# Project HERMES: Designing Recommender Systems That Respect Human Values

**29.6.2026 - | Kempelenov inštitút inteligentných technológií**

**Recommender systems shape far more than what we click. They influence what we learn, what we desire, and how we see ourselves. As AI-driven platforms become deeply embedded in everyday life (from the videos we watch to the news we consume), their impact on human agency becomes impossible to ignore. Project HERMES was created to confront this reality head-on: to explore how recommender systems can be designed in ways that genuinely respect human values and optimize for requirements of different stakeholders.**

Why Human Values Matter in Recommender Systems

Technology has never been neutral. From Robert Moses's infamous bridge designs to today's algorithmic feeds, technological artifacts reflect the intentions, assumptions, and biases of their creators. But modern AI systems amplify this dynamic dramatically. They are adaptive, personalized, and capable not only of reflecting human values but subtly shaping them.

Chatbots influence how students reason. Social media feeds steer what we find desirable. Recommender systems can even nudge vulnerable users toward harmful content, such as material related to eating disorders. These systems operate at scale, meaning that a single design decision can ripple across millions of lives.

This is why Value Sensitive Design (VSD) is so important. VSD offers a principled framework for embedding human values as fairness, autonomy, privacy, or justice into the design process from the very beginning. It acknowledges that values differ across cultures and contexts, and that ethical considerations must evolve alongside the technology itself.

## The Challenge: Many Stakeholders, Many Values

Recommender systems don't affect just one group. They shape experiences for: Users, who want relevance, safety, and autonomy; Developers, who need clarity, feasibility, and measurable metrics; Businesses, who prioritize growth, engagement, and efficiency; and Non-users, who may still be impacted by algorithmic amplification.

These groups often hold conflicting priorities. And because many human values lack clear metrics, translating them into technical requirements is notoriously difficult. Project HERMES set out to tackle this challenge through a structured, participatory, and value-driven methodology.

## Inside Project HERMES: A Methodology Grounded in Participation

We began by identifying every group potentially affected by the recommender system. Through

questionnaires and workshops, stakeholders were clustered based on: 1) how strongly their values are impacted; 2) how much influence they have over design decisions. This produced four categories, highlighting groups that are highly affected yet often underrepresented, i.e., groups that became a priority for the workshops.

Next came a conceptual phase. Using a 31-value framework inspired by Stray et al., participants ranked values on a five-point scale resulting in a top-ten list for each stakeholder group, revealing both shared priorities and striking differences. During workshops participants engaged in scenario-based discussions, surfacing tensions, proposing solutions, and articulating what "value-aligned design" should look like in practice.

To translate abstract values into actionable design implications, we used a simplified value hierarchy inspired by van de Poel: Values → Norms → Design Requirements → Technical Solutions

## What We Learned: Divergent Priorities, Shared Responsibility

Across all groups, clear patterns emerged. Business stakeholders emphasized organizational goals and operational efficiency. External stakeholders focused on user well-being, societal impact, and ethical safeguards. Developers gravitated toward values that could be operationalized as transparency, accuracy, explainability, while often struggling with broader socio-technical concerns. These differences underscore a central insight of Project HERMES: Building trustworthy recommender systems requires navigating value tensions, not eliminating them.

The project demonstrated that designing ethical recommender systems is not just a technical challenge but it's a socio-technical one.

## What We Found Down the Road: Replicability of the Multi-Stakeholder Recommender System Research is Poor

AI's rapid growth has accelerated research output in computer science, pushing the field toward faster publication cycles, widespread preprint culture, and extensive code sharing. While these practices aim to improve transparency and reproducibility, they can unintentionally undermine replicability. This opens the deeper question of whether research results can truly be trusted. Our case study of five publications on multi-objective recommender systems shows how reusing shared code and evaluation pipelines can propagate logical and implementation errors across an entire research line. Instead of independent verification, agreement emerges around the same flawed artifacts, creating a false sense of progress and masking structural vulnerabilities in current experimental practices.

Reproducibility and replicability, though often conflated, serve different purposes: rerunning shared code is not the same as independently reconstructing a method, and easy-to-use artifacts can discourage genuine replication. Moreover, tools meant to promote transparency can become "black boxes" once their outputs appear stable, researchers tend to stop questioning. In the context of Trustworthy AI, where reliability, robustness, and accountability are central. Strengthening AI requires not only better algorithms, but also a more critical examination of how research is

conducted, validated, and shared.

**Article prepared by HERMES research team.**

https://kinit.sk/project-hermes-designing-recommender-systems-that-respect-human-values