gudowsky and Mahshid Sotoudeh. 2017. Into blue skies—a transdisciplinary foresight and co-creation method for adding robustness to visioneering. *NanoEthics* 11 (2017), 93–106. <https://doi.org/10.1007/s11569-017-0284-7>
- [19] David Gunning and David Aha. 2019. DARPA's explainable artificial intelligence (XAI) program. *AI Magazine* 40, 2 (2019), 44–58. <https://doi.org/10.1609/aimag.v40i2.2850>
- [20] Kasper Hald, Katharina Weitz, Matthias Rehm, and Elisabeth André. 2021. “An Error Occurred” - Trust Repair With Virtual Robot Using Levels of Mistake Explanation. In *Proceedings of the 9th International Conference on Human-Agent Interaction*, Kohei Ogawa, Tomoko Yonezawa, Gale M. Lucas, Hirotaka Osawa, Wafa Johal, and Masahiro Shiomi (Eds.). ACM, 218–226. <https://doi.org/10.1145/3472307.3484170>
- [21] Paul Henman. 2020. Improving public services using artificial intelligence: possibilities, pitfalls, governance. *Asia Pacific Journal of Public Administration* 42, 4 (2020), 209–221. <https://doi.org/10.1080/23276665.2020.1816188>
- [22] Robert R Hoffman, Gary Klein, and Shane T Mueller. 2018. Explaining explanation for “explainable AI”. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Vol. 62. SAGE Publications Sage CA: Los Angeles, CA, 197–201. <https://doi.org/10.1177/1541931218621047>
- [23] Robert R Hoffman, Shane T Mueller, Gary Klein, and Jordan Litman. 2018. Metrics for Explainable AI: Challenges and Prospects. (2018). [arXiv:1812.04608](http://arxiv.org/abs/1812.04608) <http://arxiv.org/abs/1812.04608>
- [24] Ahto Kalja, Janari Pöld, Tarmo Robal, and Uuno Vallner. 2011. Modernization of the e-government in Estonia. In *2011 Proceedings of PICMET'11: Technology Management in the Energy Smart World (PICMET)*. IEEE, 1–7.
- [25] Areti Karamanou, Evangelos Kalampokis, and Konstantinos A. Tarabanis. 2022. Linked Open Government Data to Predict and Explain House Prices: The Case of Scottish Statistics Portal. *Big Data Res.* 30 (2022), 100355. <https://doi.org/10.1016/j.bdr.2022.100355>
- [26] Anne Kaun, Anders Olof Larsson, and Anu Masso. 2023. Automating public administration: citizens' attitudes towards automated decision-making across Estonia, Sweden, and Germany. *Information, Communication & Society* 27, 2 (2023), 314–332. <https://doi.org/10.1080/1369118X.2023.2205493>
- [27] Frank C Keil. 2006. Explanation and understanding. *Annual Review of Psychology* 57 (2006), 227–254. <https://doi.org/10.1146/annurev.psych.57.102904.190100>
- [28] Peeter Leets. 2022. *Augmenting public sector data-driven decision support systems with expert knowledge: case of OTT*. Master's thesis. University of Tartu, Faculty of Social Sciences, Johan Skytte Institute of Political Studies.
- [29] Veiko Lember, Rainer Kattel, and Piret Tõnurist. 2018. Technological capacity in the public sector: The case of Estonia. *International Review of Administrative Sciences* 84, 2 (2018), 214–230. <https://doi.org/10.1177/0020852317735164>
- [30] Q Vera Liao, Milena Pribić, Jaesik Han, Sarah Miller, and Daby Sow. 2021. Question-driven design process for explainable AI user experiences. *arXiv preprint arXiv:2104.03483* (2021).

- [31] Q Vera Liao, Yunfeng Zhang, Ronny Luss, Finale Doshi-Velez, and Amit Dhurandhar. 2022. Connecting algorithmic research and usage contexts: a perspective of contextualized evaluation for explainable AI. In *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, Vol. 10. 147–159. <https://doi.org/10.1609/hcomp.v10i1.21995>
- [32] Scott M Lundberg, Gabriel Erion, Hugh Chen, Alex DeGrave, Jordan M Prutkin, Bala Nair, Ronit Katz, Jonathan Himmelfarb, Nisha Bansal, and Su-In Lee. 2020. From local explanations to global understanding with explainable AI for trees. *Nature machine intelligence* 2, 1 (2020), 56–67. <https://doi.org/10.1038/s42256-019-0138-9>
- [33] Scott M. Lundberg and Su-In Lee. 2017. A Unified Approach to Interpreting Model Predictions. In *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, December 4-9, 2017, Long Beach, CA, USA*, Isabelle Guyon, Ulrike von Luxburg, Samy Bengio, Hanna M. Wallach, Rob Fergus, S. V. N. Vishwanathan, and Roman Garnett (Eds.), 4765–4774. <https://proceedings.neurips.cc/paper/2017/hash/8a20a8621978632d76c43dfd28b67767-Abstract.html>
- [34] Nicholas Maltbie, Nan Niu, Matthew Van Doren, and Reese Johnson. 2021. XAI tools in the public sector: a case study on predicting combined sewer overflows. In *ESEC/FSE '21: 29th ACM Joint European Software Engineering Conference and Symposium on the Foundations of Software Engineering, Athens, Greece, August 23-28, 2021*, Diomidis Spinellis, Georgios Gousios, Marsha Chechik, and Massimiliano Di Penta (Eds.). ACM, 1032–1044. <https://doi.org/10.1145/3468264.3468547>
- [35] Nijat Mehdiyev, Constantin Houy, Oliver Gutermuth, Lea Mayer, and Peter Fettke. 2021. Explainable Artificial Intelligence (XAI) Supporting Public Administration Processes - On the Potential of XAI in Tax Audit Processes. In *Innovation durch Informationssysteme - WI als zukunftsweisende Wissenschaft, 16. Internationale Tagung Wirtschaftsinformatik (WI 2021), March 09-11, 2021, Universität Duisburg-Essen, Germany*. AISel. <https://aisel.aisnet.org/wi2021/SSmartCity/Track08/5>
- [36] Silvan Mertes, Tobias Huber, Katharina Weitz, Alexander Heimerl, and Elisabeth André. 2022. Ganterfactual—counterfactual explanations for medical non-experts using generative adversarial learning. *Frontiers in artificial intelligence* 5 (2022), 825565. <https://doi.org/10.3389/frai.2022.825565>
- [37] Tim Miller. 2019. Explanation in artificial intelligence: Insights from the social sciences. *Artificial Intelligence* 267 (2019), 1–38. <https://doi.org/10.1016/j.artint.2018.07.007>
- [38] Gianluca Misuraca, Colin Van Noordt, et al. 2020. AI Watch-Artificial Intelligence in public services: Overview of the use and impact of AI in public services in the EU. *JRC Research Reports JRC120399* (2020).
- [39] Robert Munro Monarch. 2021. *Human-in-the-Loop Machine Learning: Active learning and annotation for human-centered AI*. Simon and Schuster.
- [40] Donald Norman and Stephen W. Draper. 1986. *User Centered System Design: New Perspectives on Human-Computer Interaction*. Lawrence Erlbaum Associates, USA.
- [41] Nortal. 2021. OTT – tekoälyn vauhdittamaa menestystä Viron julkisella sektorilla – nortal.com. <https://nortal.com/insights/ott-an-ai-powered-success-story-in-the-public-sector/>. [Accessed 12-09-2023].
- [42] Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. 2016. Why should I trust you?: Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, August 13-17, 2016*, Balaji Krishnapuram, Mohak Shah, Alexander J. Smola, Charu C. Aggarwal, Dou Shen, and Rajeev Rastogi (Eds.). ACM, 1135–1144. <https://doi.org/10.1145/2939672.2939778>
- [43] Mark O Riedl. 2019. Human-centered artificial intelligence and machine learning. *Human Behavior and Emerging Technologies* 1, 1 (2019), 33–36. <https://doi.org/10.1002/hbe2.117>
- [44] Tjeerd AJ Schoonderwoerd, Wiard Jorritsma, Mark A Neerinx, and Karel Van Den Bosch. 2021. Human-centered XAI: Developing design patterns for explanations of clinical decision support systems. *International Journal of Human-Computer Studies* 154 (2021), 102684. <https://doi.org/10.1016/j.ijhcs.2021.102684>
- [45] Ben Shneiderman. 2000. The limits of speech recognition. *Commun. ACM* 43, 9 (2000), 63–65. <https://doi.org/10.1145/348941.348990>
- [46] Jesper Simonsen and Toni Robertson. 2012. *Routledge international handbook of participatory design*. Routledge, New York.
- [47] William H Voorberg, Viktor JJM Bekkers, and Lars G Tummers. 2015. A systematic review of co-creation and co-production: Embarking on the social innovation journey. *Public management review* 17, 9 (2015), 1333–1357. <https://doi.org/10.1080/14719037.2014.930505>
- [48] Bernd W Wirtz, Jan C Weyerer, and Carolin Geyer. 2019. Artificial intelligence and the public sector—applications and challenges. *International Journal of Public Administration* 42, 7 (2019), 596–615. <https://doi.org/10.1080/01900692.2018.1498103>
- [49] Ziheng Zhang and Francesco Zurlo. 2020. Positioning participatory engagement in participatory design. In *HCI International 2020-Late Breaking Papers: User Experience Design and Case Studies: 22nd HCI International Conference, HCII 2020, Copenhagen, Denmark, July 19–24, 2020, Proceedings 22*. Springer, 367–379. https://doi.org/10.1007/978-3-030-60114-0_25