Personalized XAI for AI-driven Personalization

Cristina Conati[1] , **Katrien Verbert**[2] , **Oswald Barral**[2,3] and **Martijn Millecamp**[4]
[1,3]Department of Computer Science, UBC
[2,4]Department of Computer Science, KU Leuven

Abstract

Existing research on Explainable AI (XAI) suggests that having AI systems explain their inner workings to their end users can help foster transparency, interpretability, and trust. However, there are also results suggesting that such explanations are not always wanted by or beneficial for all users. These results indicate that research in XAI needs to go beyond one-size-fits-all explanations and investigate AI systems that can personalize explanations of their behaviors to the user's context and specific needs. This paper summarizes existing research and results toward personalized XAI, focusing on the role of long-term user traits for personalization and discusses directions for future research in this exciting new area

1 Introduction

Research on Explainable AI (XAI) aims to enable AI systems to explain their inner workings to their users, as a way to increase transparency, interpretability, and trust. Although there are encouraging results for this endeavor [Herlocker and J., 2000; Kulesza *et al.*, 2015; Ribeiro *et al.*, 2018], there are also findings suggesting that explanations are not always wanted by or beneficial for all users [Bunt *et al.*, 2007; Bunt *et al.*, 2012; Ehrlich *et al.*, 2011; Wang *et al.*, 2019]. There is a general agreement and some formal evidence that the need for explanations in AI systems may depend on context, e.g., the type of AI application and criticality of the targeted tasks [Bunt *et al.*, 2012], but there is also evidence that, given the same context, *user differences* such as cognitive abilities and personality traits play a role in defining when and how explanations may be useful and effective [Millecamp *et al.*, 2019; Millecamp *et al.*, 2021; Kouki *et al.*, 2019; Naveed *et al.*, 2018; Schaffer *et al.*, 2019; Conati *et al.*, 2021].

These results call for the need to investigate *personalized XAI*, namely how to create AI systems that understand to whom, when and how to deliver effective explanations of their actions and decisions.

AI-driven personalization has been an active field of research for several decades, spanning fields such as recommender systems, intelligent-tutoring systems, conversational agents, and affect-aware systems. To provide personalization, an AI system needs to have an adaptive loop in which it ac-

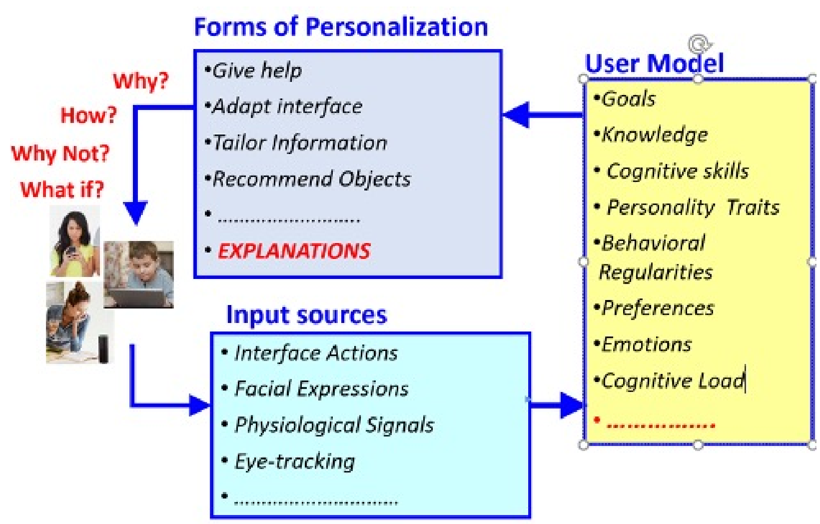

Figure 1: Adaptive Loop for AI-Driven personalization, with explanations added as a form of personalization.

quires a model of its user by inferring relevant user properties (see User Model in Figure 1) from available observations (see Input Sources in Figure 1) and decides how to personalize its behavior accordingly, to favor at best the goal of the interaction (see Forms of Personalization in Figure 1). Each box in Figure 1 provides some non-exhaustive examples of user properties, input sources and forms of personalization that have been explored in the literature, depending on the type of application and tasks that the AI system aims to support.

We see explanations as yet another element of personalization in the adaptive loop, where the system ascertains if and how to justify its behavior to the user based on its best understanding of user properties relevant to evaluate the need for explanation. What these relevant properties are is still largely unknown, hence, a key step toward personalized XAI is research to fill this gap.

Two general types of user properties have shown to be relevant for personalization: *long-term traits* that do not usually change over short periods of time (such as cognitive abilities and personality traits); and transient *short-term states* such as attention, interest and emotions.

We argue that, given a specific AI application, different types and forms of explanations may work best for different users, and even for the same user at different times, depending to some extent on both their long-term traits and short-term states. As such, our long-term goal is to develop personalized XAI tools that adapt dynamically to the user's needs by taking both these types of user factors into account.

In this paper, we focus on research investigating the impact of *long-term traits*, and how they may drive personalization. We present a general methodology to address these two questions, followed by two examples of how it was applied to gain insights on which long-term traits are relevant for personalizing explanations in a music recommender system and in an intelligent tutoring system (ITS). We then discuss how to move forward from these insights, and present research paths that should be explored to make personalized XAI happen.

2 Related Work on User Differences in XAI

The majority of the work on the role of individual differences in XAI pertain to *long-term traits*, with results supporting the importance of these individual differences for personalized XAI. For instance, Kouki et al. [2019] report a crowd-sourced study showing that users prefer item-centric to user-centric or socio-centric explanations, although preference for the latter type is modulated by levels of the Neuroticism personality trait. Naveed et al. [2018] looked at user decision-making style (rational vs. intuitive) in a study where they mocked up explanations reflecting different approaches to generate recommendations for buying a camera (e.g. content-based vs item-based) and found that it impacted which explanations were preferred. Looking at an AI agent that helps users play a decision game, Schaffer et al. [2019] found that explanations of the agents' suggestions were useful only for users who reported low ability at the game. In the next section, we illustrate a methodology for the systematic exploration of the role of long-term traits in XAI and two examples of findings based on this methodology.

3 Evaluating the Impact of User Long-Term Traits for Personalized XAI

The methodology underlying this research (summarized in Figure 2) entails having two versions of an AI-driven system: one that can provide explanations for its behaviors, and one that does not[1]. The added value of the explanations is measured in terms of changes in both *user performance* (e.g. time on task, task accuracy) and user *subjective experience* with and without explanations. The latter is crucial to assess constructs such as confidence and trust in the system, which are complementary to performance for assessing if explanations improve the user experience with the AI.

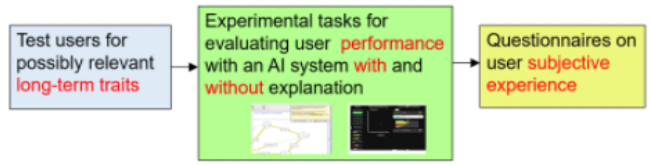

Figure 2: Overall methodology for evaluating the impact of user long-term traits on explanation effectiveness.

The possible impact of long-term user traits is gauged by measuring them with tests or questionnaires, and then include these measures in the statistical analysis of the impact of explanations on user performance and subjective experience. To avoid overwhelming the users with too many tests and cluttering the statistical analysis with too many measures, the traits tested should be selected based on pre-existing evidence that they can potentially impact how a user deals with explanations. The rest of this section presents two examples where this methodology revealed an impact of long-term user traits on explanation effectiveness and consequent insights for personalized XAI.

3.1 Long-Term Traits and Explanations in a Music Recommender

Millecamp et al. [2019] applied the methodology described in the previous section to investigate the impact of long-term traits on the effectiveness of explanations in a music recommender system. They implemented two versions of the system using the Spotify Web API. In both versions, a user can get personalized recommendations by expressing their music preferences on six musical attributes, using the sliders shown in Figure 3, part b. The system then generates a list of recommended songs that best match these preferences (Figure 3, part c), and the user can provide feedback on each recommendation by giving it a thumb up/down, and they can also listen to each song. In one version of the system, the recommendation process ends here.

In the second version, the user can ask *why* each recommended song was selected (see the "why" button in Figure 3, part c). In response, the system presents an explanation in the form of a grouped bar chart (Figure 3d) that shows how

[1] This is the simplest incarnation of the methodology, it is of course possible to have multiple versions of explanation functionalities in one study

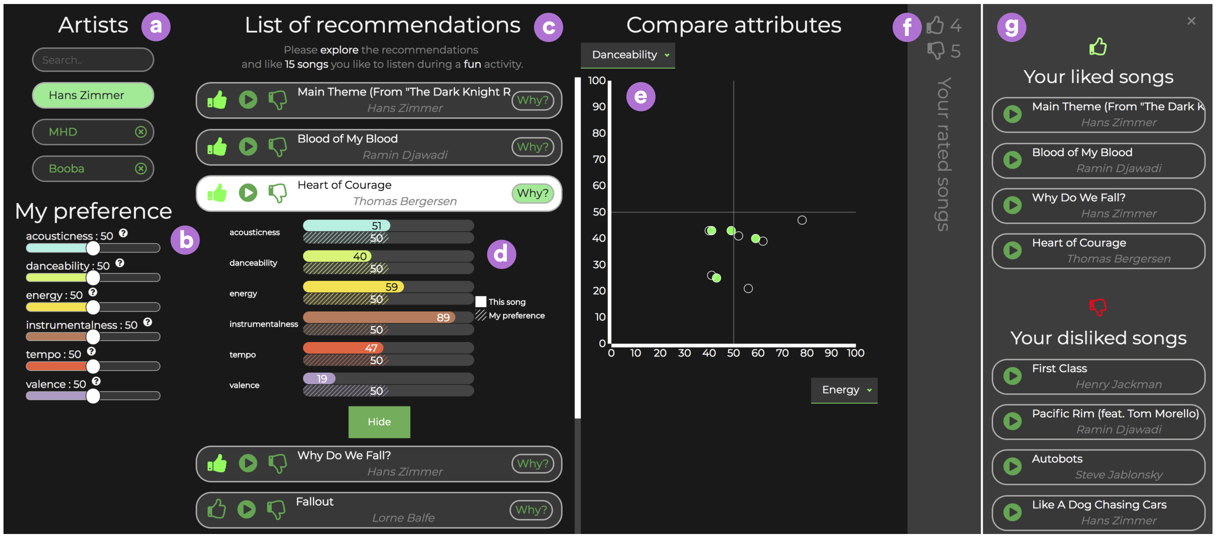

Figure 3: Explanation interface for music recommendations. The relevant areas are described in the text.

this song matches the expressed user preferences by displaying a comparison of the six preference attributes between the selected song and the user's selections. The user can also access a visualization showing how all songs compare for two specific attributes, namely the scatterplot shown in Figure 3d, where the user can select the attributes they want for the X and Y axes.

The two versions of the music recommender, with and without explanations, were evaluated using the methodology in the previous section, and a within-subject study. 71 users, recruited via Amazon Mechanical Turk were asked to create two different playlists, each with one of the two versions. Performance was evaluated in terms of the user's confidence in the playlists they created, which users self-reported after completing each playlist. Users also filled out a questionnaire that captures their perception of different aspects of the system, including the intention to use the system again, satisfaction with the recommended songs, trust in the system.

The user traits evaluated in this study, all measured through existing standard tests (see [Millecamp *et al.*, 2019] for details) included:

- Two perceptual abilities, *visual working memory* and *visualization literacy*, as well as the *locus of control* personality trait, (the extent to which one believes to be in control of their circumstances), all chosen because they have been shown to impact how people process visualizations like the ones used for the explanations.
- Level of *music expertise*, which affects how the user interacts with music recommendations and thus could also have an impact on how they process the consequent explanations.
- *Need for Cognition* (N4C) a trait defining one's inclination towards effortful cognitive activities, included because processing explanations can be cognitively effortful.

The study results revealed a significant [2] interaction effect between N4C and the presence/absence of explanation on *user confidence* in their playlists. Participants with lower N4C reported higher confidence when generating the playlist with the system with the explanations than with-out. Post-study interviews suggest that this effect is possibly due to the explanations helping the low N4C participants think about the recommendations they received more than they would do spontaneously, which in turn increased their confidence in their final playlist selections.

This result has a clear implication for personalization: if the music recommender knows that a user has lower N4C, it could monitor if they access the available explanations, and proactively offer them if they do not, given the positive effect of explanations for this type of users detected in the study. How can the system know if a user has low N4C will be discussed in a later section.

3.2 Long-Term Traits and Explanations in an ITS

The previous study targeted a task, music selection, which is not very cognitively demanding. Also, the AI to be explained, content-based filtering, is relatively intuitive. It is important to understand the need for personalized XAI in other contexts, for instance when the user tasks are more effortful and the AI to be explained is more complex.

Conati et al. [2021] started filling this gap by investigating the effect of long-term traits on explanations within the ACSP applet, an AI-driven interactive simulation that provides tools to explore and learn the Arc Consistency 3 (AC-3) algorithm for constraint satisfaction problems. The ACSP generates AI-driven hints to help a user leverage the available tools more effectively when it assesses that the user is not learning well from the interaction (the left in Figure 4 shows a screenshot of the ACSP interface with a hint). The assessment is done via a user modeling framework (FUMA, Framework for User Modeling and Adaptation [Kardan and Conati,

[2] statistical significance at $p < 0.05$

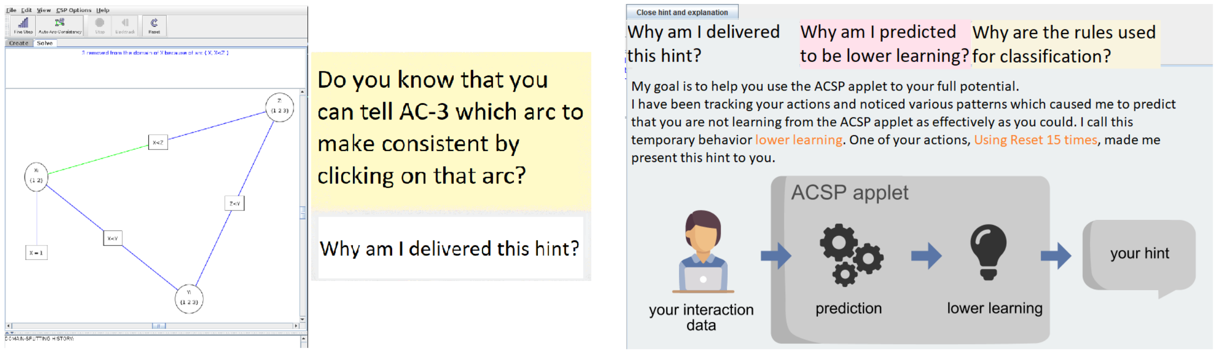

Figure 4: Left: Screenshot of the ACSP interface with a hint (yellow dialogue box) and access to its explanation. Right: First page of the explanation interface, which answers the question, *why am I delivered this hint* as shown by the selected tab at the top of the interface

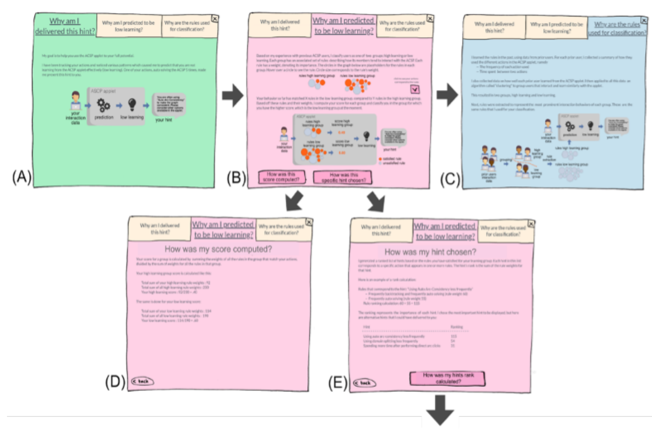

Figure 5: Flow Chart of Explanation Navigation (A) Why am I delivered this hint? (B) Why am I predicted to be lower learning? (C) Why are the rules used for classification? (D) How was this score computed? (E) How was this specific hint chosen? The arrow from (E) link to "How was my hint's rank calculated?" not shown.

2015] that uses unsupervised clustering and association rule mining on existing data of users interacting with a target system to discover classes of behaviors conducive or detrimental to learning. Supervised machine learning is then applied to the resulting clusters and accompanying association rules to build classifiers that predict in real-time whether a user is learning from the interaction and, if not, what are the behaviors responsible for this outcome. The ACSP uses FUMA predictions of low learning and the corresponding behaviors to generate personalized hints that guide the student towards a more effective usage of the available tools (dialogue box to the left in Figure 4).

A formal evaluation [Kardan and Conati, 2015] showed that the ACSP AI-driven hints improve student learning. Conati et al. [2021] wanted to ascertain if explaining to students the AI underlying the hints can further increase their uptake and effectiveness. Thus, they implemented an explanation functionality that conveys the motivations (*why*) and processes used (*how*) to generate each hint.

Explanations are automatically derived from the user model. The complexity of the underlying AI was handled by following three guiding principles from Kulesza et al.[2015], aiming to make the ACSP explanations to be *iterative* (i.e., accessible at different levels of detail), *sound* and *not overwhelming*. Determining how to convey coherent, clear and non-overwhelming information on the elements of the FUMA model required a lengthy process of iterative design and pilot evaluations, which identified three self-contained *why* explanations, as well as three *how* explanations.

These explanations aim to help the ACSP users gain a *global understanding* of the AI driving the ACSP hints, as well as a *local understanding* of the specific hints generated. The explanations were structured in an explanation interface where the user can choose how to navigate through them. This interface can be accessed once a hint is delivered, by clicking on the button "Why am I delivered this hint?" (dialogue box to the left of in Figure 4), which brings up the explanation window shown to the right of Figure 4, the entry point for the explanation interface.

The interface is structured around three tabs, each providing an incremental part of the rationale for hint computation, namely: "*Why am I delivered this hint?*" (Figure 5A); "*Why am I predicted to be lower learning?*" (Figure 5B) and "Why are these rules used for classification? (Figure 5C). In the second *why* explanations (Figure 5B), the user can access more details on *how* three specific aspects were computed: "*How* *was this score computed?*", "*How was this specific hint chosen?*" (Figure 5D–5E), and "*How was my hint's rank calculated?*" (not shown). The content of each tab is a combination of text and graphics (see [Conati *et al.*, 2021] for details)

The complexity of the explanation interface enhances the need to ascertain how different users interact with it, and how the interface could be personalized to suit specific user differences. A formal evaluation was conducted to answer these questions for long-term user traits by applying the methodology summarized earlier. The study compared *student learning* and their *perception of the ACPS hints* for two groups of students, one that worked with the ACSP with explanations and one without, and verified if these outcomes are impacted by a set of possibly relevant long-term traits. These include:

- Three *cognitive abilities* that may affect how users process the content of explanations: *Reading Proficiency* for the textual parts; *Visual Working Memory* and *Perceptual Speed*[3] for the graphical parts .
- A set of traits that may affect the perception of the explanations and hints, based on existing literature: *Need for Cognition*, included because of the impact on explanation effectiveness found in [Millecamp *et al.*, 2019]; the five *personality dimensions* of the Big 5 Model, (Agreeableness, Conscientiousness, Extraversion, Neuroticism, and Openness) included because at least one of them was found to have an impact on explanation preference in [Kouki *et al.*, 2019]; two dimensions of *Curiosity* added because some users had mentioned curiosity when asked reasons for wanting explanations in pilot studies.

All user traits were measured with standard instruments (see [Conati *et al.*, 2021] for details). The study results show a significant main effect of explanation interface: the group of students who had access to the explanations reported *higher trust* in the ACSP hints, feeling *more helped* by them, and more willing to *want the ACSP hints* in the future, compared to the group without explanation. The results also revealed significant effects providing evidence that the ACSP explanations might be more effective if they are personalized to specific student characteristics:

- student level of *reading proficiency* influences whether the ACSP explanations help reduce *student confusion* about the hints: students with higher reading proficiency reported less confusion with the hints when they had access to hint explanations than when they didn't. The opposite is true for students with lower reading proficiency, likely due to having to process the large amount of text currently used in the explanations, This result suggests that explanations could be personalized for students with lower reading ability by reducing the amount of text and possibly replacing it with more figures or animations.
- Student level of *conscientiousness*, which defines one's tendency to follow rules and instructions, impact whether the ACSP explanations improve *student learning*: students with low conscientiousness learned more with explanations of the ACSP hints than without. The opposite is true for students with high contentiousness. These findings have two implications for personalization. First, the system should actively encourage low conscientiousness users to access the explanations, because they are beneficial for them. Second, there is a need to investigate how to reduce the negative effects of explanations for users with high contentiousness and personalize the explanations accordingly.
- Students' level of *Need for Cognition (N4C)*, impacts how much they look at the explanations: students with high N4C paid more attention to explanation than their lower-level counterparts. This finding aligns with the fact that high N4C users are inclined toward effortful cognitive activities. Here too there is a clear implication for personalization: it is worthwhile exploring ways to make the ACSP proactively encourage low N4C users to pay more attention to its explanations, given that they improve user perception of the ACSP hints, and that these hints have been shown to promote learning [Kardan and Conati, 2015].

[3]Perceptual speed was included because it is the skill more directly related to processing visual information at large, included the diagrams used in the ACSP explanations

4 Personalized XAI: What's Next

The results in the previous sections are exciting because they provide evidence on the importance of personalization in XAI, and on how this personalization could be done. These findings, however, only scratch the surface of the knowledge necessary to develop this important new area. In this section, we discuss some of the research directions that are key to extend this knowledge.

4.1 How to Assess the Relevant User Traits

In all the work discussed in this paper, measures for the relevant user traits are obtained directly from the users via standard tests, questionnaires, etc. While personalization could be enabled by entering this information into the system, it is not realistic to assume that is always feasible to measure user traits in advance.

The alternative is to enable the system to infer the relevant user traits during interaction, as is done in the Adaptive Loop in Figure 1. Existing research has shown that long-term traits can be inferred in real-time from interaction data. For instance, Küster et al., [2018] predicted some personality traits from touchscreen interaction, and Conati et al., [2017] predicted some perceptual abilities from eye-tracking data as users were processing data visualizations. These results constitute an encouraging proof-of-concept that this level of user modeling is feasible, however, feasibility might depend on the type of interface, tasks and data sources available. Thus, it is important to investigate whether the long-term traits that are relevant for personalizing XAI in a specific system can be predicted during interaction, early enough to support personalization.

Millecamp, Conati and Verbert [2021] report preliminary results on predicting N4C from eye-tracking data of users creating a playlist with a variation of the recommender system with explanations described in a previous section. They show that a logistic regression classifier beats a majority class baseline in predicting if a user has high or low N4C. The accuracy reaches an acceptable level (67%) only after seeing most of the interaction data. Although this prediction would come too late for personalizing explanations in the current task, it can be leveraged if the user engages in subsequent tasks. In fact, it will be worthwhile to investigate how prediction accuracy changes if there is data to train classifiers over multiple tasks, as in [Barral *et al.*, 2020].

We are currently working on predicting the user traits found to be relevant for personalizing explanations in the ACSP applet (N4C, reading proficiency and contentiousness), by leveraging eye-tracking and action logs as data sources. Our preliminary results indicate that accurate early prediction is possible for reading proficiency by using eye-tracking data (68% accuracy after only 30 sec of interaction). and for conscientiousness by using log data (85% accuracy after about 2 min in tasks that tends to last over 15'), with no improvement when the data sources are combined. These results underline the relevance of exploring a variety of data sources for user modeling in personalized XAI.

In the long-term, as more datasets that connect user traits with XAI interfaces become available, it will be possible to explore if these datasets can be combined to train classifiers that are task and application-independent.

4.2 Looking at Short-Term User States

This paper focused on the role of long-term user traits for personalized XAI. As we discuss in the introduction, another important class of user properties is short-term states that change dynamically during interaction.

Short-term states that have been shown to impact user experience in HCI include confusion, cognitive load, and various affective states such as boredom, surprise and frustration, and there is substantial work on AI-driven personalization based on these states. In general, short-term states can function as local triggers for the system to evaluate when an intervention is needed, whereas long-term traits provide richer information of why the user needs help and how to provide it. For instance, in the context of personalized XAI for the ACSP applet discussed in the previous section, a system may rely on detection of confusion to realize that a user is having difficulty processing explanations, thus triggering a possible intervention to not give access to an explanation the next time. But if the system can also predict low reading proficiency, then it can consider the alternative action of offering an explanation format with less text, rather than removing explanations altogether. Only predicting low reading proficiency still enables the system to generate this more specific personalization, but it may cause taking the action when it is not necessary, because a low reading proficiency user may not always be confused by a textual explanation.

One research direction moving forward is to investigate the interplay between long-term traits and short-term states in personalized XAI, ideally via ablation studies with user models for different combinations of these two types of user properties.

4.3 How to Personalize the Explanations

Another large space for future research in Personalized XAI pertains to forms of personalization (Figure 1), namely investigating how explanation effectiveness depends on the relationship between user properties and properties of the explanations, such as *type* (e.g., why vs how the system generated its predictions), *delivery format* (e.g., level of detail, text vs graphics) and amount of *user control* in accessing the explanations (i.e., the degree to which the user chooses to access the explanation vs having the system offering them proactively)

The dimension of *user control* is still largely unexplored, and calls for studies that investigate the impact of making explanations user-initiated (e.g., via the click of a button as in [Millecamp *et al.*, 2019] and [Conati *et al.*, 2021]), permanently visible on the interface (as in [Millecamp *et al.*, 2021]), or system-initiated.

There are already some interesting results on *type* and *delivery format*. For instance, Wang et al. [2019] discuss initial results on the specific value of *why* and *why-not* explanations. Both Cotter et at. [2017] and Conati et al [2021] compared usage of *why* vs *how* explanations (in a recommender system and the ACSP applet, respectively). Tsai and Brusilovsky [2019] evaluated twelve visual explanations and three text-based explanations in a recommender system, reporting a preference for visual explanation over text-based explanation. But, there is much need for additional research to cover several other types of explanations [Wang *et al.*, 2019; Millecamp *et al.*, 2019], and delivery formats, along with how these dimensions interact with user differences across domains and applications.

4.4 AI-driven vs. User-Driven XAI Personalization

Research in AI-driven personalization often begs the question of why should the system do the personalization, as opposed to enabling users to perform the personalization themselves (aka *customization*). The main reason is that there is extensive research in HCI showing that users don't always want or know how to customize (see [Lalle and Conati, 2019] for an overview). Tsai and Brusilovsky [2019] provide initial evidence that this is also the case in XAI, reporting that users' preference for visual vs textual explanations did not match which of the two was more effective.

AI-driven personalization, however, does not mean that the system unilaterally decides. Instead, it should be seen as a dialogue between the system and the user, with the objective to help the user understand how to best use the system's features, always leaving the user with the ultimate control on what to do. In the context of Personalized XAI, this will involve investigating how to design effective interface tools for the user to access and personalize the explanations, but also enabling a system to monitor if and how the user leverages these tools and provide support as needed.

References

[Barral *et al.*, 2020] O. Barral, S. Lallé, G. Guz, A. Iranpour, and C. Conati. Eye-tracking to predict user cognitive abilities and performance for user-adaptive narrative visualizations, 2020.

[Bunt *et al.*, 2007] A. Bunt, J. McGrenere, and C. Conati. Under-standing the utility of rationale in a mixed-initiative sys-tem for gui customization. In *Proceedings of User Modeling 2007 (Lecture Notes in Computer Science*, page 147–156, 2007.

[Bunt *et al.*, 2012] A. Bunt, M. Lount, and C. Lauzon. Are explana-tions always important? a study of deployed, low-cost in-telligent interactive systems. In *Proceedings of IUI 2012, 2012 ACM International Conference on Intelligent User In-terfaces*, page 169–178, 2012.

[Conati *et al.*, 2017] C. Conati, S. Lallé, MdA Rahman, and D. Toker. Further results on predicting cognitive abilities for adaptive visualizations. In *Proceedings of IJCAI 2017, 26th International Joint Conference on Artificial Intelligence*, page 1568–1574, 2017.

[Conati *et al.*, 2021] Cristina Conati, Oswald Barral, Vanessa Putnam, and Lea Rieger. Toward personalized xai: A case study in intelligent tutoring systems, 2021.

[Cotter *et al.*, 2017] K. Cotter, J. Cho, and E. Rader. Explaining the news feed algorithm: an analysis of the 'news feed fyi. In *blog, in: Proceedings of the 2017 CHI, Conference on Human Factors in Computing Systems, Extended Ab-stracts pp.1553–1560*, 2017.

[Ehrlich *et al.*, 2011] K. Ehrlich, S. Kirk, J. Patterson, J. Rasmussen, S. Ross, and Gruen D. Taking advice from intelligent systems: the double-edged sword of explanations. In *Proceed-ings IUI 2011, 16th international conference on Intelligent user interfaces*, pages 125–134,, 2011.

[Herlocker and J., 2000] J.Konstan Herlocker and Riedl J. J. Explaining col-laborative filtering recommendations. In *Proceedings of the 2000 ACM Conference on Computer Supported Coopera-tive Work*, volume CSCW '00, page 241–250, 2000.

[Kardan and Conati, 2015] S. Kardan and C. Conati. Providing adaptive support in an interactive simulation for learning: An ex-perimental evaluation. In *Proceedings of CHI 2015, the 33rd Annual ACM Conference on Human Factors in Computing Systems*, page 3671–3680, 2015.

[Kouki *et al.*, 2019] Pigi Kouki, James Schaffer, Jay Pujara, John O'Donovan, and Lise Getoor. Personalized explanations for hybrid recommender systems. In *Proceedings of the 24th International Conference on Intelligent User Interfaces*, pages 379–390, 2019.

[Kulesza *et al.*, 2015] T. Kulesza, M. Burnett, W. Wong, and S. Stumpf. Principles of explanatory debugging to personalize interactive machine learning. In *Proceedings of IUI 2015, 20th International Conference on Intelligent User Interfaces*, pages 126–137,, 2015.

[Küster *et al.*, 2018] L. Küster, C. Trahms, and J. Voigt-Antons. Pre-dicting personality traits from touchscreen based interac-tions. In *Proceedings of Tenth International Conference on Quality of Multimedia Experience*, page 1–6, 2018.

[Lalle and Conati, 2019] S. Lalle and C. Conati. The role of user differences in customization: a case study in personalization for info-vis-based content. In *Proceedings of IUI 2019, 24th Interna-tional Conference on Intelligent User Interfaces*, page 329–339, 2019.

[Millecamp *et al.*, 2019] M. Millecamp, N. Htun, C. Conati, and K. Verbert. To explain or not to explain: the effects of personal characteristics when explaining music recommendations. In *Proceedings of IUI 2019, the 24th International Conference on Intelligent User Interfaces*, page 397–407, 2019.

[Millecamp *et al.*, 2021] Martijn Millecamp, Cristina Conati, and Katrien Verbert. Classifeye: classification of personal characteristics based on eye tracking data in a recommender system interface. In *Joint Proceedings of the ACM IUI 2021 Workshops co-located with 26th ACM Conference on Intelligent User Interfaces (ACM IUI 2021)*, volume 2903. CEUR Workshop Proceedings, 2021.

[Naveed *et al.*, 2018] Sidra Naveed, Tim Donkers, and Jürgen Ziegler. Argumentation-based explanations in recommender systems: Conceptual framework and empirical results. In *Adjunct Publication of the 26th Conference on User Modeling, Adaptation and Personalization*, pages 293–298, 2018.

[Ribeiro *et al.*, 2018] Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. Anchors: High-precision model-agnostic explanations. In *Proceedings of the AAAI conference on artificial intelligence*, volume 32, 2018.

[Schaffer *et al.*, 2019] James Schaffer, John O'Donovan, James Michaelis, Adrienne Raglin, and Tobias Höllerer. I can do better than your ai: expertise and explanations. In *Proceedings of the 24th International Conference on Intelligent User Interfaces*, pages 240–251, 2019.

[Tsai and Brusilovsky, 2019] C. Tsai and P. Brusilovsky. Evaluating visual ex-planations for similarity-based recommendations: user per-ception and performance. In *Proceedings of the 27th ACM Conference on User Modeling, Adaptation and Personaliza-tion*, page 22–30, Larnaca, Cyprus, 2019.

[Wang *et al.*, 2019] D. Wang, Q. Yang, A. Abdul, and Lim B. De-signing theory-driven user-centric explainable ai. In *Proceedings of the 2019 CHI Conference on Human Fac-tors in Computing Systems*, page 1–15, 2019.