HERMES

D2.1

Validated self-assessment and checklist for Trustworthy RS

Project Title	Trustworthy Multi-Objective and Multi-Stakeholder Recommenders
Contract No.	09103-03-V04-00336
Project start date	1.7.2024
Duration	24 months

Funded by
the European Union
NextGenerationEU

PLÁN OBNOVY

Grant agreement no.:	09I03-03-V04-00336
Project acronym:	HERMES
Project website:	https://kinit.sk/project/hermes
Project full title:	Trustworthy Multi-Objective and Multi-Stakeholder Recommenders
Project start date:	July 2024 (24 months)
Work Package:	WP2 – Trustworthy RS
Version:	1.0
Authors:	Michal Kompan

Project funded by VAIA - Research and innovation authority, under the call 09I03-03-V04, Grant agreement no. PPM_09I03-03-V04-00336		
Dissemination Level		
PU	Public	x
NP	Non-public, only for members of the consortium (including the Agency Services)	

Table of Contents

1. Introduction

This document presents the checklist for Trustworthy RS which is based on the integration of the Human values to the RS design. Deliverable presents the checklist and 2 applications assessed based on the proposed attributes.

The deliverable summarizes the methodology which is in detail described in the research paper submitted to the Expert Systems With Applications Journal.

2. Human Values as a basis for Trustworthy RS

Technological artifacts have always reflected the values and intentions of their creators, and today's AI-based systems intensify this dynamic by not only embedding human assumptions and biases but also actively shaping users' experiences, preferences, and even their ways of thinking. While debates about technological neutrality long predate AI - seen in examples like Robert Moses's discriminatory bridge designs - the adaptive, personalized, and opaque nature of modern AI makes its influence on human agency far more pervasive. As chatbots guide how students reason and recommender systems continuously predict and steer what we find desirable, these "black-box'' systems can subtly affect our choices without our awareness and can scale human biases across millions of users. Because such technologies increasingly mold our values, preferences, and sense of self, approaches like Value Sensitive Design are crucial for addressing their ethical impact from the outset.

Value Sensitive Design (VSD) provides a principled, ethically grounded framework for integrating human values, especially morally significant ones like fairness, privacy, autonomy, and justice. It considers the entire technology design process, recognizing both that some values have inherent moral standing and that their meanings vary across cultural contexts. Its methodology weaves together conceptual analyses of values and stakeholders, empirical research on how people prioritize and interpret those values, and technical investigations into how system features support or undermine them - an iterative, interdependent process that evolves as technologies and stakeholder interactions evolve. Because AI systems like modern recommender platforms adapt continuously to user behavior and operate within complex sociotechnical environments.

Recommender systems, now deeply embedded in domains from e-commerce to social media, shape not only individual user experiences but also broader societal dynamics, raising significant ethical concerns as they influence how people receive information and perceive the world. Their impact spans multiple stakeholders: users, creators, platforms, and even non-users whose often-conflicting values make multi-objective and multi-stakeholder design especially challenging. These tensions can have real consequences, such as when algorithmic amplification exposes vulnerable users to harmful content, including material related to eating disorders. Evaluating and designing such systems is further complicated by domain specificity, institutional opacity, the difficulty of balancing heterogeneous objectives, and the lack of established metrics for many human values. As a result, operationalizing stakeholder values remains an open problem, with current efforts hindered by limited participatory engagement and insufficient mechanisms for translating diverse ethical considerations into concrete system design choices.

3. Methodology

The process began by identifying all stakeholder groups potentially affected by the RS and refining this list through an initial questionnaire and workshop, after which stakeholders were clustered using a VSD-inspired analysis based on how strongly their values might be impacted and how much influence they had over design decisions. This process produced four categories, capturing different combinations of vulnerability and influence, and guiding subsequent recruitment efforts that prioritized some categories to ensure that highly affected yet often underrepresented perspectives were included. Recruited stakeholders then completed a value-ranking questionnaire, enabling comparison of value priorities across groups and informing the design of the workshops that followed.

Next, a conceptual phase aimed at identifying the human values most relevant to the recommender system's design and deployment, beginning with a questionnaire completed by the industry partner's technical team to clarify the system's purpose, motivations, stakeholders, and socio-technical context. Building on Stray et al.'s value framework, the researchers used a list of 31 values as a starting point to guide reflection and surface potential tensions. A value-ranking questionnaire based on this list was administered to all participants, who rated each value on a five-point scale; aggregated group-level rankings were used to identify each stakeholder group's top ten values. This shortlist informed the design of workshop scenarios that intentionally combined multiple values and perspectives to expose trade-offs in multi-objective, multi-stakeholder recommender systems, while still allowing additional, previously overlooked values to emerge organically through discussion.

Workshops formed the core of the study and followed a shared participatory methodology applied across different stakeholder groups: industry representatives in Workshop 2, external stakeholders identified earlier in Workshop 3, and technical collaborators in Workshop 4. Conducted mostly in person at the industry partner's premises, each three-hour session brought together members of the research team and invited participants, with one stakeholder joining remotely; all workshops were audio-recorded and supplemented with field notes for later analysis. Because participants varied widely in their familiarity with AI, recommender systems, and VSD, each workshop opened with a brief theoretical introduction covering AI ethics, human values in technology, and the rationale behind VSD, ensuring a common foundation for the subsequent activities.

The recordings were examined using an annotation framework in which each session was first analysed by a designated researcher and then cross-checked across the team, using a structured table to capture stakeholder positions, key discussion excerpts, associated values, inferred norms, design requirements, proposed solutions, and points of tension. The analysis combined deductive coding of values based on the Stray et al. Framework, with inductive identification of norms, requirements, and technical ideas emerging from stakeholder dialogue, with particular attention to conflicting expectations that signalled value tensions. The individual analyses were then merged into a summary organized around design requirements, which served as the central analytical unit linking stakeholder values to actionable system implications. To further structure this translation from abstract values to concrete design choices, the team adopted a simplified value hierarchy inspired by van de Poel, distinguishing values, norms, design requirements, and technical solutions as successive layers from ethical commitments to implementable system features.

4. Results

Across the three stakeholder groups, participants expressed distinct priorities shaped by their roles and proximity to the recommender system. Identified differences reflect broader patterns in which business actors emphasize organizational goals, external stakeholders foreground user-centric and societal concerns, and developers gravitate toward values that can be operationalized through technical solutions rather than broader socio-technical considerations.

	Platform provider group (COM)	**External stakeholders (SP, US, REC)**	**Developers (DEV)**
1.	Fairness, Equality and Equity	Safety and Security	Usefulness
2.	Accuracy (Factuality)	Privacy	Transparency and Explainability
3.	Usefulness	Usefulness	Control
4.	Transparency and Explainability	Liberty	Privacy
5.	Privacy	Fairness, Equality and Equity	Liberty
6.	Inspiration and Awe	Accuracy (Factuality)	Safety and Security
7.	Freedom	Freedom	Agency
8.	Sustainability	Control	Accuracy (Factuality)
9.	Well-being	Knowledge and Informativeness	Accessibility and Inclusiveness
10.	Control	Care, Compassion and Empathy	Accountability

5. Stakeholder-validated Check-list

From the perspective of recommender-system designers or the platform itself, it is essential not only to understand the diverse values and objectives of different stakeholders but also to determine how these can be meaningfully operationalized and embedded into the system. Accordingly, for each requirement we outlined a set of potential solutions that translate the underlying norms and values into concrete design choices, which serves as a checklist for guiding RS design. Based on the participative methodology, the check-list itself was validated with all stakeholders during the participative workshops organized during the project.

- Recommender system takes a specific set of selected categories as one of inputs and considers it during recommendation (e.g., LGBTQ+ friendly).
- The platform automatically extracts the relevant attributes from the freetext (service/product description) to extend item's metadata.
- The recommendation system and or the chat-bot are connected to the relevant authorities.
- Once a conversation is flagged as potentially harmful, the platform does not continue with recommendation.
- User registration is required to see some specific content.
- Recommender system implements several user models which reflect short and long term user preferences.
- The platform collects general statistics which allows a recommender system to recommend relevant items for anonymous stakeholders.
- Recommender system provides recommendations with desired values of novelty and serendipity.
- The platform offers the option to select between various user profiles (different roles).
- Incognito mode on the level of the platform is implemented.
- The platform empowers users to edit the information from the user model which should be used or not for recommendations.
- Turn off all personalization altogether permanently.
- The user is able to choose which categories that interest him.
- Implement diversity into a recommender system. A freshly identified interest has to be confirmed by the user.
- Do not use profiling of users by using the vulnerable and sensitive information about them.
- Avoid paid content (e.g., advertisement) in recommendations that conflicts with the interest of users.
- Add more-like-this a less-like-this function.
- The platform gathers statistics that allows to evaluate whether a specific stakeholder is "discriminated/disadvantaged" and provides mechanisms to prevent such biases.
- Recommender system provides explanations how the recommendation works at least the message "This item is recommended based on your previous purchases and interaction."
- Provide simple explanations such as "you might also like ..." or "others have also purchased ..."

- Recommender system provides personalized (i.e., adjusted to the level needs and with examples relevant to their interaction history) explanation to those who request it.
- A defined set of items (mature content) is not recommended even despite its availability on the web.
- Provide a base layer of positive explanation of how the recommender system is useful for the stakeholders.

6. Real-world applications assessment

Based on the defined checklist, 2 real-world applications have been assessed. The check-list can be used in both stages. For the platform and developers in the design and the implementation phase as well as retrospectively to evaluate the characteristics of existing RS applications. The deep knowledge about the platform or algorithm are needed for some of the characteristics, which are not publicly available.

Booking.com

Requirement	Available
Recommender system takes a specific set of selected categories as one of inputs and considers it during recommendation (e.g., LGBTQ+ friendly).	yes
The platform automatically extracts the relevant attributes from the freetext (service/product description) to extend item's metadata.	N/A
The recommendation system and or the chat-bot are connected to the relevant authorities.	N/A
Once a conversation is flagged as potentially harmful, the platform does not continue with recommendation.	N/A
User registration is required to see some specific content.	no
Recommender system implements several user models which reflect short and long term user preferences.	yes
The platform collects general statistics which allows a recommender system to recommend relevant items for anonymous stakeholders.	N/A
Recommender system provides recommendations with desired values of novelty and serendipity.	yes
The platform offers the option to select between various user profiles (different roles).	no
Incognito mode on the level of the platform is implemented.	no
The platform empowers users to edit the information from the user model which should be used or not for recommendations.	no
Turn off all personalization altogether permanently.	yes
The user is able to choose which categories that interest him.	yes
Implement diversity into a recommender system. A freshly identified interest has to be confirmed by the user.	no
Do not use profiling of users by using the vulnerable and sensitive information about them.	N/A
Avoid paid content (e.g., advertisement) in recommendations that conflicts	N/A

with the interest of users.	
Add more-like-this a less-like-this function.	no
The platform gathers statistics that allows to evaluate whether a specific stakeholder is "discriminated/disadvantaged" and provides mechanisms to prevent such biases.	N/A
Recommender system provides explanations how the recommendation works at least the message "This item is recommended based on your previous purchases and interaction."	no
Provide simple explanations such as "you might also like ..." or "others have also purchased ..."	yes
Recommender system provides personalized (i.e., adjusted to the level needs and with examples relevant to their interaction history) explanation to those who request it.	no
A defined set of items (mature content) is not recommended even despite its availability on the web.	N/A
Provide a base layer of positive explanation of how the recommender system is useful for the stakeholders.	N/A

Martinus.sk

Requirement	Available
Recommender system takes a specific set of selected categories as one of inputs and considers it during recommendation (e.g., LGBTQ+ friendly).	N/A
The platform automatically extracts the relevant attributes from the freetext (service/product description) to extend item's metadata.	yes
The recommendation system and or the chat-bot are connected to the relevant authorities.	N/A
Once a conversation is flagged as potentially harmful, the platform does not continue with recommendation.	N/A
User registration is required to see some specific content.	no
Recommender system implements several user models which reflect short and long term user preferences.	no
The platform collects general statistics which allows a recommender system to recommend relevant items for anonymous stakeholders.	N/A
Recommender system provides recommendations with desired values of novelty and serendipity.	yes
The platform offers the option to select between various user profiles (different roles).	no
Incognito mode on the level of the platform is implemented.	no
The platform empowers users to edit the information from the user model which should be used or not for recommendations.	no
Turn off all personalization altogether permanently.	no
The user is able to choose which categories that interest him.	yes
Implement diversity into a recommender system. A freshly identified interest has to be confirmed by the user.	no
Do not use profiling of users by using the vulnerable and sensitive information about them.	N/A
Avoid paid content (e.g., advertisement) in recommendations that conflicts with the interest of users.	N/A
Add more-like-this a less-like-this function.	no
The platform gathers statistics that allows to evaluate whether a specific stakeholder is "discriminated/disadvantaged" and provides mechanisms to prevent such biases.	N/A
Recommender system provides explanations how the recommendation works at least the message "This item is recommended based on your previous	yes

purchases and interaction."	
Provide simple explanations such as "you might also like ..." or "others have also purchased ..."	yes
Recommender system provides personalized (i.e., adjusted to the level needs and with examples relevant to their interaction history) explanation to those who request it.	no
A defined set of items (mature content) is not recommended even despite its availability on the web.	no
Provide a base layer of positive explanation of how the recommender system is useful for the stakeholders.	no

7. Conclusions

Applying the VSD approach to trustworthy recommender systems enables the translation of human values into concrete design requirements and technical solutions. Drawing on a qualitative case study conducted with an industry partner from the e-commerce sector, we created a list of solutions which uncover value tensions that would otherwise remain implicit in the design process of recommender systems. The validated checklist serves as tool for ensuring operationalization of basic requirements from all stakeholders.

The detailed methodology, results, limitations and discussion are presented in the paper “Applying value sensitive design to multi-objective and multi-stakeholder recommender systems in e-commerce” submitted to the Expert Systems With Applications Journal.