IEEE *Access*
Multidisciplinary : Rapid Review : Open Access Journal

Personas for Artificial Intelligence (AI) An Open Source Toolbox

ANDREAS HOLZINGER[1,2], (Senior Member, IEEE), MICHAELA KARGL[1], BETTINA KIPPERER[1], PETER REGITNIG[1], MARKUS PLASS[1] AND HEIMO MÜLLER [1], (Member, IEEE)

[1]Medical University Graz, Austria (e-mail: andreas.holzinger@medunigraz.at)
[2]Alberta Machine Intelligence Institute, Canada

Corresponding author: Andreas Holzinger (e-mail: andreas.holzinger@medunigraz.at).

ABSTRACT Personas have successfully supported the development of classical user interfaces for more than two decades by mapping users' mental models to specific contexts. The rapid proliferation of Artificial Intelligence (AI) applications makes it necessary to create new approaches for future human-AI interfaces. Human-AI interfaces differ from classical human-computer interfaces in many ways, such as gaining some degree of human-like cognitive, self-executing, and self-adaptive capabilities and autonomy, and generating unexpected outputs that require non-deterministic interactions. Moreover, the most successful AI approaches are so-called "black box" systems, where the technology and the machine learning process are opaque to the user and the AI output is far not intuitive. This work shows how the personas method can be adapted to support the development of human-centered AI applications, and we demonstrate this on the example of a medical context. This work is - to our knowledge - the first to provide personas for AI using an openly available *Personas for AI toolbox*. The toolbox contains guidelines and material supporting persona development for AI as well as templates and pictures for persona visualisation. It is ready to use and freely available to the international research and development community. Additionally, an example from medical AI is provided as a best practice use case. This work is intended to help foster the development of novel human-AI interfaces that will be urgently needed in the near future.

INDEX TERMS Artificial Intelligence, Human–AI Interface, Personas

I. INTRODUCTION

Artificial Intelligence (AI) applications are becoming increasingly common in almost all domains of human life from agriculture [1] to zoology [2]. In this paper, we provide a generic example from the medical domain, and start with a provocative question: *Why should human–AI interfaces (Fig. 1) be in any way different from the classical human–computer interface as it was described in [3], [4], and [5] almost four decades ago?*

The answer can be found in (a) the advances in statistical machine learning over the last 10 years, (b) a tremendous downsizing of hardware while increasing performance, and (c) the tremendously high penetration rate of intelligent sensing in our daily lives. This has indeed brought us a new AI spring, and many successful AI applications have entered daily life, which we accept as invisible helpers day by day and no longer even perceive as great AI. Good examples are the intelligent interaction systems in our smartphones that go far beyond voice recognition or facial recognition. These human–AI interfaces are rather unobtrusive and some services, which adopt the concept of 'invisible UIs', do not even have a graphical user interface, but only provide a command-line for text input. However, such 'interfaces' should not be underestimated: they learn from our interactive behavior, store every interaction and can react adaptively, even make predictions. They are gaining some degree of human-like cognitive, self-executing, and self-adaptive capabilities and autonomy, and generate unexpected outputs
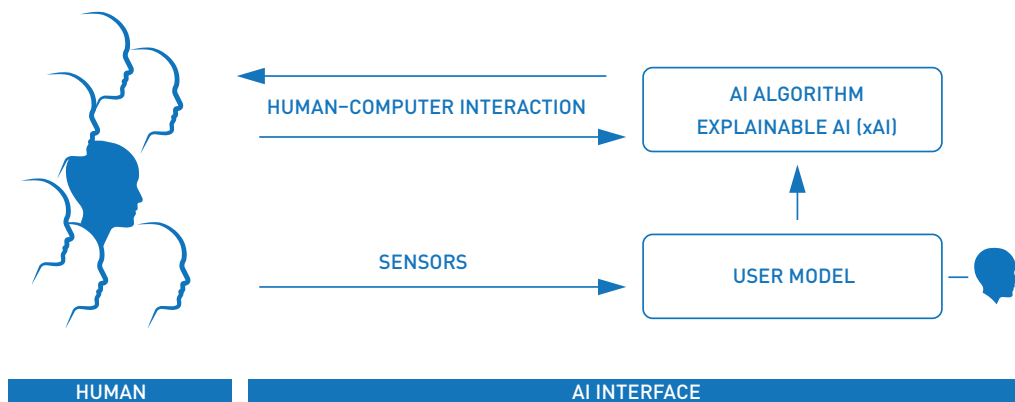
FIGURE 1. Human–AI interfaces differ in many ways from classical human–computer interfaces: they learn unobtrusively from our interaction behavior, store every interaction and can react adaptively and even make predictions about our next behaviour. They acquire some degree of human-like cognitive, self-executing, and self-adaptive capabilities and autonomy, and produce unexpected outputs that require non-deterministic interactions.

that require non-deterministic interactions. Whereas a classic human-computer interface was just an input-output tool for operating a computer to solve a specific problem, today's computers accompany people in a variety of ways around the clock. We diligently click every accept button and experience a completely different understanding of the digital computer, from the 'Von Neumann calculating machine' with switch board and punch card input/output, adding numbers [6] – to become a daily universal, even indispensable companion that completely changes the role of the user *from subscriber to producer, from reader to publisher and from consumer to developer of content* [7]. The convenience that there is "an app for everything" and "AI inside" everywhere may lead to discomfort and the eeriness that results from the invasive behavior of these applications, which is why privacy-oriented solutions must also be incorporated [8]. It is the simplicity and convenience of such human-AI interfaces with the increase in machine decisions that require a deeper understanding of the human experience with algorithms in general and the psychology of 'Human-AI Interaction (HAII)' in particular to investigate the symbolic and empowering effects of the achievements of AI-driven media on user perception and experience [9].

Human–AI Interaction (HAII) [10] differs from traditional 'Human-Computer Interaction (HCI)' [11] in several aspects, including (Fig. 1):

- AI systems can show human-like behaviour, e.g. in communications systems such as chat bots [12].
- AI systems can provide certain autonomy, e.g. humanoid robots in health support [13].
- AI systems can exhibit contextual understanding to some degree, as in advanced natural language translation systems based on neural machine translation [14].
- AI systems can demonstrate classification problem solving capabilities beyond human level, e.g. in the medical domain [15].

- AI systems enable intelligent interaction, such as voice input or facial recognition, and can adapt to the user by continuously learning from the user's behaviour, e.g. Facebook's user models [16].
- AI systems can generate output that is non-deterministic and unexpected, e.g. co-creating musical content [17].
- AI systems can not only be an assisting tool for a human operator but collaboratively work with humans as teammates [18], [19].
- AI systems can augment human intelligence, or can utilise human machine hybrid intelligence (human-in-the-loop AI systems [20], [21]).
- AI systems are mostly "black box" systems, where the machine learning technology and the learning process are opaque to the users and the AI output is not re-traceable and not or at least hard to verify [22] - see more details below.

Within the last four decades, researchers and practitioners have developed a lot of valuable methods and tools to support human-centered design (or user-centered design) of conventional computer systems [23]. However, to address the changing characteristics of AI systems as just described and to effectively support human-centered AI, many of these existing HCI methods need to be extended and adapted, and new approaches and methods need to be created to support the design and development of human-centered AI in general and human–AI interfaces in particular [10].

The *Personas* method, a well-known and successful HCI method, originally introduced by Alan Cooper in 1999 for user-centered interaction design [24], we believe is also well applicable in the context of AI systems. Personas are archetypes of users that help designers and developers focus on the needs and goals of target users throughout the product development process [25], [26]. The benefits of personas in the design and development of complex user interfaces are well known because personas closely approximate the mental

IEEE *Access*

model of various end users.

A *mental model* is an internal mental representation of the perceived real world and the relationships between its various parts and a person's intuitive perception of his or her own actions and their consequences [27].

Personas usually encompass aspects such as context and environment, tasks and workflows, skills and knowledge, personal traits, goals, values, motivations but also frustrations. To adapt the personas method to the context of HAII, we identified the following additional aspects describing the user's attitude regarding AI solutions as specifically relevant for personas for AI:

- Trust (How much trust does the user have in the decisions/output of the AI system?)
- Acceptance (Does the user accept (and follow) the decision of the AI system?)
- Assent (Is the user willing to accept/use the support by the AI system?)

Taking into account these peculiarities of AI applications and the interfaces described above, the guiding research question for our work was: How can the 'traditional' personas method be adapted for AI, specifically in a medical context, aiming at providing an adapted personas method for human-AI interface development fostering human-centered design approach [28].

To illustrate our approach, we describe the practical implementation for creating personas for AI solutions in the context of medicine. Medicine is where the demand for AI solutions will be strongest in the future and where there is already a lot of room for improvement [29]. We use digital pathology as a best practice example. Experience from this very demanding field of application is transferable to other domains, however not necessarily vice versa. In terms of human-computer interaction, medicine presents a paradox: On the one hand, the application health is considered the most important and therefore best supported field, but on the other hand, the systems currently in use are committed to classical and very basic interfaces, and only very few innovative technologies have achieved market penetration so far [30].

These facts are a perfect motivator for our work, because there is a further important aspect which we must emphasize: Future AI systems must have an ability to "explain themselves", i.e., there is a legal requirement for medical AI systems for enabling re-traceability, transparency, and explainability [31], [32]. Moreover, it will become mandatory to ensure the quality of these explanations, i.e., to ensure that a machine explanation is also understood by a human expert. Technical explainability highlights technically decision-relevant parts of machine representations and machine models, i.e., parts that contributed to model accuracy in training or to a particular prediction. However, it does not refer to a human model. For this purpose, the term causability [33] is used, which refers to a human model. The term causability was introduced in reference to the well-known

term usability [34]. Whereas explainability is about implementing transparency and traceability, causability is about measuring the quality of explanations, i.e., the measurable extent to which an explanation of a statement achieves a certain level of causal understanding for a user with effectiveness, efficiency, and satisfaction in a given context of use [35]. Causability, then, is the measurable extent to which an explanation achieves a certain level of causal understanding for a human. Another major challenge in the medical domain is that multiple modalities contribute to a single outcome, requiring multi-modal causability [36]. All these issues call for a new human–AI interface and not merely a transfer of existing traditional concepts.

When it comes to AI applications, user requirements go beyond the requirements and needs known from traditional HCI, which relate to functionality, usability, safety, physiology, psychology, and user experience [37]. In addition to these user needs from classic HCI, higher level user needs related, for example, to emotion, decision-making authority and explainability as well as ethical issues are added in HAII [10], [38].

AI systems should always provide the user with situation awareness and a human-controllable interface in order to ensure that the user is in control and the ultimate decision maker [10]. Specifically in high-stake domains, such as in financial decision-making, justice or medical diagnoses, verification of the AI application's result by the domain expert is required and therefore it is crucial that these users understand the underlying rationale and certainty of the result provided by the AI application [39]. The research field of explainable AI (xAI) aims at generating explanations about the AI application's behaviour [40]. However, since the perceived quality of an explanation is strongly dependent on the context and the user [41], it is crucial that these explanations are created with the users in mind to be understandable and useful. Thus, human-centered design and the development of AI applications are essential to achieve solutions, which are both usable and comprehensible.

This paper is structured as follows: We provide a detailed description of our approach for developing personas to support human-centered design of AI applications in section III, and discuss the results briefly in section IV. The next section (section II) gives an overview of the evolution and the usage of the personas method as well as a brief introduction to the field of explainable AI (xAI).

II. RELATED WORK

One of the first principles taught in statistics lectures is that correlation does not equal causation. Unfortunately, this is also one of the first things to be forgotten, especially in the current AI boom. This has increasingly become a problem because the current AI hype is based on the great success that statistical data-driven machine learning has achieved in the last three decades [42]. This success is based in particular on the progress of non-symbolic AI methods, such as neural networks (deep learning) [43]. These learning models are

trained to behave in a predictable way, but do not allow insight into the learned solution paths. Knowledge is thus implicitly represented and not accessible to a human expert. For example, words of a natural language are mapped onto high-dimensional vectors, making them incomprehensible to humans [44]. Especially with the most successful learning models today, it is very difficult to almost impossible for human experts to understand how a neural network has reached a result - we speak of so-called "black-box" models [45]. Therefore, experts from the fast-growing field of "explainable AI" (xAI) [46] are trying to develop methods and approaches to make such black-box models comprehensible, transparent and thus interpretable and explainable for humans. Explainable AI is not a new field. The problem of explainability is as old as AI itself [47], indeed it is the result of AI itself [48]. Of course, it would be wrong to claim that explanations are needed for everything at all times. In fact, the exact opposite is true, and that is why AI is currently so successful with its statistical learning methods: abstract algorithms find patterns in large, complex and high-dimensional data sets that no human could ever discover. That is good. However, there are certain domains and certain situations in which a comprehensible explanation is necessary. In particular, this is true in problematic situations of human decision-making. Here, an explanatory component can help to give human decision makers at least a chance to check the plausibility of a result. One example is medicine. Here, solutions are helpful that make it possible to make decisions comprehensibly transparent, understandable and explainable. Especially in security-relevant domains, the question inevitably arises: "Can we trust our results"? [49]. Here, explainable AI is not only useful and necessary, but also represents a huge opportunity for AI solutions in general, because it can reduce the alleged opacity of AI and build up the necessary trust. This is precisely what can sustainably promote acceptance among future users [50].

In 1999, Alan Cooper introduced personas, hypothetical archetypes of users, as a method to represent users throughout the design process of a software-based product [24]. Personas are helpful tools for representing users, since they enable designers and developers to empathize with these imaginary users in the same way they would empathize with real persons [51], [52]. Usually, when creating a persona, not the whole person is described, but the focus is put only on relevant aspects (such as relevant attitudes, skills...) and specific context associated with these aspects [53]. However, since personas are prone to activate and reinforce stereotypes [54], it is necessary to ensure that the diversity of people is accounted for in the way users are represented in personas. To support a comprehensive representation, Marsden and Proebster (2019) suggest to take into account the multiple identities of a person which helps minimising stereotyping and highlighting facets that are easily overlooked [55]. Personas' ability to specifically highlight certain facets of people are especially useful to help designers and developers to take on the perspective of underrepresented or easily over-

looked users. For example, personas, which were specifically designed to represent users' diversity known from gender difference research, have been successfully used to detect gender-inclusiveness issues in software [56], [57].

Over the years, the evolution of the personas method has brought about different approaches towards personas [58]: The personas described by Cooper [24] were *goal oriented personas*, distinguished from one another based on their different goals. In the early 2000's, *role-based personas*, which are defined by their roles, were introduced by Pruitt and others [59], [60]. Nielsen described scenario oriented *engaging personas*, whereby these personas do have needs based on their individual characteristics, and their goals are based on these needs and appear only in the context of a specific scenario [61]. Blythe described fully fictional personas, so-called *pastiche personas*, not established on user data but entirely grounded on fictional characters from literature or film [62].

Before the concept of personas was introduced by Cooper as a method for the development of software-based products [24], personas had been widely used also in other contexts such as general product design, marketing, communication, and service design [26]. The persona method was also further adapted, for example to better suit the software engineering process [63], or to fit to the development of products associated with social and political goals rather than market introduction [64]. There is no single way of creating and using personas, neither in literature nor in practice, but personas are developed and used in various ways [65], [66], [67], [68], [69].

With respect to the purpose personas are developed for, it can be distinguished between *user personas* (also called *design personas*) and *buyer personas* (also called *customer personas* or *marketing personas*) [60]. User personas are developed to understand the (future) users, the context of use and the user interactions in order to design and develop products, which are meaningful and easy to use, while customer personas are developed to understand the purchase motivations, habits and values of the (future) buyers, in order to successfully sell a product [53]. This paper is about user personas.

Even though the concept and idea of personas is by now popular within the product and software development, it is also not uncontroversial. Chang *et al.* [66] state that designers often create personas only implicitly for themselves instead of sharing them within their workspace in order to get everyone's input. Furthermore, Marsden and Haag [54] emphasize the importance of empirical data to create personas, which also helps to make the potential customers vivid and lively [25]. Salminen *et al.* [70] argue that even in the era of online analytical data, personas are still a useful option, as they *'give faces to data'*. Salminen *et al.* have shown that data-driven personas do have the ability to change decision makers' preconceptions of user segments [71].

Usually, with the exception of *pastiche personas*, personas are based on data and information collected about real peo-

ple. Sometimes, personas are developed together with the users [72], [73]. The classical approach for data collection for personas' development is using qualitative methods such as ethnographic interviews, open-ended survey questions, or contextual inquiries and field studies [24], [59], [53]. In the last 15 years, a collection of large amounts of quantitative data (for example from web analytics, social media, online customer data, and online surveys) has become popular and utilising these data together with techniques such as machine learning has led to so-called digital data-driven persona development mainly used in marketing and customer research [74]. So-called *hybrid personas* are created by utilising quantitative data from online analytics together with qualitative insights [70].

In literature concerning AI applications, the need for human-centered design is mainly described with respect to explainable AI (xAI), since it has been recognised that different user groups do have different needs regarding explanations [75], [76].

III. DEVELOPING PERSONAS FOR AI

In this section, we describe our approach for creating personas to support human-centered design and the development of AI applications. Based on the procedures for creation of personas proposed in literature [77], [53], we have elaborated a 5-step process for the development of user personas for AI (see Fig. 2).
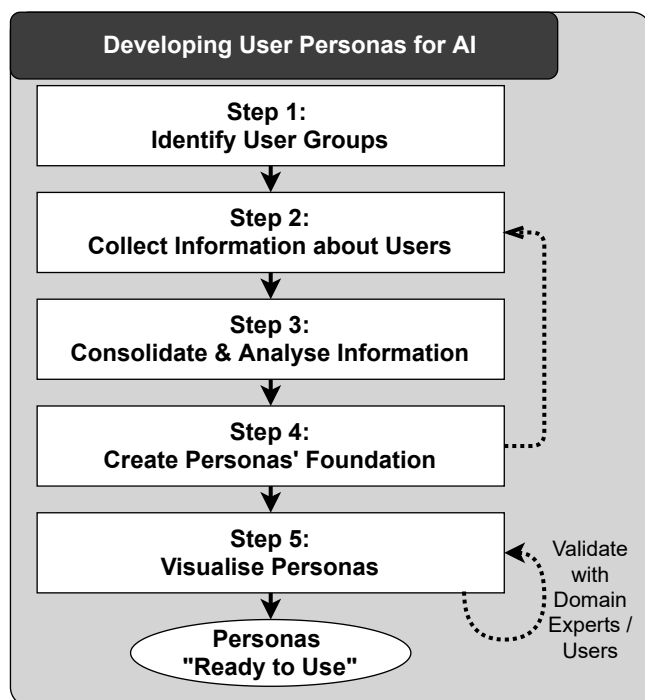

FIGURE 2. 5-step process for development of user personas for AI.

Although a large part of the process of developing personas for AI is similar to 'traditional' persona development, there are aspects and details, which are especially important when developing personas for AI. In the following paragraphs we will point out these aspects and explain the practical implementation of our approach on the example use case of creating personas for explainable AI applications for digital pathology:

In our use case, we developed 9 user personas representing user groups, which are relevant to consider during the design and development process of AI solutions for digital pathology. These personas were validated in feedback loops with respective domain experts and are currently used in practice in the ongoing design and development of the user interface of an application for AI-based analysis of digitized histology slides (so called Whole Slide Images (WSI)) in digital pathology. The process for developing the user personas was carried out by two researchers over a period of 3 months and included 7 in-depth interviews with representatives of relevant user groups, online surveys (a total of 8 different questionnaires was created for these surveys), several feedback loops with domain experts and complementary research. Approaches applied and activities undertaken in our use case are described in more detail in the following subsections.

A. STEP 1: IDENTIFICATION OF (POTENTIAL) USER GROUPS

The aim of this first step towards the development of personas is to come up with a comprehensive list of groups of people, who will (potentially) use the AI application. Usually for applications in a business domain, these user groups map with job descriptions, while for applications in a consumer domain, these user groups map with lifestyles [77]. Each of these user groups identified in this first step is the seed for a distinct persona. Hence, at this point it is important to apply a wide view rather than restricting the list to the most obvious end-users. Specifically when developing personas for AI applications in domains where causability is required, all groups of people, who will need to understand and interpret the results delivered by the AI application, shall be included as (potential) user groups.

Ideally, the identification of (potential) user groups for an AI application should be based on data to avoid misleading results. However, if no data is available, the identification of (potential) user groups can be based on the assessments of domain experts.

In our example use case, identification of (potential) user groups of explainable AI applications for digital pathology was based on domain experts' assessments: A multidisciplinary group of domain experts (including staff of pathology institutes, manufacturers of digital pathology solutions, consultants and researchers in the field of digital pathology) participated in a brainstorming and discussion session about: "Who will need to understand the rationale behind the results provided by an AI solution for digital pathology and thus will need explanations for the results provided by this AI solution"? These experts came up with a consolidated list of 10 user groups including pathology laboratory staff (pathologists, AI laboratory technician, quality

manager), researchers, people working at the manufacturer of the AI solution (software developer, quality manager, sales, customer support) and auditors for market admittance of medical devices (see Fig. 3).

B. STEP 2: COLLECTION OF INFORMATION ABOUT THE USERS

The aim of this information collection activity is threefold, whereby the first two aspects are similar to 'traditional' persona development but the third aspect is specific for developing personas for AI:

- First, to get to know (potential) users personally: find out their goals and motivations, learn about their frustrations and hopes, their skills, education and knowledge as well as their personal traits and aspirations.
- Second, to get to know the users' tasks and to discover in which context they would probably use the AI solution.
- Third, to find out the users' attitudes towards working with new technologies and innovations (- this is relevant in application cases and domains, where AI is perceived as new and innovative), as well as to find out the users' attitudes towards machine decisions: Under which conditions would they trust the decision/result of an AI application? Would they follow the decision/result of an AI application? Would they be willing to accept support by an AI application?

Typically in this kind of information collection activity, members of the respective user group are asked about their *personal* education, tasks, skills, knowledge, goals, motivations, values, and frustrations. In addition to these questions, which provide a 'direct' view on members of the respective user group, in our use case we have asked also about skills, knowledge, goals, motivations and frustrations of a *typical* member of the respective user group, which provides an 'indirect' view on that user group seen through the eyes of a peer. This approach has two advantages: First, these 'indirect' questions make it easier for people to state negative feelings and aspects such as frustrations and difficulties. In addition, these questions about a 'typical' member of the respective user group also help to level out bias in the collected information. The probability of introducing bias through information collection is especially high in (online) surveys, as those, who answer questionnaires and participate in surveys, at least according to our experience, usually tend to be the more active, extrovert and open-minded members of a group. For example, to collect data about the user group of pathologists in our use case, we have included an invitation to fill in an online questionnaire in a newsletter, which was sent to pathologists. From the questionnaire responses, which we have got from pathologists, we could observe indications for that bias effect, for example, by looking at how respondents rated statements about affinity to new technologies (see Fig. 4) and statements about attitudes towards changes (see Fig. 5) with regard to themselves and with regard to 'typical' pathologists.

Usually for developing personas, it is suggested to conduct ethnographic interviews or contextual inquiries for collecting information about the users [77], [24], [59]. However, we have realised that this approach is not always feasible in practice: Such on-site interviews might not be possible either due to external framework conditions (as for example the COVID-19 pandemic situation), or stakeholders might not be available for interviews, for example, due to time constraints. To tackle these issues, we wanted to find out, whether all user information needed for persona development could also be collected successfully by other ways of data gathering involving no interviews. To research this, we have developed and applied four alternative approaches for information collection in our use case. As depicted in Fig. 6, in two of these approaches we have attempted to collect all necessary information directly from members of the respective user group via interviews and/or questionnaires. In the other two approaches, parts of the information have been collected through (internet) research. In the following paragraphs these four approaches are described in more detail.

1) Approach 1: Comprehensive Interviews

This is the 'classical' approach for user information collection, as described in literature [77], [24], [59]. The main method used in this information collection approach are semi-structured one-on-one interviews with members of the user group. These interviews include open-ended questions covering all relevant topics needed for the persona development: description of the job (tasks, workflows and work context), education, experience, skills and knowledge needed for the job, job-related goals and challenges, as well as motivational factors and annoying/frustrating aspects, attitudes about AI, personal traits and characteristics of 'typical' members of the user group. Usually, all necessary information for persona development can be collected through these interviews. However, optionally short questionnaires can be used to collect additional information, in case this is deemed to be necessary to broaden the view.

In our use case, we took this approach to collect information for two of the user groups identified in step 1 of the persona development process: We conducted face-to-face as well as online interviews with quality managers of three different pathology institutes and telephone and online interviews with people working as auditors at a notified body and at a certification organisation for medical products. Each of these interviews took about 50-90 minutes, with two people from interviewer side involved: one person asking the questions and interacting with the interviewee, and another person observing the interview and taking notes. In addition to the interviews, we collected information from further three auditors at notified bodies and certification organisations via short questionnaires.

At first glance, it seems that the interview approach for information collection is rather costly, since in addition to the time needed for conducting the interview, there is also considerable time needed for post-processing (e.g. transcription

User Groups for AI Application in Computational Pathology
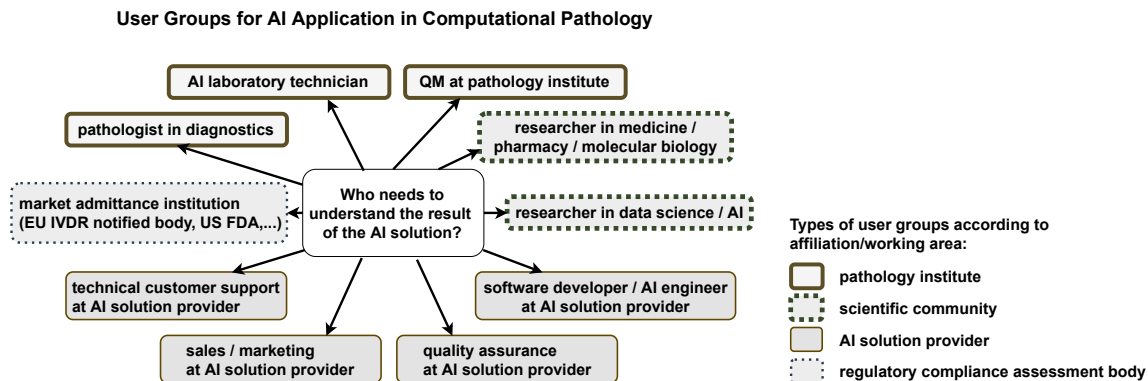
FIGURE 3. User groups for AI applications in digital pathology identified in a brainstorming and discussion session with domain experts from pathology institutes, industry and research.
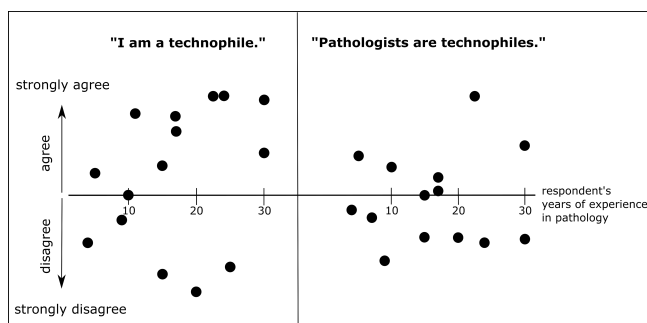
FIGURE 4. Pathologists' ratings of the statements "I am a technophile" and "Pathologists are technophiles."

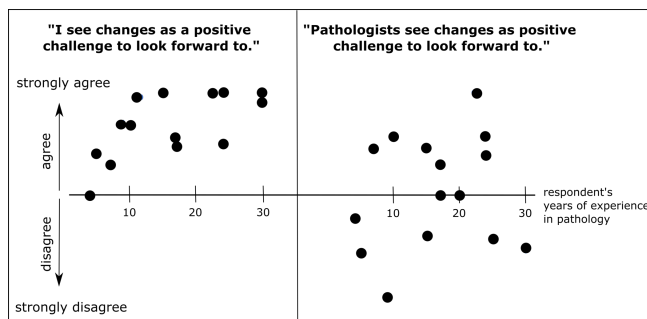**FIGURE 5.** Pathologists' ratings of the statements "I see changes as a positive challenge to look forward to." and "Pathologists see changes as a positive challenge to look forward to."

or coding) of the interview results. Nevertheless, we would recommend to conduct interviews for information collection whenever this is possible, as interviews provide much deeper insights and much more detailed information than questionnaires. Besides the actual answers of the interviewee to the interview questions, also other aspects observed in the course of the interview, such as asides made by the interviewee or the tone of voice used and aspects specifically stressed by the interviewee, give valuable insights, which cannot be obtained through questionnaires [78], [79].

2) Approach 2: Comprehensive Questionnaires

This approach for information collection tries to mimic the interview situation by using a comprehensive questionnaire with many open-ended questions covering all information topics needed for the development of personas. The goal is to minimise any bias introduced by the formulation of the questions and to allow the respondents to set the focus of their answers on aspects and details that are important from their point of view rather than influencing them by pre-defined answer options.

However, the drawback of this approach is the fact that people are not fond of spending a lot of time on questionnaires and people do not like writing much text when answering questionnaires. Therefore, it is extremely difficult to convince people to fill in such a long, time consuming questionnaire. Usually, only a small number of responses to this comprehensive questionnaire can be gained from well-disposed and motivated survey participants.

Furthermore, since the information obtained through questionnaires is not as rich and detailed as the information obtained through interviews, a small number of responses to the comprehensive questionnaire is not sufficient for the development of personas. Therefore, to gain additional input from more members of the user group, we used a short questionnaire in addition to the long comprehensive questionnaire. This short questionnaire provides an 'indirect view' by asking to list important skills and main motivational and annoying/frustrating aspects of the respective job as well as rating statements about 'typical' members of the user group.

In our use case, we applied this approach for information collection from software developers, one of the user groups identified in step 1 of the persona development process. The comprehensive questionnaire we created for collecting all information needed for persona development included 12 open questions. We could obtain 4 responses, all from people we knew personally, and respondents reported that they needed about 20 minutes to complete this questionnaire. With the additional short questionnaire we could obtain input
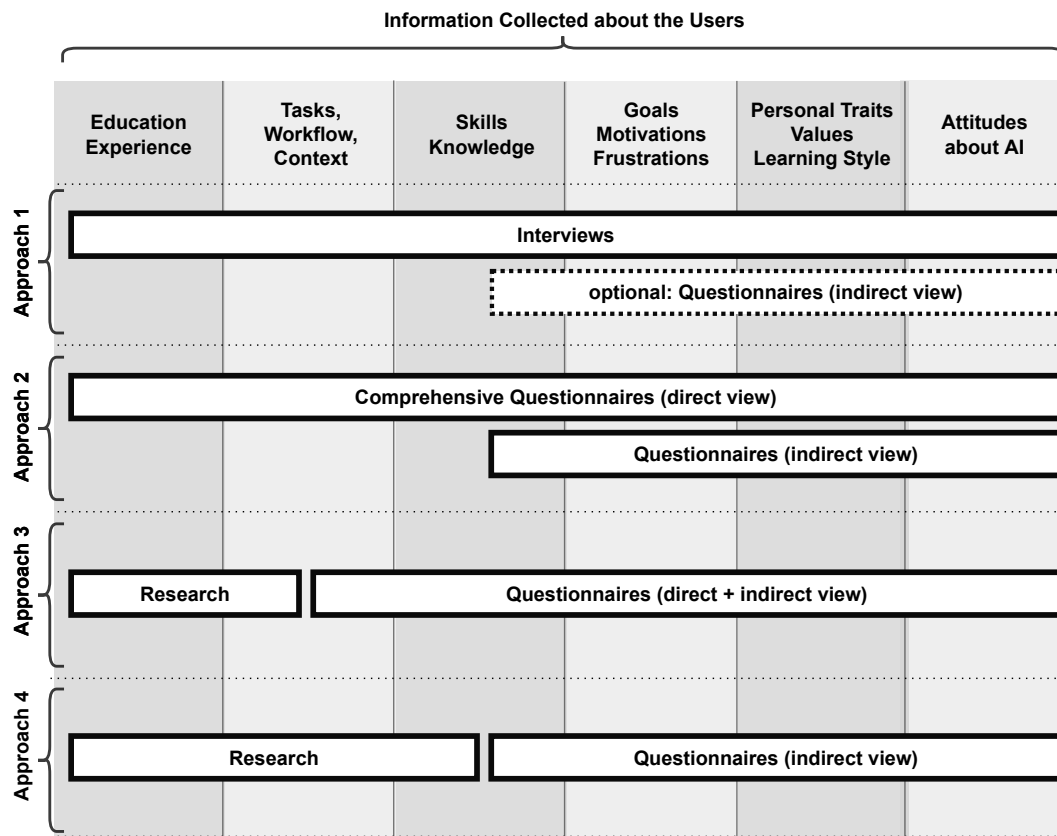
FIGURE 6. Different approaches for user information collection.

from further 10 software developers.

3) Approach 3: Combined View Questionnaires + Supplemental Research

This approach for information collection avoids the need for two separate questionnaires and combines 'direct' and 'indirect' views on the user group by including questions regarding the respondents themselves and questions regarding the respondents' perception of a 'typical' member of the user group. When designing this questionnaire, the aim was to create a questionnaire that takes less than 10 minutes to fill in. To achieve this, the number of open questions is reduced, and part of the information needed for developing personas is not covered by the questionnaire. Instead, information about job-related tasks and workflows as well as information about the typical education for this job is collected by additional research from literature, online resources or through on-site visits at a typical workplace.

In our use case, we applied this approach for collecting information about the user group of pathologists. In that case, descriptions of the education and training required for becoming a pathologist as well as detailed descriptions of the tasks of a pathologist could easily be retrieved from the internet. In addition, we had detailed knowledge regarding the workflows and work-context from previous work involv-

ing contextual inquiries and on-site visits in a pathology laboratory [80].

4) Approach 4: Short Indirect View Questionnaires + Extended Research

The goal of this approach is to further reduce the time needed for filling in the questionnaire. Therefore, only those parts of the information, which must be retrieved directly from members of the user group and cannot be retrieved through research, are included in the questionnaire. The focus of the information, which is retrieved via questionnaire directly from members of the user group, is mainly on the more person-related aspects, such as goals, motivations, frustrations, personal traits and attitudes towards AI. To minimise the number of questions and at the same time minimise bias, the questionnaire asks about the respondents' perception of a 'typical' member of the user group and thus provides an 'indirect' view on the user group. The job-related aspects, such as education/career, tasks, workflows and partly also skills and knowledge are collected via research. Valuable information about the characteristic tasks and activities to be done in a specific job as well as the skills, knowledge and education needed for a specific job can be found in respective job advertisements and articles of educational institutions. Interesting insights into typical careers and education paths

of people working in a specific job can be obtained from social media such as, for example, LinkedIn or Xing.

In our use case, we have taken this approach to collect information for sales representatives for software solutions in digital pathology, one of the user groups identified in step 1 of the persona development process.

5) Findings Regarding Different Approaches for User Data Collection

Experience from practical implementation of different information gathering approaches in our use case confirms the fact described by methodological literature [78], [79] that interviews provide much deeper insights and much more detailed information than questionnaires. Therefore, we would recommend to use interviews for user data collection whenever possible.

However, in our use case we were able to collect all user information needed for persona development also by data gathering approaches involving no interviews, which is valuable when data collection via ethnographic means is costly or impossible.

While we found that the approach to replace ethnographic user interviews by a comprehensive questionnaire with open questions covering all information topics was practically not feasible, the information needed for persona development could be obtained successfully through a combination of questionnaires and complementary research.

C. STEP 3: CONSOLIDATION AND ANALYSIS OF THE COLLECTED INFORMATION

The goals of this step in the development process of personas are (1) to get an overview of the information collected, (2) to distill the important findings from the heap of information collected, and (3) to decide, based on these findings, which personas to develop. The process applied to reach these goals is depicted in Fig. 7.

First, to get an overview of the available information, all collected information must be gathered in one place. Depending on the kind of information collected, this central information storage can, for example, be a database or a simple spreadsheet document. Thereby, throughout the whole process of organising, structuring, splitting and condensing the information, it must be taken care that for each piece of information the connection to the origin is preserved, so that it is, for example, possible to find out grouping characteristics later on in the process.

Once the collected information has been gathered in a central information storage, different methods are applied for consolidation and analysis depending on whether the information is structured or unstructured:

- *Visualisation Diagrams* (for example bar charts or scatter plots) support consolidation and analysis of structured information, such as categorical or numerical information obtained through closed questions in an interview or questionnaire. Examples of such visualisation diagrams created from questionnaire answers of
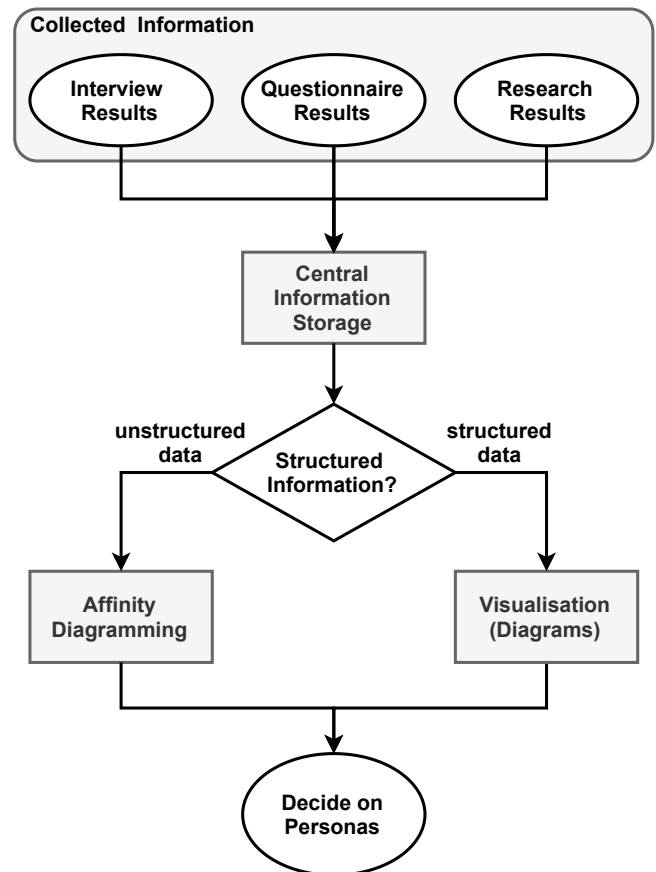
FIGURE 7. Process applied for consolidation and analysis of the collected information.

pathologists in our use case are shown in Fig. 5, 4, and 8.

- *Affinity Diagramming* supports consolidation and analysis of unstructured information such as, for example, information obtained through open questions in an interview or questionnaire. Thereby, the collected information is split into single aspects, all these single pieces of information are grouped in clusters based on their content relationships, and each of these clusters gets a label that summarises the information contained in that specific cluster.

From the Affinity Diagram or from the diagrams visualising the structured data it may become apparent that the members of a user group cluster into several subgroups regarding certain aspects, such as, for example, personal traits, education, approach towards new technologies, working style, etc. It is up to the persona development team to assess whether or not these differences are important and will have an influence on the usage of the product. For all aspects, where such differences might have an influence on the usage of the product, it should be taken care that each cluster seen in the data is represented by a distinct persona. When developing personas for AI applications, clusters with respect to users' attitude about AI or users' attitude about

new technologies should always be regarded as important, since these attitudes will have an influence on the usage of the AI application. Therefore, when the data indicate such differences in the users' attitudes, these different attitudes must be represented by the created personas.

To illustrate this, we take a practical example from our use case: The diagram shown in Fig. 8, which visualises the answers of pathologists to the statement "Pathologists like to work with new technologies and innovations," indicates that there are pathologists who like to work with new technologies and pathologists who do not. The user's attitude towards working with new technologies is regarded to be an important aspect that needs to be taken into account when designing or evaluating AI solutions for digital pathology, since these applications are currently perceived as brand new innovative technology and not yet commonly used in practice. Therefore, in our use case we decided to develop two different personas for pathologists in diagnostics: A pathologist persona type 1, who likes to work with new technologies and innovations, and a pathologist persona type 2, who is not as enthusiastic about new technologies.
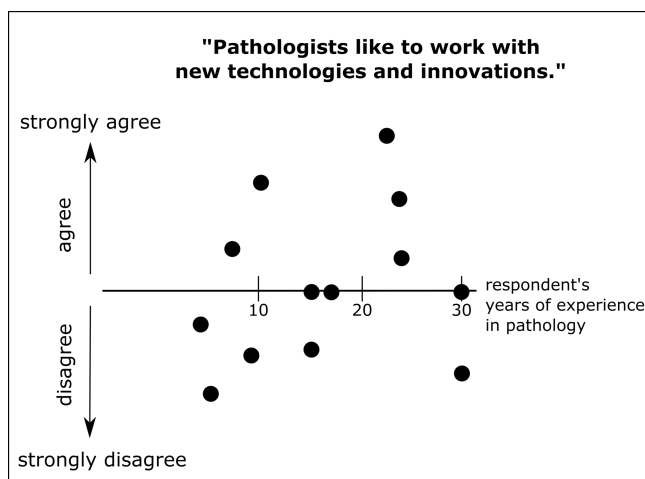
FIGURE 8. Diagram visualising the answers of pathologists regarding the statement "Pathologists like to work with new technologies and innovations."

D. STEP 4: CREATING THE FOUNDATION FOR PERSONAS

In this step of the development process of personas, a so-called foundation document is created for each persona. The foundation document contains all information about a specific persona in a structured way. There are many different templates and structures for foundation documents proposed in the literature - examples can be found in [59], [60], [81]. A section of the foundation document, which is specific for a persona developed for AI, is the attitude of the persona about AI and the attitude of the persona about new technology (if AI is perceived as new/innovative technology in the respective application domain). In our use case, we have structured the foundation documents, which were created to

form the basis of the user personas for AI solutions for digital pathology, into the following sections:

- Work (tasks, workflows, context)
- Education/knowledge/skills
- Personal traits
- Motivational factors
- Frustrations/hurdles
- Goals/values
- Attitudes towards AI / attitudes towards new technology

The foundation document serves mainly two purposes: First, this structured representation of all the information collected for a persona makes it easy to check for completeness of the information. If important pieces of information for a persona is missing, additional research can be done to fill these gaps. Second, this foundation document builds the basis of any further usage of the persona, such as for example visualisations of the persona (as described in the next section) or development of use cases and scenarios for the persona.

E. STEP 5: VISUALISING PERSONAS

In the final step of the process for creating a persona, the fictional person is made a tangible realistic character to help people empathize with the persona. To achieve this and bring the persona to life, the persona is visualised in a nice 1-page layout, including the persona's name, a picture showing the persona, and a narrative text about the persona's interests, values, lifestyle, attitudes and behavioural patterns, based on the information from the persona's foundation document. When visualising a persona created for AI, it is important that this persona sheet gives an indication about the persona's attitude towards AI and (if AI is perceived as new/innovative technology in the respective application domain) conveys also the persona's attitude towards new technologies. Examples of such persona sheets are shown in Fig. 9 and Fig. 10.

One central element for visualising the persona as a tangible person is the picture of the persona. Salminen *et al.* (2021) found that a realistic picture/photo visualising the persona increases empathy for the persona [82]. Furthermore, to avoid confusion, it is recommended not to use pictures of different "persons" to visualise one persona [83]. However, it should be kept in mind that pictures can increase stereotyping and are perceived differently by people with diverse cultural background and gender [84]. Therefore, pictures for visualising a persona must be chosen carefully, so that the chosen picture does not support stereotyping and helps to communicate the characteristic aspects as described in the persona's foundation document. However, we found that it is rather difficult to get photos, which are suitable for persona visualisation and published under a licence that allows their usage for this purpose. To solve this problem, we have created a set of fictional pictures suitable for visualisation of personas for AI, using the code from *thispersondoesnotexist.com* [85], [86]. Besides the fictional picture of the persona, also a fictional name, age and other fictional data (such as, for example, hobbies, family, ...) are added to the persona sheet

IEEE *Access*

Laura **LENG**

Pathologist

Age: 45

Education: Study of Medicine

6 COMPUTER LITERACY   **10** MEDICAL EXPERTISE

skill scale from 0 (Fundamental Awareness) to 10 (Expert)

SKILLS & KNOWLEDGE

clinical medicine

gastrointestinal pathology

accuracy and precision

structured working style

conscientiousness

MOTIVATIONS

rejoice in her work

make correct diagnoses to help patients

apply innovative technologies in diagnostics

FRUSTRATIONS

lack of collegiality   work overload

unnecessary delays and idle time

insufficient clinical information

lack of time for reading specialist literature

VALUES & ASPIRATIONS

stay up-to-date with new research results

open minded towards helpful innovations

continuous learning   family and nature

PERSONALITY AND BEHAVIOUR

Laura is kind of a person you would call a technophile. She is enthusiastic about innovations and new technological developments. Laura thinks that life long learning, and staying informed about new developments and research results is essential. Thus, Laura does not only have up-to-date technology gadgets like smart lights and home automation appliances in her house, but she is also open minded to try out and use innovative technologies at her workplace as a pathologist. From time to time she would wish that research results and new technologies could find their way into the laboratory's routines more quickly.

OCCUPATION

Laura has been working as a pathologist for more than 15 years. She had spent some years working abroad before she came to her current position as specialist for gastrointestinal pathology at a large hospital in Austria. Her daily routine work includes macropathology (cutting and grossing specimens) as well as microscopic examination of specimens in order to make diagnoses. Several times a week she participates in tumorboards to discuss oncology cases with colleagues from other medical disciplines. Laura likes her job. However, sometimes she feels a bit depressed about the lack of collegiality and feuding within the team.

OBJECTIVES AND GOALS

Laura wants to make correct diagnoses, so that each patient can receive the right therapies. Therefore she aims to keep up with the state-of-the-art diagnostic means and research results in her medical field for the sake of the patients.

ATTITUDE TOWARDS AI

Laura uses AI-based products, such as, for example, a voice assistant or recommender apps, quite frequently at home. However, she thinks that AI solutions for medicine must follow much stricter quality criteria. Therefore, in diagnostics she would only trust products, which are certified for that purpose.

FIGURE 9. Example visualisation of the pathologist persona.

Max **MOSER** Software Developer

Age: 36 * **Education:** Study of Computer Science

COMPUTER EXPERTISE
10

MEDICAL LITERACY
2

skill scale from 0 (Fundamental Awareness) to 10 (Expert)

SKILLS & KNOWLEDGE

analytical thinking · stamina and endurance
creativity · accuracy

MOTIVATIONS

the joy of solving problems · positive feedback
appreciation of his work by others · new technologies
interesting tasks · attractive salary

FRUSTRATIONS

unclear specifications · administrative workload
underestimation of needed resources
unrealistic deadlines

VALUES & ASPIRATIONS

good work/life balance · rejoice in his work
good team for working together and discussing
self-determined working hours and work flow

PERSONALITY AND BEHAVIOUR

Max has been interested in technical gadgets since his childhood. His friends have been asking for his advice regarding mobile phones, smart watches, computers or personal assistants, because Max is always informed about new technologies and knows state of the art-products. When Max uses a new device, he rather relies on his intuition and experience than reading the instruction manual.
Max spends a large part of his leisure time engaging in computer games. Once a week, he practices bouldering. He likes cooking and loves to try out new recipes.

OCCUPATION

Already during his studies at the university, Max started working part-time as software tester in the development-department of a large company. Since 6 years, he has been employed full-time as a software developer at a company that offers software solutions for the medical sector.
As part of a team of 8 software developers, Max is responsible for maintenance and implementation of new features into the company's software-product for analysis of histopathological images. Tasks and issues are distributed according to a plan among the developers in the team. Developers work on the assigned tasks on their own, but they are in close contact with each other. Since Max and the others are working from home most of the time, a short virtual meeting with the team mates is scheduled every day.

OBJECTIVES AND GOALS

Max aims to fulfil the programming tasks assigned to him in time and in a good quality. At the same time, he tries to maintain a good work/life balance.

ATTITUDE TOWARDS AI

Max is very enthusiastic about the new possibilities offered by AI-based products and services. He has got plenty of smart home gadgets in his flat and finds it convenient to use personalised recommender apps. Max thinks that AI technology can bring a lot of benefits also to the medical sector. Thus he is glad that in his job as a programmer he can contribute to practical implementation of this technological revolution.
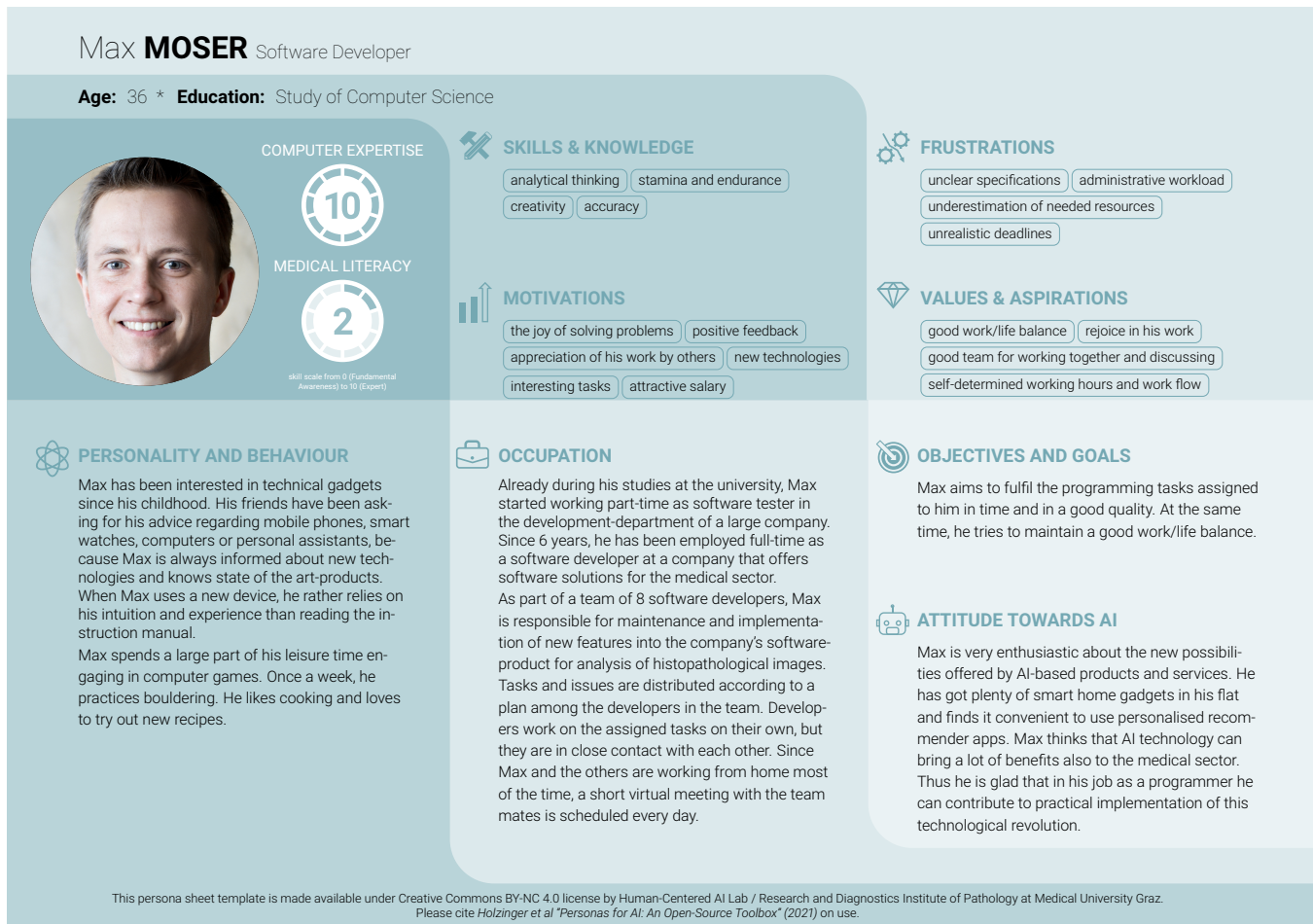
FIGURE 10. Example visualisation of the software-developer persona.

to make the persona more realistic. All these fictional parts must be chosen carefully as well. It must be taken care that they do not reinforce stereotyping, fit to the character and support communication of the persona's characteristics, so that the persona description conveys all important aspects gathered in the persona's foundation document. Furthermore, to validate the persona sheet we would recommend to obtain feedback from domain experts, or show the persona sheet to people from the respective user group and check whether they feel plausibly represented (as described in [55]).

IV. RESULTS

Our 5-step approach for developing personas for AI is similar to the 'traditional' personas development, but throughout the process the following aspects are specific for AI:

- In step 1 (Identification of (potential) user groups) not only the end-users of the AI solution but all groups of people, who need to understand/interpret the results delivered by the AI solution, shall be taken into account as (potential) user groups.
- In step 2 (Collection of information about the users) Information about the users' attitude towards AI and

(if AI is perceived as new and innovative technology in the respective application domain) also information about the users' attitude towards new technologies shall be collected in addition to the data/information usually gathered for 'traditional' personas.

- In step 3 (Consolidation and analysis of the collected information) clusters seen in the data with respect to users' attitude about AI or users' attitude about new technologies shall always be regarded as important, as these attitudes will have an influence on the usage of the AI solution. Therefore, such differences indicated by the data must be represented by the created personas.
- In step 4 (Creating the foundation for personas) the foundation document of a persona created for AI contains a specific section holding the information about the persona's attitude towards AI and (if AI is perceived as new and innovative technology in the respective application domain) the persona's attitude towards new technologies.
- In step 5 (Visualising Personas) it is important that the persona sheet conveys the persona's attitude towards AI and the persona's attitude towards new technologies.

Furthermore, in this paper we have also described possible ways of collecting user information for persona development by questionnaires and complementary research, when information collection via ethnographic means such as interviews or contextual inquiries is difficult or not possible, as during a pandemic. We have applied our approach in practice to develop 9 personas for AI applications in digital pathology and validated these personas in feedback loops with respective domain experts. Currently, these personas are used in the design and development process of the user interface for an AI solution supporting the analysis of Whole Slide Images in digital pathology. For this application, from the 9 personas developed in total, the personas representing pathologists and AI-laboratory technicians were defined as primary personas. These personas formed the basis for developing use cases and scenarios as well as deducing user requirements for the explanation component in the user interface. During our many years of experience developing applications for medicine and recent experience developing medical AI solutions, we quickly realized that our findings would be very valuable to researchers, developers, and users in the international community. Therefore, we have now summarized our findings in this paper and created our GitHub repository *Personas for AI*. As shown in Fig. 11, this repository contains a step-by-step guide to persona development, which includes recommendations and caveats based on our own practical experience. In addition, this repository provides clear descriptions of all the methods we recommend using, and supporting material and tools for each step of the persona development process. Our GitHub repository includes:

- Helpful impulses for identification of (potential) user groups for an AI application in step 1 of the persona development process.
- Concrete examples of interview guidelines and questionnaires, which can be used as inspiration and starting point for elaborating a successful strategy for the collection of user information in step 2 of the persona development process.
- Valuable practical tips for efficient and effective implementation of Affinity Diagramming to consolidate and analyse the collected user information in step 3 of the persona development process.
- A detailed example to showcase how the data and information, which forms the foundation of a specific persona, can be organised and structured in practice in step 4 of the persona development process.
- Inspiring practical examples of persona visualisations, which have been created to support user centered design and development of AI solutions for digital pathology in step 5 of the persona development process.
- Ready-to-use and easy to customise LaTeX and Microsoft PowerPoint templates for creating appealing visualisations of personas in portrait or landscape format in step 5 of the persona development process.
- A set of more than 4000 artificially created images,

which have been manually curated and annotated to comprise different genders and age groups as well as a wide range of emotions and expressions. As this picture set provides such a great variety of freely available pictures, which are suitable for persona visualisation, it is an invaluable resource for visualising personas in step 5 of the persona development process.

The toolbox is valuable for both, novice and expert persona developers: The easy to follow step-by-step guideline for persona development, the clear descriptions of the recommended methods and the practical examples are specifically beneficial for novice persona creators. However, although expert persona developers are familiar with the process and methods, they may still find the notes regarding AI-specific peculiarities useful. In this regard, the ready-to-use persona sheet templates and the large set of pictures of faces facilitate persona visualisation for both, novice and expert persona developers. The repository *Personas for AI* is publicly available at https://github.com/human-centered-ai-lab/PERSONAS.
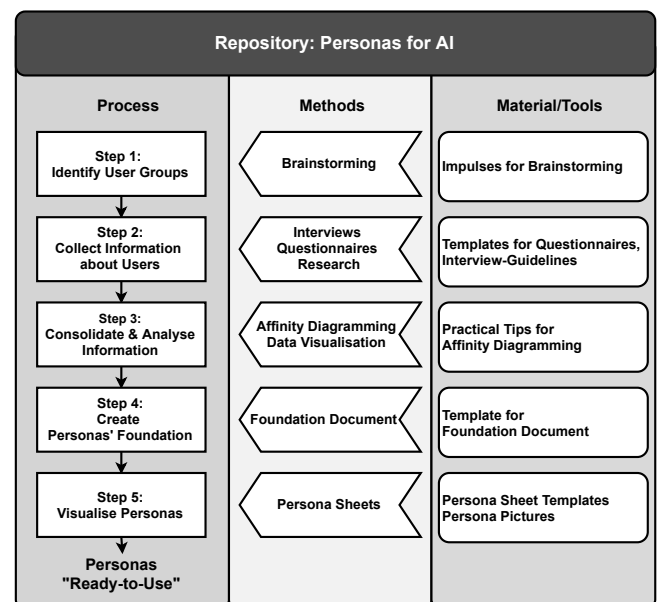
FIGURE 11. The repository *Personas for AI* contains a step-by-step guide for persona development, clear descriptions of the recommended methods, and useful tools for practical implementation, publicly available at https://github.com/human-centered-ai-lab/PERSONAS.

V. CONCLUSION
In human-centered design of software applications, the personas method is used to keep design and development of the product focused on the individual users' needs, abilities and preferences in order to create useful products which are usable in an easy, secure and trustworthy way. When it comes to AI applications, the needs of the users go beyond those for conventional software applications. AI applications are designed to assist users by, for example, reducing workload, improving task performance, or providing advice. To ensure

human–centred AI, they must be developed with users, their needs, their intended use and the context in mind, to be useful, usable, reliable, safe and trustworthy [87].

Causability must be added to the users' needs, since specifically in high-stake domains, such as, for example, in medicine, users need to understand the rationale and the trustworthiness underlying the results delivered by an AI application [88], [41]. Therefore, it is of utmost importance that AI applications are designed and developed with the users in mind, in order to achieve human-centered AI solutions with high security, usability and causability.

In this paper, we have described our 5-step approach for developing personas to support human-centered design of AI applications, and we have introduced practical examples from personas development for AI solutions for digital pathology to illustrate our approach. Based on our experience, we created free tools for developing personas for AI and made these publicly available in a repository as a contribution to support human-centered design and the development of AI applications. This repository *Personas for AI* is specifically targeted at the research community and small (start-up) companies, for which the available commercial solutions and tools supporting persona development are often not accessible.

REFERENCES

[1] D. P. C. Peters, A. Rivers, J. L. Hatfield, D. G. Lemay, S. Liu, and B. Basso, "Harnessing ai to transform agriculture and inform agricultural research," *IT Professional*, vol. 22, no. 3, pp. 16–21, 2020.

[2] M. S. Norouzzadeh, A. Nguyen, M. Kosmala, A. Swanson, M. S. Palmer, C. Packer, and J. Clune, "Automatically identifying, counting, and describing wild animals in camera-trap images with deep learning," *Proceedings of the National Academy of Sciences*, vol. 115, no. 25, pp. E5716–E5725, 2018. [Online]. Available: https://www.pnas.org/content/115/25/E5716

[3] D. A. Norman, "Design principles for human-computer interfaces," in *Proceedings of the SIGCHI conference on Human Factors in Computing Systems (CHI '83)*, 1983, pp. 1–10.

[4] S. K. Card, T. P. Moran, and A. Newell, *The psychology of Human-Computer Interaction*. Hillsdale (NJ): Erlbaum, 1983.

[5] B. Shneiderman, *Designing the user interface: Strategies for effective human-computer interaction*. Reading (MA): Addison-Wesley, 1987.

[6] R. C. Tumbleson, "Calculating machines," *Electrical Engineering*, vol. 67, no. 1, pp. 6–12, 1948.

[7] E. R. Harper, T. Rodden, Y. Rogers, and A. Sellen, *Being Human: Human-Computer Interaction in the year 2020*. Cambridge (UK): Microsoft Research, 2008.

[8] J. S. Seberger, M. Llavore, N. N. Wyant, I. Shklovski, and S. Patil, "Empowering resignation: There's an app for that," in *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, 2021, pp. 1–18.

[9] S. S. Sundar, "Rise of machine agency: A framework for studying the psychology of human–ai interaction (haii)," *Journal of Computer-Mediated Communication*, vol. 25, no. 1, pp. 74–88, 2020.

[10] W. Xu, M. J. Dainoff, L. Ge, and Z. Gao, "From human-computer interaction to human-ai interaction: New challenges and opportunities for enabling human-centered ai," *arXiv:2105.05424 [cs.HC]*, 5 2021. [Online]. Available: http://arxiv.org/abs/2105.05424

[11] A. Dix, J. Finlay, G. D. Abowd, and R. Beale, *Human-computer interaction*. Harlow: Prentice Hall, 1993.

[12] A. Przegalinska, L. Ciechanowski, A. Stroz, P. Gloor, and G. Mazurek, "In bot we trust: A new methodology of chatbot performance measures," *Business Horizons*, vol. 62, no. 6, pp. 785–797, 2019.

[13] M. Manca, F. Paternò, C. Santoro, E. Zedda, C. Braschi, R. Franco, and A. Sale, "The impact of serious games with humanoid robots on

[14] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[15] A. Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau, and S. Thrun, "Dermatologist-level classification of skin cancer with deep neural networks," *Nature*, vol. 542, no. 7639, pp. 115–118, 2017.

[16] A. Ortigosa, R. M. Carro, and J. I. Quiroga, "Predicting user personality by mining social interactions in facebook," *Journal of computer and System Sciences*, vol. 80, no. 1, pp. 57–71, 2014.

[17] R. Louie, A. Coenen, C. Z. Huang, M. Terry, and C. J. Cai, "Novice-ai music co-creation via ai-steering tools for deep generative models," in *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 2020, pp. 1–13.

[18] A. Calero Valdez, A. Schaar, M. Ziefle, A. Holzinger, S. Jeschke, and C. Brecher, "Using mixed node publication network graphs for analyzing success in interdisciplinary teams," in *Active Media Technology, Lecture Notes in Computer Science LNCS 7669*, R. Huang, A. A. Ghorbani, G. Pasi, T. Yamaguchi, N. Y. Yen, and B. Jin, Eds. Heidelberg, Berlin: Springer, 2012, pp. 606–617.

[19] S. Robert, S. Büttner, C. Röcker, and A. Holzinger, "Reasoning under uncertainty: Towards collaborative interactive machine learning," in *Machine Learning for Health Informatics: State-of-the-Art and Future Challenges*, A. Holzinger, Ed. Cham: Springer International Publishing, 2016, pp. 357–376.

[20] A. Holzinger, "Interactive machine learning for health informatics: When do we need the human-in-the-loop?" *Brain Informatics*, vol. 3, no. 2, pp. 119–131, 2016.

[21] A. Crisan and M. Correll, "User ex machina: Simulation as a design probe in human-in-the-loop text analytics," in *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, 2021, pp. 1–16.

[22] A. Holzinger, P. Kieseberg, E. Weippl, and A. M. Tjoa, "Current advances, trends and challenges of machine learning and knowledge extraction: From machine learning to explainable ai," in *Springer Lecture Notes in Computer Science LNCS 11015*. Cham: Springer, 2018, pp. 1–8.

[23] K. Vredenburg, J.-Y. Mao, P. W. Smith, and T. Carey, "A survey of user-centered design practice," in *Proceedings of the SIGCHI conference on Human factors in computing systems*, 2002, pp. 471–478.

[24] A. Cooper and P. Saffo, *The Inmates Are Running the Asylum*. USA: Macmillan Publishing Co., Inc., 1999.

[25] T. Miaskiewicz and K. A. Kozar, "Personas and user-centered design: How can personas benefit product design processes?" *Design Studies*, vol. 32, no. 5, pp. 417–430, 2011. [Online]. Available: http://dx.doi.org/10.1016/j.destud.2011.03.003

[26] L. Nielsen, "Design personas – new ways, new contexts," *Persona Studies*, vol. 4, pp. 1–4, 11 2018.

[27] A. Collins and D. Gentner, "How people construct mental models," in *Cultural models in language and thought*, D. Holland and N. Quinn, Eds. Cambridge: Cambridge University Press, 1987, pp. 243–265.

[28] A. Holzinger and H. Mueller, "Toward human-ai interfaces to support explainability and causability in medical ai," *IEEE COMPUTER*, vol. 54, no. 10, pp. 78–86, 2021.

[29] C. LeRouge, J. Ma, S. Sneha, and K. Tolle, "User profiles and personas in the design and development of consumer health technologies," *International journal of medical informatics*, vol. 82, no. 11, pp. e251–e268, 2013.

[30] A. Blandford, "Hci for health and wellbeing: challenges and opportunities," *International Journal of Human-Computer Studies*, vol. 131, no. 11, pp. 41–51, 2019.

[31] D. Schneeberger, K. Stoeger, and A. Holzinger, "The european legal framework for medical ai," in *International Cross-Domain Conference for Machine Learning and Knowledge Extraction, Springer LNCS 12279*. Cham: Springer, 2020, pp. 209–226.

[32] K. Stoeger, D. Schneeberger, and A. Holzinger, "Medical artificial intelligence: The european legal perspective," *Communications of the ACM*, vol. 64, no. 11, pp. 34–36, 2021.

[33] A. Holzinger, G. Langs, H. Denk, K. Zatloukal, and H. Mueller, "Causability and explainability of artificial intelligence in medicine," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 9, no. 4, pp. 1–13, 2019.

[34] A. Holzinger, "Usability engineering methods for software developers," *Communications of the ACM*, vol. 48, no. 1, pp. 71–74, 2005.

[35] A. Holzinger, A. Carrington, and H. Müller, "Measuring the quality of explanations: The system causability scale (scs). comparing human and

mild cognitive impairment older adults," *International Journal of Human-Computer Studies*, vol. 145, p. 102509, 2021.

machine explanations," *KI - Künstliche Intelligenz (German Journal of Artificial intelligence), Special Issue on Interactive Machine Learning*, vol. 34, no. 2, pp. 193–198, 2020.

[36] A. Holzinger, B. Malle, A. Saranti, and B. Pfeifer, "Towards multi-modal causability with graph neural networks enabling information fusion for explainable ai," *Information Fusion*, vol. 71, no. 7, pp. 28–37, 2021.

[37] E. L.-C. Law, V. Roto, M. Hassenzahl, A. Vermeeren, and J. Kort, "Understanding, scoping and defining user experience: a survey approach," in *Proceedings of the SIGCHI conference on human factors in computing systems*, 2009, pp. 719–728.

[38] H. Mueller, M. T. Mayerhofer, E.-B. V. Veen, and A. Holzinger, "The ten commandments of ethical medical ai," *IEEE COMPUTER*, vol. 54, no. 7, pp. 119–123, 2021.

[39] A. Holzinger, "From machine learning to explainable ai," in *2018 World Symposium on Digital Intelligence for Systems and Machines (IEEE DISA)*. IEEE, 2018, pp. 55–66.

[40] L. Longo, R. Goebel, F. Lecue, P. Kieseberg, and A. Holzinger, "Explainable artificial intelligence: Concepts, applications, research challenges and visions," in *Machine Learning and Knowledge Extraction*, vol. 12279 LNCS. Cham: SpringerInternational Publishing, 2020, pp. 1–16.

[41] J. Zhou, A. H. Gandomi, F. Chen, and A. Holzinger, "Evaluating the quality of machine learning explanations: A survey on methods and metrics," *Electronics 2021*, vol. 10, no. 593, 2021.

[42] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," *Nature*, vol. 323, no. 6088, pp. 533–536, 1986.

[43] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge (MA): MIT Press, 2016.

[44] M. Kusner, Y. Sun, N. Kolkin, and K. Weinberger, "From word embeddings to document distances," in *Proceedings of the 32nd International conference on machine learning - Volume 37*, ser. ICML'15. JMLR.org, 2015, pp. 957–966.

[45] R. Guidotti, A. Monreale, S. Ruggieri, F. Turini, F. Giannotti, and D. Pedreschi, "A survey of methods for explaining black box models," *ACM computing surveys (CSUR)*, vol. 51, no. 5, p. 93, 2019.

[46] H. Hagras, "Toward human-understandable, explainable ai," *Computer*, vol. 51, no. 9, pp. 28–36, 2018.

[47] O. Helmer and N. Rescher, "On the epistemology of the inexact sciences," *Management science*, vol. 6, no. 1, pp. 25–52, 1959.

[48] E. H. Shortliffe, R. Davis, S. G. Axline, B. G. Buchanan, C. C. Green, and S. N. Cohen, "Computer-based consultations in clinical therapeutics: explanation and rule acquisition capabilities of the mycin system," *Computers and biomedical research*, vol. 8, no. 4, pp. 303–320, 1975.

[49] K. Holzinger, K. Mak, P. Kieseberg, and A. Holzinger, "Can we trust machine learning results? artificial intelligence in safety-critical decision support," *ERCIM News*, vol. 112, no. 1, pp. 42–43, 2018.

[50] S. G. Vilbergsdottir, E. T. Hvannberg, and E. L.-C. Law, "Assessing the reliability, validity and acceptance of a classification scheme of usability problems (cup)," *Journal of Systems and Software*, vol. 87, pp. 18–37, 2014.

[51] M. Pröbster, M. E. Haque, and N. Marsden, "Perceptions of personas: The role of instructions," *2018 IEEE International Conference on Engineering, Technology and Innovation, ICE/ITMC 2018 - Proceedings*, 8 2018.

[52] N. Marsden, M. Pröbster, M. E. Haque, and J. Hermann, "Cognitive styles and personas: Designing for users who are different from me," in *Proceedings of the 29th Australian Conference on Computer-Human Interaction*, ser. OZCHI '17. New York, NY, USA: Association for Computing Machinery, 11 2017, pp. 452–456.

[53] L. Nielsen, *Personas - User Focused Design*. London: Springer, 2019.

[54] N. Marsden and M. Haag, "Stereotypes and politics: Reflections on personas," in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, CHI '16*. New York, NY, USA: Association for Computing Machinery, 2016, pp. 4017–4031. [Online]. Available: https://doi.org/10.1145/2858036.2858151

[55] N. Marsden and M. Proebster, "Personas and identity: Looking at multiple identities to inform the construction of personas," in *Conference on Human Factors in Computing Systems CHI 2019 - Proceedings*. Association for Computing Machinery, 5 2019, pp. 1–14.

[56] M. Burnett, S. Stumpf, J. Macbeth, S. Makri, L. Beckwith, I. Kwan, A. Peters, and W. Jernigan, "Gendermag: A method for evaluating software's gender inclusiveness," *Interacting with Computers*, vol. 28, pp. 760–787, 11 2016.

[57] A. Shekhar and N. Marsden, "Cognitive walkthrough of a learning management system with gendered personas," in *Proceedings of the 4th*

[58] C. H. Goh, N. Kulathuramaiyer, and T. Zaman, "Riding waves of change: A review of personas research landscape based on the three waves of HCI," in *Information and Communication Technologies for Development - 14th IFIP WG 9.4 International Conference on Social Implications of Computers in Developing Countries, ICT4D 2017, Yogyakarta, Indonesia, May 22-24, 2017, Proceedings*, ser. IFIP Advances in Information and Communication Technology, J. Choudrie, M. S. Islam, F. Wahid, J. M. Bass, and J. E. Priyatma, Eds., vol. 504, 2017, pp. 605–616. [Online]. Available: https://doi.org/10.1007/978-3-319-59111-7_49

[59] J. Pruitt and J. Grudin, "Personas: practice and theory," *DUX 03: Proceedings of the 2003 conference on Designing for user experiences*, pp. 1–15, 6 2003.

[60] J. S. Pruitt and T. Adlin, *The Persona Lifecycle - Keeping People in Mind Throughout Product Design*. USA: Elsevier Inc., 2006.

[61] L. Nielsen, *Engaging Personas and Narrative Scenarios - a study on how a user-centered approach influenced the perception of the design process in the e-business group at AstraZeneca*. PhD dissertation, 2004. [Online]. Available: https://www.personas.dk/wp-content/uploads/2019/11/samlet-udgave-til-load-1.pdf

[62] M. Blythe and A. Dearden, "Representing older people: Towards meaningful images of the user in design scenarios," *Universal Access in the Information Society*, vol. 8, pp. 21–32, 6 2009. [Online]. Available: http://teaching.shu.ac.uk/aces/amd/

[63] S. T. Acuña, J. W. Castro, and N. Juristo, "A HCI technique for improving requirements elicitation," *Information and Software Technology*, vol. 54, no. 12, pp. 1357–1375, 2012.

[64] A. Wilson, S. D. Paoli, P. Forbes, and M. Sachy, "Creating personas for political and social consciousness in hci design," *Persona Studies*, vol. 4, pp. 25–46, 11 2018.

[65] I. R. Floyd, M. C. Jones, and M. B. Twidale, "Resolving incommensurable debates: a preliminary identification of persona kinds, attributes, and characteristics," *Artifact*, vol. 2, pp. 12–26, 4 2008.

[66] Y. N. Chang, Y. K. Lim, and E. Stolterman, "Personas: From theory to practices," in *Proceedings of the 5th Nordic Conference on Human-Computer Interaction: Building Bridges*, ser. NordiCHI '08, vol. 358, no. August 2015. New York, NY, USA: Association for Computing Machinery, 2008, pp. 439–442.

[67] L. Nielsen and K. S. Hansen, "Personas is applicable- a study on the use of personas in denmark," in *Conference on Human Factors in Computing Systems - Proceedings*. Association for Computing Machinery, 2014, pp. 1665–1674.

[68] L. Nielsen, K. S. Hansen, J. Stage, and J. Billestrup, "A template for design personas: Analysis of 47 persona descriptions from danish industries and organizations," *International Journal of Sociotechnology and Knowledge Development*, vol. 7, pp. 45–61, 1 2015.

[69] P. Losana, J. W. Castro, X. Ferre, E. Villalba-Mora, and S. T. Acuña, "A systematic mapping study on integration proposals of the personas technique in agile methodologies," *Sensors*, vol. 21, no. 18, 2021. [Online]. Available: https://www.mdpi.com/1424-8220/21/18/6298

[70] J. Salminen, B. J. Jansen, J. An, H. Kwak, and S. gyo Jung, "Are personas done? evaluating their usefulness in the age of digital analytics," *Persona Studies*, vol. 4, pp. 47–65, 11 2018.

[71] J. Salminen, S. G. Jung, S. Chowdhury, D. R. Robillos, and B. Jansen, "The ability of personas: An empirical evaluation of altering incorrect preconceptions about users," *International Journal of Human Computer Studies*, vol. 153, 9 2021.

[72] D. Cabrero, H. Winschiers-Theophilus, and H. Mendonca, "User-created personas – a micro-cultural lens into informal settlement's youth life," in *Human Work Interaction Design. Work Analysis and Interaction Design Methods for Pervasive and Smart Workplaces. 4th IFIP 13.6 Working Conference, (HWID) 2015, London, UK, June 25-26,2015, Revised Selected Papers*, ser. IFIP Advances in Information and Communication Technology, J. A. Nocera, B. R. Barricelli, A. Lopes, P. Campos, and T. Clemmensen, Eds. Springer International Publishing, 2015, pp. 57–70.

[73] N. Marsden, J. Hermann, and M. Pröbster, "Developing personas, considering gender: A case study," in *Proceedings of the 29th Australian Conference on Computer-Human Interaction*, ser. OZCHI '17. New York, NY, USA: Association for Computing Machinery, 11 2017, pp. 392–396.

[74] J. Salminen, K. Guan, S.-G. Jung, and B. J. Jansen, "A survey of 15 years of data-driven persona development," *International Journal of Human–Computer Interaction*, pp. 1–24, 4 2021. [Online]. Available: https://www.tandfonline.com/doi/full/10.1080/10447318.2021.1908670

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2022.3154776, IEEE Access

IEEE *Access*

Holzinger *et al.*: Personas for AI: An Open Source Toolbox

[75] C. Meske, E. Bunde, J. Schneider, and M. Gersch, "Explainable artificial intelligence: Objectives, stakeholders, and future research opportunities," *Information Systems Management*, vol. 39, no. 1, pp. 53–63, 2022.

[76] V. Arya, R. K. E. Bellamy, P.-Y. Chen, A. Dhurandhar, M. Hind, S. C. Hoffman, S. Houde, Q. V. Liao, R. Luss, A. Mojsilović, S. Mourad, P. Pedemonte, R. Raghavendra, J. Richards, P. Sattigeri, K. Shanmugam, M. Singh, K. R. Varshney, D. Wei, and Y. Zhang, "One explanation does not fit all: A toolkit and taxonomy of AI explainability techniques," *CoRR2019*, vol. abs/1909.03012, 9 2019. [Online]. Available: http://arxiv.org/abs/1909.03012

[77] A. Cooper and R. Reimann, *About Face 2.0 - The Essentials of Interaction Design*. John Wiley & Sons, 2003. [Online]. Available: https://flylib.com/books/en/2.153.1/

[78] K. Williamson, "Chapter 16 - questionnaires, individual interviews and focus group interviews," in *Research Methods*, 2nd ed., K. Williamson and G. Johanson, Eds. Chandos Publishing, 2018, pp. 379–403. [Online]. Available: https://www.sciencedirect.com/science/article/pii/B9780081022207000169

[79] C. Phellas, A. Bloch, and C. Seale, "Structured methods: interviews, questionnaires and observation," in *Researching Society and Culture*, 3rd ed., C. Seale , Ed. United Kingdom: SAGE Publications Ltd, 2011, ch. 11, pp. 181–205.

[80] M. Kargl, P. Regitnig, H. Mueller, and A. Holzinger, "Towards a better understanding of the workflows: Modeling pathology processes in view of future AI integration," in *Artificial Intelligence and Machine Learning for Digital Pathology*, ser. Lecture Notes in Computer Science, A. Holzinger, R. Goebel, M. Mengel, and H. Müller, Eds. Cham: Springer International Publishing, 2020, vol. 12090, pp. 102–117.

[81] J. W. Castro and S. T. Acuña, "Extension of personas technique for the requirements stage," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 7545 LNCS, pp. 94–103, 2012.

[82] J. Salminen, S.-G. Jung, J. M. Santos, A. M. S. Kamel, and B. J. Jansen, "Picturing it!: The effect of image styles on user perceptions of personas," in *Proceedings CHI '21*. Association for Computing Machinery (ACM), 5 2021, pp. 1–16.

[83] J. Salminen, S. gyo Jung, J. An, H. Kwak, L. Nielsen, and B. J. Jansen, "Confusion and information triggered by photos in persona profiles," *International Journal of Human Computer Studies*, vol. 129, pp. 1–14, 9 2019.

[84] M. Pröbster, J. Hermann, and N. Marsden, "Personas and persons – an empirical study on stereotyping of personas," in *Proceedings of Mensch Und Computer 2019*, ser. MuC'19. New York, NY, USA: Association for Computing Machinery, 9 2019, pp. 137–146.

[85] T. Karras, S. Laine, and T. Aila, "A style-based generator architecture for generative adversarial networks," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019, pp. 4396–4405.

[86] T. Karras, S. Laine, M. Aittala, J. Hellsten, J. Lehtinen, and T. Aila, "Analyzing and improving the image quality of stylegan," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020, pp. 8107–8116.

[87] B. Shneiderman, "Human-centered artificial intelligence: Reliable, safe and trustworthy," *International Journal of Human–Computer Interaction*, vol. 36, no. 6, pp. 495–504, 2020.

[88] A. Holzinger, "Explainable ai and multi-modal causability in medicine," *Wiley i-com Journal of Interactive Media*, vol. 19, no. 3, pp. 171–179, 2020.

MICHAELA KARGL received the M.S. degree in electrical engineering with a focus on biomedical engineering from Graz University of Technology, Austria, in 1993 and the B.S. degree in information and computer engineering with a focus on human-computer interaction from Graz University of Technology, Austria, in 2017. She works as a Research Associate at the Diagnostic and Research Institute of Pathology at the Medical University Graz. Her main research interests lie in user research, user experience (UX) and usability.

BETTINA KIPPERER received the masters degree (Mag.) in English and American studies from the Karl-Franzens-University, Graz, in 2013. Additionally, she accomplished trainings and education in marketing and communications, intercultural management and worked long-term as a language service provider. She is a member of the Information Science and Machine Learning Group at the Diagnostics and Research Center of Molecular BioMedicine at the Medical University Graz, Austria, where her subjects of interest are user research, usability studies and reviewing.

MARKUS PLASS received the B. S. and the M. S. degree in biomedical engineering from Graz University of Technology, Austria. He is a member of the Information Science and Machine Learning Group at the Diagnostic and Research Center of Molecular BioMedicine at the Medical University Graz, Austria. At the university institute, he is in charge of digital pathology. His main research interests lie in digital pathology, machine learning and process optimization of digital work flows.

PETER REGITNIG received his M.D. from Graz University, Austria. He is working as Professor at the Diagnostic and Research Institute of Pathology at the Medical University of Graz. As Head of the diagnostic branch he is involved in the organization and optimization of routine processes. He is a board certified in pathology, cytology and molecular pathology with main expertise in breast- and gynecopathology and cytology. He was actively involved in implementations of laboratory information systems in the past. In recent years, his interests and activities in the application of digital pathology and AI applications for the use in histopathology have grown.

ANDREAS HOLZINGER is Visiting Professor for explainable AI at the University of Alberta, Canada since 2019 and Head of the Human-Centered AI Lab at the Medical University Graz, Austria. He received his Ph.D. in cognitive science from Graz University and his second Ph.D. in computer science from Graz University of Technology. He was elected ordinary member in the Academia Europaea, the European Academy of Sciences in the section Informatics, for his pioneering work in interactive machine learning with the human-in-the-loop. He is a full member of the European Lab for Learning and Intelligent Systems.

HEIMO MÜLLER received his Ph.D. in mathematics from Vienna University of Technology with a work on data space semantics. He is Head of the Information Science and Machine Learning Group at the Diagnostic and Research Center of Molecular BioMedicine at the Medical University Graz, Austria, and he is Scientific Officer at BBMRI-ERIC. His main research interests are in digital pathology and AI applications.

• • •