Personalisation Under Constraints: From Recommender Systems to Clinical AI

Abstract

Shlomo Berkovsky's research spans foundational work in user modelling and recommender systems through to recent contributions in clinical AI, conversational systems, and human-AI collaboration. This overview synthesises major thematic streams across his work: (i) mediation and transfer of user models across heterogeneous representations and domains; (ii) privacy-preserving recommendation, including obfuscation and differential privacy; (iii) interaction-efficient recommendation and diversity-aware ranking; (iv) trust, reliance, and the behavioural dynamics of human interaction with recommenders and machine teammates; (v) personalised persuasion and behaviour-change technologies in health; (vi) physiological and behavioural sensing for user modelling; and (vii) translation of AI in healthcare, particularly the argument that effectiveness depends on interaction and workflow integration. Across streams, a consistent posture is evident: algorithms are treated as components of socio-technical systems operating under constraints (limited feedback, privacy requirements, cognitive limits, and high-stakes environments), so evaluation must jointly address technical performance and human factors. The synthesis concludes with open challenges suggested by this trajectory, including trust calibration under uncertainty, governance of implicit sensing, and evaluation frameworks connecting model behaviour to workflow and outcome-level effects in healthcare.

1. Introduction

Personalised technologies increasingly mediate everyday decisions: what information we see, which items we purchase, and how we engage with health services. The scientific community has responded with sustained advances in recommender algorithms, user modelling, and interactive intelligent systems. Yet as recommenders and AI systems move from laboratory settings to operational deployments, persistent constraints come to the fore. Users rarely provide abundant explicit feedback; privacy expectations shape what can be collected and retained; and the success of a system often depends on trust, interpretability, and fit to the surrounding workflow rather than raw predictive accuracy.

Across 2 decades, Berkovsky's work offers a coherent vantage point on these issues. Early contributions focus on mediation: how user models can be translated across techniques and representations in heterogeneous personalisation ecosystems [1,4,5]. In parallel, the work addresses privacy and data governance in collaborative filtering, ranging from obfuscation trade-offs to distributed architectures and differential privacy guarantees [2,8,9,12]. Subsequent research develops interaction-efficient recommendation paradigms explicitly treating user effort as a constraint and addresses ranking quality beyond accuracy through diversity optimisation [10,11,13]. A major human-centred strand studies trust and reliance dynamics in recommenders and in human-AI teaming, including cross-cultural variation [14,15,18,19]. Another strand links personalisation with persuasive technology in behaviour change, emphasising engagement mechanisms and realistic usage contexts [6,7,17]. Consistently with the constraints lens, Berkovsky has explored physiological and behavioural sensing, particularly eye tracking and biosignals, to infer personality traits and cognitive states during interaction [16,20,22,27]. These themes converge in healthcare and clinical AI, addressing conversational agents, documentation burden, clinician attitudes, and the broader position that clinical AI must be judged by interaction quality and workflow impact rather than algorithmic accuracy [21,23,26,28]. 2 recent papers extend the trajectory by consolidating personalised AI coaching as a research area and by articulating opportunities and risks in human-AI collaboration [29,30].

This overview synthesises Berkovsky's research and emphasises (i) recurrent conceptual commitments, (ii) methodological patterns, and (iii) open challenges that appear as personalisation and AI systems are embedded in social and clinical practice.

2. User model mediation and cross-domain personalisation

A foundational idea in Berkovsky's early work is that personalisation operates within heterogeneous environments. Systems represent preferences differently (e.g., ontologies, feature spaces, rating matrices), and users' behaviour is fragmented across contexts. Rather than forcing a single canonical representation, Berkovsky proposed mediation mechanisms to translate user modelling information across techniques [1] and later formalised mediation across representations [5]. The paper on mediation of user models for personalisation develops the core architecture: user model information can be transformed and aligned so that personalisation remains effective even when systems differ in representation and inference [4].

This emphasis on mediation naturally extends to cross-domain recommendation. Cross-domain mediation examines how preference information can be transferred between domains to support recommendation where data is sparse [3]. Beyond predictive accuracy, the problem also about coverage, cold-start relief, and practical reuse of behavioural traces; all concerns central to deployability. Even where modern methods use different technical idioms (transfer learning, representation learning), the mediation work remains conceptually valuable because it forces explicit reasoning about what is shared between domains and what should remain domain-specific.

The chapter on cross-domain recommender systems reflects maturation of the area and synthesises design and evaluation considerations at the field level [25]. Read together, the mediation stream of work shows continuity: from foundational modelling mechanisms for heterogeneous systems to consolidation of cross-domain recommendation as an established research direction.

3. Privacy-preserving personalisation

Privacy appears in Berkovsky's recommender research as a first-class design constraint rather than a post-hoc compliance requirement. The privacy-enhanced distributed collaborative filtering work addresses disclosure risks in decentralised settings while preserving recommendation accuracy [2]. This is notable because it treats privacy as an architectural property: where computation happens and what information flows between parties shapes risk.

The work on data obfuscation in collaborative filtering provides a clear empirical framing of the privacy-utility trade-off: perturbing user data can protect privacy but degrades recommendation quality [8]. Such analyses remain relevant because many privacy interventions in deployed systems, such as noise addition, coarsening, retention limits, are heuristic and designers need quantitative expectations about their costs.

A later shift is the adoption of stronger privacy paradigms. In [9], Berkovsky reframes privacy through governance: reduce risk by designing systems that function without storing user-level data long-term. The differential privacy framework for matrix factorisation brings formal guarantees into the recommender pipeline [12]. Differential privacy is not a universal solution, as utility loss, parameter sensitivity, and threat-model fit are non-trivial, but it raises the scientific standard by requiring explicit privacy budgets and clarifying what is protected.

A broader interpretation of this stream is that privacy is treated as a driver of algorithm and system design. This posture becomes particularly salient in healthcare contexts, where data sensitivity and governance requirements are elevated, even when the specific technical mechanisms differ from consumer recommendation settings.

4. Interaction-efficient recommendation

A central practical challenge in recommenders is interaction scarcity: users rarely provide dense explicit feedback, and repeated elicitation can impose unacceptable cognitive and time costs. Berkovsky's work tackles this constraint through minimal-interaction paradigms, which propose a discovery framing in which the system is evaluated by its ability to help users find relevant content with minimal feedback [11]. Another paper on minimal-interaction content discovery extends the framing and formalises it within interactive intelligent systems [13]. The main significance is the explicit coupling of algorithm design with behavioural realism: personalisation must succeed under sparse, noisy signals typical of online deployments.

Ranking quality beyond accuracy is also addressed, most notably through diversity. Diversity has multiple motivations: reducing redundancy, supporting exploration, and improving usefulness in settings where a single recommended items is insufficient. The paper on optimal greedy diversity provides an optimisation approach that makes diversity an explicit objective [10]. Importantly, this work operationalises experiential constructs (variety, coverage) into algorithmic terms that can be analysed, compared, and tuned.

Together, minimal-interaction and diversity research reinforce a theme that recommender systems should be studied as interactive systems whose success depends on experience quality under realistic constraints, beyond point-wise predictive performance.

5. Trust, reliance, and behavioural dynamics

As AI systems increasingly influence decisions, the question shifts from "is the system accurate" to "will people use it appropriately". Berkovsky's work addresses this through empirical studies and conceptual framing of trust and reliance.

2 papers study trust factors in movie recommenders and the dynamics of trust under different system performance regimes [14,15]. The focus on dynamics is important: trust is experience-dependent and users may under- or over-rely depending on explanations, perceived performance, and consistency of recommendations. In practical terms, this suggests that system design must manage expectations and communicate uncertainty or limitations, rather than assuming that better accuracy automatically yields better user outcomes. Trust is also examined across cultural contexts. The cross-cultural analysis of trust in recommenders provides evidence that trust-related behaviours and interpretations vary across populations, complicating generalisation from narrow samples [18]. This matters for both scientific external validity and real-world deployment, where systems operate across diverse user groups.

The human-machine teaming lens is extended in work on trust in a machine teammate, tracking how perceptions translate into decision-making [19]. This connects to later conceptual work on human-AI collaboration, which articulates challenges and opportunities that arise when both the AI system and the human are fallible [30]. By treating imperfection as a default setting, the focus shifts to coordination mechanisms: when should the AI defer, how should uncertainty be communicated to the user, how can accountability be maintained, and what interaction protocols reduce error amplification?

This stream bridges recommender system interaction aspects with clinical AI concerns: appropriate reliance is a core outcome when AI is embedded in decision workflows, especially in high-stakes settings.

6. Personalisation and persuasion for behaviour change

Berkovsky's work in persuasive technology demonstrates how personalisation can function as an intervention mechanism in health. The conceptual paper personalization-and-persuasion paper argues that persuasive strategies should be adapted to individuals, merging the logic of user modelling with behaviour-change design [7]. This idea anticipates contemporary interest in tailored and adaptive interventions and foregrounds a persistent methodological challenge: identifying which adaptations are beneficial for which users under which conditions. The physical activity motivating games line offers concrete intervention designs. The paper on virtual rewards for real activity demonstrates how game mechanics and incentive structures can be integrated with behaviour tracking to encourage engagement and physical activity [6].

Engagement mechanisms beyond games are explored in the study of push notifications in diet apps, which analyses how notifications influence engagement timing and task completion [17]. This work is important because many interventions fail at the level of sustained participation: users disengage, ignore prompts, or fail to integrate tools into daily routines. At the same time, behaviour-change research must be interpreted cautiously: engagement improvements do not automatically imply clinical benefit, and causal attribution is difficult due to confounding, attrition, and novelty effects. The scientific contribution lies in focussing on mechanisms of engagement that are testable and actionable for system design.

7. Physiological and behavioural sensing for user modelling

A distinctive research stream uses physiological and behavioural signals to infer traits and states relevant to personalisation. The first paper in this stream demonstrates that gaze behaviour during interaction can encode stable individual differences [20]. This is significant because it suggests an alternative to explicit questionnaires, as traits can be inferred through ordinary interaction signals. The trait inference agenda is extended to physiological responses to stimuli, showing that biosignals can support personality modelling under controlled exposure to images and videos [22]. These approaches can be viewed as an attempt to address the feedback scarcity problem: if systems can infer traits and states implicitly, they may reduce the need for explicit elicitation and could support more adaptive interfaces.

The paper on eye-tracking personality prediction in recommendation interfaces directly links sensing to recommendation settings [27]. It suggests that personality-related signals may manifest in how users attend to and process recommendation interfaces, potentially affecting acceptance, choice behaviour, and trust. In addition, cognitive state inference is addressed through work on indexing cognitive load using blood volume pulse features [16], providing a pathway to adaptation that responds to mental workload beyond preference profiles.

This sensing stream is scientifically promising but practically constrained. Physiological signals are sensitive, can be confounded by context, and raise privacy concerns that exceed conventional behavioural logging. The presence of strong privacy-focused themes in Berkovsky's research [2,8,9,12] provides a conceptual counter-weight: extracting more information about users can improve adaptation, but governance and consent become correspondingly central.

8. Clinical AI translation: conversational agents and documentation burden

More recently, Berkovsky's work increasingly centres on healthcare, where constraints and stakes are elevated: clinical time pressure, patient safety, regulatory obligations, and complex socio-technical workflows. A systematic review of personalisation in healthcare conversational agents maps the landscape of approaches and challenges, highlighting that personalisation is not just about content selection but also about appropriateness, safety, and presentation [23].

A set of contributions addresses clinical documentation burden and the possibility of AI support. The paper on challenges of developing a digital scribe articulates non-algorithmic barriers: privacy, workflow fit, safety, clinician control, and risk of shifting burden [21]. The co-design study with general practitioners advances a grounded design agenda for an AI documentation assistant embedded in consultations [24]. These works are notable for treating documentation AI as a clinical-system design problem rather than a narrow speech benchmark. Clinician perspectives on automation are examined in [28] that frames acceptance as dependent on alignment with professional roles and systems that support clinical agency. This resonates with the trust and teaming work [19] and human-AI collaboration framing [30]: systems should be designed for calibrated reliance and coordinated action.

In the position paper [26], Berkovsky crystallises a broad position: in clinical settings, predictive performance is necessary but insufficient; effective clinical AI must improve interaction and decision workflows. This mirrors recommender system insights about trust and experience, but the clinical context sharpens consequences: inappropriate reliance can harm patients, and poorly integrated tools can increase cognitive load or documentation burden.

9. Research synthesis

In addition to primary empirical and technical contributions, Berkovsky's record includes synthesis and agenda-setting work. The paper on research directions in session-based and sequential recommendation contributes to field-level framing, emphasising challenges and promising directions in a domain that has become central to modern recommenders [24]. The cross-domain recommender chapter similarly consolidates design and evaluation knowledge for a mature field [25].

At the frontier of current interest, the systematic review on personalised AI coaching technology synthesises an emerging area at the intersection of personalisation, persuasion, and AI-enabled guidance [29]. While AI coaching spans diverse contexts, the review provides a structured overview of what has been attempted, where evidence is thin, and how personalisation is operationalised. This sits naturally alongside the human-AI collaboration paper [30]: coaching systems are, in effect, long-term human-AI partnerships, so issues of trust calibration, transparency, and mutual adaptation are central to their real-world effectiveness.

10. Cross-cutting themes and summary

Across domains, several consistent themes appear. A repeated pattern is treating *constraints as core drivers of design*: heterogeneity of user models [1,4,5], privacy and governance [2,8,9,12], scarcity of interaction [11,13], cultural variability in trust [18], and workflow and safety constraints in clinical environments [21,24,28]. This posture increases external validity by aligning research problems with deployment realities.

Many contributions implicitly or explicitly shift evaluation away from accuracy-only optimisation toward *focus on interaction aspects*: user effort, perceived usefulness, trust dynamics, and decision workflow effects [11,13,14,15]. In healthcare, this becomes explicit as a design imperative [26]. The human-AI collaboration framing extends the shift to the team level, focusing on coordination under mutual fallibility [30].

The body of work offers a *methodological combination* of algorithmic design (diversity optimisation, privacy frameworks), HCI/behavioural studies (trust dynamics, cross-cultural analyses), and synthesis work. This variety is consistent with the view that personalisation is socio-technical: credible advances require both technical and human-centred evidence. In summary, Berkovsky's research contributes to an increasingly influential view of AI in general and personalisation particularly: systems should be evaluated not only by predictive performance but by how they support people under constraints.

References

[1] Berkovsky, S., Kuflik, T., Ricci, F. Cross-Technique Mediation of User Models. **Adaptive Hypermedia and Web-Based Systems**, 21–30, 2006.

[2] Berkovsky, S., Eytani, Y., Kuflik, T., Ricci, F. Enhancing Privacy and Preserving Accuracy of a Distributed Collaborative Filtering. **RecSys**, 9–16, 2007.

[3] Berkovsky, S., Kuflik, T., Ricci, F. Cross-Domain Mediation in Collaborative Filtering. **User Modeling** (conference), 365–369, 2007.

[4] Berkovsky, S., Kuflik, T., Ricci, F. Mediation of User Models for Enhanced Personalization in Recommender Systems. **User Modeling and User-Adapted Interaction** 18(3), 245–286, 2008.

[5] Berkovsky, S., Kuflik, T., Ricci, F. Cross-Representation Mediation of User Models. **User Modeling and User-Adapted Interaction** 19(1–2), 35–63, 2009.

[6] Berkovsky, S., Coombe, M., Freyne, J., Bhandari, D., Baghaei, N. Activity Motivating Games: Virtual Rewards for Real Activity. **CHI**, 243–252, 2010.

[7] Berkovsky, S., Freyne, J., Oinas-Kukkonen, H. Influencing Individually: Fusing Personalization and Persuasion. **ACM Transactions on Interactive Intelligent Systems** 2(2), e9, 2012.

[8] Berkovsky, S., Kuflik, T., Ricci, F. The Impact of Data Obfuscation on the Accuracy of Collaborative Filtering. **Expert Systems with Applications** 39(5), 5033–5042, 2012.

[9] Vallet, D., Friedman, A., Berkovsky, S. Matrix Factorization without User Data Retention. **PAKDD**, 569–580, 2014.

[10] Ashkan, A., Kveton, B., Berkovsky, S., Wen, Z. Optimal Greedy Diversity for Recommendations. **IJCAI**, 1742–1748, 2015.

[11] Kveton, B., Berkovsky, S. Minimal Interaction Search in Recommender Systems. **IUI**, 236–246, 2015.

[12] Friedman, A., Berkovsky, S., Kaafar, M.A. A Differential Privacy Framework for Matrix Factorization Recommender Systems. **User Modeling and User-Adapted Interaction** 26(5), 425–468, 2016.

[13] Kveton, B., Berkovsky, S. Minimal Interaction Content Discovery in Recommender Systems. **ACM Transactions on Interactive Intelligent Systems** 6(2), e15, 2016.

[14] Berkovsky, S., Taib, R., Conway, D. How to Recommend? User Trust Factors in Movie Recommender Systems. **IUI**, 287–300, 2017.

[15] Yu, K., Berkovsky, S., Taib, R., Conway, D., Zhou, J., Chen, F. User Trust Dynamics: Investigation Driven by Differences in System Performance. **IUI**, 307–317, 2017.

[16] Zhou, J., Arshad, S.Z., Berkovsky, S., Chen, F. Indexing Cognitive Load Using Blood Volume Pulse Features. **CHI**, 2269–2275, 2017.

[17] Freyne, J., Yin, J., Brindal, E., Hendrie, G., Berkovsky, S., Noakes, M. Push Notifications in Diet Apps: Influencing Engagement Times and Tasks. **International Journal of Human-Computer Interaction** 33(10), 833–845, 2017.

[18] Berkovsky, S., Taib, R., Hijikata, Y., Braslavski, P., Knijnenburg, B. A Cross-Cultural Analysis of Trust in Recommender Systems. **UMAP**, 285–289, 2018.

[19] Yu, K., Berkovsky, S., Taib, R., Zhou, J., Chen, F. Do I Trust My Machine Teammate? An Investigation from Perception to Decision. **IUI**, 460–468, 2019.

[20] Berkovsky, S., Taib, R., Koprinska, I., Wang, E., Zeng, Y., Li, J., Kleitman, S. Detecting Personality Traits Using Eye-Tracking Data. **CHI**, 2019.

[21] Quiroz, J.C., Laranjo, L., Kocaballi, A.B., Berkovsky, S., Rezazadegan, D., Coiera, E. Challenges of Developing a Digital Scribe to Reduce Clinical Documentation Burden. **NPJ Digital Medicine** 2, e114, 2019.

[22] Taib, R., Berkovsky, S., Koprinska, I., Wang, E., Zeng, Y., Li, J. Personality Sensing: Detection of Personality Traits Using Physiological Responses to Image and Video Stimuli. **ACM Transactions on Interactive Intelligent Systems** 10(3), e18, 2020.

[23] Kocaballi, A.B., Berkovsky, S., Quiroz, J.C., Laranjo, L., Tong, H.L., Rezazadegan, D., Briatore, A., Coiera, E. The Personalization of Conversational Agents in Healthcare: Systematic Review. **JMIR** 21(11), e15360, 2019.

[24] Kocaballi, A.B., Ijaz, K., Laranjo, L., Quiroz, J.C., Rezazadegan, D., Tong, H.L., Willcock, S., Berkovsky, S., Coiera, E. Envisioning an Artificial Intelligence Documentation Assistant for Future Primary Care Consultations: A Co-design Study with General Practitioners. **JAMIA** 27(11), 1695–1704, 2020.

[25] Ferrari Dacrema, M., Cantador, I., Fernandez-Tobias, I., Berkovsky, S., Cremonesi, P. Design and Evaluation of Cross-Domain Recommender Systems. In: Ricci, F., Rokach, L., Shapira, B. (eds.) **Recommender Systems Handbook**, 485–516. Springer, 2022.

[26] Berkovsky, S., Coiera, E. Moving Beyond Algorithmic Accuracy to Improving User Interaction with Clinical AI. **PLOS Digital Health** 2(3), e0000222, 2023.

[27] Chen, L., Cai, W., Yan, D., Berkovsky, S. Eye-Tracking Personality Prediction on Recommendation Interfaces. **User Modeling and User-Adapted Interaction** 33(1), 121–157, 2023.

[28] Fraile Navarro, D., Kocaballi, A.B., Dras, M., Berkovsky, S. Collaboration, not Confrontation: Understanding General Practitioners' Attitudes Towards Natural Language and Text Automation in Clinical Practice. **ACM Transactions on Computer-Human Interaction** 30(2), e29, 2023.

[29] Vitale, J., Cenacchi, F., Kraus, M., Orji, F., Richards, D., Orji, R., Conati, C., Berkovsky, S. Personalised AI Coaching Technology: A Systematic Mapping Review. **ACM Transactions on Interactive Intelligent Systems**, to appear, 2026.

[30] Berkovsky, S., Jacucci, G. Imperfect AI, Imperfect Humans, Effective Teams: Challenges and Opportunities of Human-AI Collaboration. **ACM Transactions on Interactive Intelligent Systems** 16(2), e10, 2026.