When Personalization Conflicts with Sustainability: A System-Level Perspective

Alan Said[1]

[1]University of Gothenburg, Gothenburg, Sweden

Abstract

Personalized systems increasingly shape what people consume, watch, learn, and purchase at global scale, yet they are typically optimized for engagement and short-term user satisfaction while their broader environmental and societal effects are rarely accounted for. This creates a structural tension: systems designed to increase interaction are incentivized to drive activity and consumption, while sustainability goals often require reduction, moderation, and efficiency. In this paper, we argue that this tension is not incidental, but follows from how personalization systems are conceived, optimized, and evaluated. We structure this tension into a set of recurring conflicts, including objective misalignment, efficiency–impact trade-offs, system-level externalities, and ecosystem blindness, and show how they manifest in real-world settings. Rather than proposing a single solution, we argue for a shift in perspective. Instead of optimizing isolated metrics, personalization systems need to be understood in terms of how they interact with the systems around them and the behaviors they induce. Our goal is to make these interactions visible and to provide a basis for developing systems that better account for their broader impact.

Keywords

Reflection, evaluation, call for action

1. Introduction

Recommender systems are no longer just ranking algorithms. They are embedded in nearly every large-scale digital platform and increasingly function as decision-making infrastructures that shape what users see, buy, learn, and consume. This influence is not marginal. It operates at the scale of billions of interactions per day across domains such as media, e-commerce, education, and communication. As a result, recommender systems do not merely reflect user preferences; they actively shape behavioral patterns over time.

Despite this central role, the dominant optimization paradigm in recommender systems remains narrowly defined. Systems are typically designed to maximize engagement metrics such as clicks, watch time, and conversion rates [1]. These metrics work well as proxies for platform success, but they are fundamentally incomplete. They capture short-term interaction rather than long-term outcomes, emphasize individual utility over system-level effects, and remain largely agnostic to environmental impact.

At the same time, sustainability has moved from a peripheral concern to a central constraint across technological domains [2]. Machine learning systems contribute to environmental impact both directly, through computation, and indirectly, through the behaviors they induce [3, 4]. This creates a structural tension. Systems optimized for engagement are incentivized to increase activity and consumption, while sustainability goals often require the opposite: reduction, moderation, and efficiency.

This tension becomes concrete in practice. In fashion e-commerce, systems designed to increase conversion can encourage behaviors such as bracketing [5], where users order multiple variants of a product and return most of them. From the user's perspective, this is a rational response to uncertainty. At the system level, it translates into increased return volumes, transportation, and waste.

In this paper, we argue that this misalignment is not the result of isolated design decisions. It follows from the objectives and evaluation practices that define modern personalization systems. To make

Joint Proceedings of the ACM UMAP Workshops 2026, UMAP 2026, June 8–11, 2026, Gothenburg, Sweden

alansaid@acm.org (A. Said)
0000-0002-2929-0529 (A. Said)

this explicit, we introduce a taxonomy of recurring conflicts that link personalization objectives to system-level sustainability outcomes. Our goal is not to propose a single solution, but to make these interactions visible and to provide a basis for developing systems that better account for their broader impact.

2. Background and Related Work

Over the past decade, recommender systems have become increasingly complex, often driven by advances in deep learning. These models deliver measurable improvements in predictive accuracy, but the gains are frequently small relative to the computational cost required to achieve them. Prior work has shown that training modern machine learning models can result in substantial energy consumption and associated carbon emissions [3, 4]. In recommender systems, this creates a persistent asymmetry: the field rewards incremental improvements in accuracy, while largely ignoring the resources required to obtain them. Recent work has begun to make this trade-off explicit, showing that even modest performance gains can incur non-trivial carbon costs [6, 7]. Empirical studies further suggest that increasing model complexity often leads to diminishing returns, while substantially increasing environmental impact [8]. Yet, these considerations remain largely absent from standard evaluation protocols.

At the same time, the impact of recommender systems cannot be understood purely in terms of model training or inference. Their effects are amplified by scale. Users spend multiple hours per day interacting with personalized content streams, and even small shifts in recommendation policies can translate into large aggregate changes in behavior. What appears as a minor improvement at the model level can therefore have significant downstream consequences. This makes recommender systems impactful not only because of their computational footprint, but because of the behaviors they shape. A growing body of work has begun to examine the relationship between recommender systems and sustainability, typically from two perspectives: systems that promote sustainable behavior, and systems that aim to reduce the environmental and societal impact of the recommendation process itself [9]. While both perspectives are important, they are often addressed in isolation. As a result, the connection between personalization objectives and system-level sustainability outcomes remains only partially understood.

This becomes particularly visible in fashion e-commerce, where return behavior represents both an economic and environmental challenge. Product returns are widespread, with estimates suggesting that more than half of ordered fashion items are returned in some settings [10]. These returns carry substantial environmental costs, including transportation emissions and, in many cases, disposal of unsold goods. Returned items are frequently not resold due to processing costs, leading to additional waste streams and CO_2 emissions [10]. A key driver of this phenomenon is bracketing, where users purchase several types of the same item, keep one, and return the rest. This behavior emerges as a reasonable response to uncertainty in online shopping [5], but it is also shaped by system design choices, such as lenient return policies and interfaces that surface similar or adjacent items. It is therefore not an edge case, but a predictable outcome of how uncertainty is managed in online retail.

Importantly, bracketing is not marginal. It is widely observed in practice and explicitly identified as a common behavior in fashion e-commerce [11]. This points to a broader dynamic: personalization systems do not only influence what users choose, but how they choose under uncertainty. In doing so, they can amplify behaviors that are locally suitable for users but globally inefficient, increasing return volumes, logistical overhead, and environmental impact.

Taken together, these strands of work highlight a gap in how recommender and personalization systems are currently studied and evaluated. Research has examined model efficiency, environmental impact, and return behavior largely in isolation. What is still missing is a perspective that connects these dimensions and treats them as part of the same system.

3. A Taxonomy of Conflicts

The misalignment between personalization and sustainability does not take a single form. It emerges through a set of recurring tensions that cut across model design, evaluation, and deployment. We structure these tensions into four categories that make explicit how personalization objectives translate into system-level sustainability outcomes.

These categories are not independent. They describe different aspects of the same underlying pattern: systems optimized for engagement produce effects that remain largely unaccounted for at the system level.

3.1. Objective Misalignment

Modern personalization systems are built around a simple objective: increase engagement. More clicks, longer sessions, more purchases. This objective is rarely questioned. It aligns with platform incentives and is easy to measure.

From a sustainability perspective, however, it is difficult to justify. Sustainability often implies doing less, not more. Reducing unnecessary consumption, limiting waste, and encouraging efficient behavior are fundamentally at odds with maximizing interaction. A streaming platform that increases watch time also increases energy consumption associated with data transfer and device usage. An e-commerce system that increases conversions may also increase returns, packaging, and transportation.

This becomes particularly visible in domains such as fashion e-commerce. Systems designed to increase conversion can encourage, e.g., bracketing. From the user's perspective, this is a reasonable response to uncertainty. From the system's perspective, it is a success. At the system level, however, it results in increased activity that is inherently inefficient.

This is not a side effect. It follows directly from the objective. As long as engagement remains the dominant metric, these outcomes are expected.

3.2. Efficiency–Impact Trade-offs

A related tension concerns how progress is measured. Improvements in recommender systems are typically reported in terms of accuracy, often at the level of small percentage gains. Achieving these gains frequently requires more complex models, larger datasets, and increased computational resources.

The implicit assumption is that higher accuracy is always desirable. In isolation, this is reasonable. In context, it becomes less convincing. Training and evaluating recommender systems can generate significant carbon emissions [6], and these costs increase rapidly with model complexity.

From a user perspective, the benefit of a 2–3% improvement in recommendation accuracy is often negligible [12]. From an environmental perspective, the cost is cumulative and very real [13]. The result is a familiar asymmetry: we optimize aggressively for improvements that are barely perceptible, while largely ignoring the resources required to achieve them.

This does not make accuracy irrelevant. It does make it incomplete [14, 15]. When improvements are evaluated without considering their cost, the system is only partially optimized.

3.3. System-Level Externalities

The most consequential effects of recommender systems often lie outside the system itself. Recommendations influence behavior, and behavior drives real-world processes such as logistics, manufacturing, and resource consumption [16]. These downstream effects are rarely captured in evaluation metrics, yet they often dominate the overall impact.

Fashion e-commerce provides a clear example. Return rates can exceed 40% [17], with a substantial share driven by bracketing. Each return involves transportation, handling, and often disposal. Even when items are resold, they rarely follow the same path as the original purchase, introducing additional overhead at every step.

Conflict	Optimizes	Ignores
Objective Misalignment	Engagement e.g., clicks, views	Resource use, long-term outcomes
Efficiency–Impact Trade-offs	Accuracy gains	Energy, carbon cost
System-Level Externalities	User actions e.g., purchases	Logistics, waste, emissions
Ecosystem Blindness	Model performance	System behavior, downstream impact

Table 1
Four recurring ways in which personalization objectives conflict with sustainability outcomes at the system level.

At scale, these effects accumulate quickly. Reducing returns by even a small percentage can lead to substantial carbon savings [18, 19]. Conversely, small increases can have disproportionately large consequences. What appears as a minor shift in recommendation output can trigger a cascade of physical processes.

Recommendation outputs are therefore not just informational. They are *operational*. They initiate actions that propagate through supply chains and infrastructure, with consequences that extend far beyond the system itself.

3.4. Ecosystem Blindness

A final tension concerns how recommender systems are evaluated. They are typically treated as self-contained algorithms, rather than as components of larger socio-technical systems [20, 21]. This leads to a form of ecosystem blindness.

In practice, recommender systems are embedded in complex environments that include data pipelines, cloud infrastructure, business processes, and user behavior. Their effects propagate through these systems in ways that are difficult to capture with traditional evaluation methods. Optimizing a recommender system in isolation can therefore lead to suboptimal, or even harmful, outcomes at the system level.

There is an important distinction here. System-level externalities describe what happens as a result of recommendations. Ecosystem blindness describes our failure to account for those effects. The issue is not only that these impacts exist, but that they are systematically excluded from evaluation. Addressing this requires a shift in perspective. What is typically measured is model performance; what ultimately matters is system behavior.

Combined, the categories summarized in Table 1 describe a consistent pattern rather than a set of isolated issues. Objective misalignment captures what systems are designed to optimize. Efficiency–impact trade-offs describe how improvements are pursued. System-level externalities make visible how these choices propagate beyond the system, while ecosystem blindness explains why these effects remain largely unaccounted for.

The result is not a single problem, but a way of building and assessing recommender systems. The misalignment with sustainability is therefore not incidental, but a consequence of how these systems are currently conceived, optimized, and evaluated in practice.

4. Where These Conflicts Appear

The tensions described above are not abstract. They become visible in concrete settings where recommender systems are deployed, often in ways that are easy to miss when focusing on model-level performance.

One recurring pattern is the pursuit of marginal improvements at disproportionate cost. Recent analyses show that training recommender systems, particularly deep learning models, can have a substantial environmental footprint [6, 7]. At the same time, the performance gains are often modest. This raises a straightforward question: do we need these improvements? Under standard evaluation metrics, the answer is typically yes. From a system-level perspective, the answer is less clear [22, 13]. What counts as progress at the model level may, in practice, be a poor trade-off once resource use is taken into account.

A more direct example can be found in fashion e-commerce, where the interaction between recommendation, user behavior, and logistics becomes visible. High return rates are not merely a side effect of online retail, but are closely tied to how products are presented and recommended. Behaviors such as ordering multiple variants of the same item with the intention of returning most of them emerge as a rational response to uncertainty, but are also reinforced by system design. From the perspective of the recommender system, these are successful interactions. At the system level, they translate into transportation, handling, repackaging, and, in many cases, disposal. The gap between these perspectives is where the environmental impact arises.

Similar patterns appear in media consumption. Personalization systems are designed to maximize engagement, often by encouraging longer sessions and continuous interaction. While this increases user retention, it also increases the energy required to deliver and consume content. Given that users already spend several hours per day interacting with digital media, further increases offer diminishing returns in terms of user value, while continuing to increase resource use. In this setting, engagement is treated as a proxy for value, even when that relationship becomes increasingly weak.

5. Rethinking Personalization Systems

The tensions outlined above are structural. Addressing them requires rethinking how personalization systems are built, evaluated, and understood.

Model choice is typically framed as a question of performance. The default assumption is that more complex models are preferable, as long as they improve accuracy. The taxonomy in Table 1 makes this assumption difficult to sustain. Once efficiency–impact trade-offs are taken seriously, model selection becomes a question of balance rather than maximization. Energy consumption and carbon cost are no longer external concerns, but part of the decision itself. In many cases, simpler models provide a more reasonable trade-off, particularly when performance differences are small. Complexity should be used where it adds value, and its cost should be visible when it does not [4].

The same shift applies to evaluation. Current frameworks focus almost exclusively on short-term metrics such as accuracy and engagement [20]. This reinforces objective misalignment and ecosystem blindness. Expanding evaluation to include environmental and system-level effects would change what counts as a good system. This is challenging. Long-term outcomes are difficult to measure, and system-level effects are often indirect. Without such measures, however, systems will continue to optimize for what is easy to observe rather than what actually matters.

Personalization systems already function as behavioral interventions. They do not simply reflect preferences; they shape how decisions are made. This becomes visible in settings such as e-commerce, where recommendation can encourage users to consider multiple similar options, reinforcing bracketing and similar behaviors. From the perspective of the system, this increases interaction and conversion. At the system level, it increases returns, handling, and waste.

Once this role is taken seriously, different design choices become possible. This may involve reducing the number of options presented, prioritizing alternatives with lower downstream impact, or, in some cases, not recommending additional items at all. These are not standard optimization targets, but they follow directly from a system-level perspective.

This does not point to a single solution. It changes how the problem is framed. The focus shifts away from optimizing isolated metrics and toward understanding how personalization interacts with the systems around it.

From this perspective, recommender systems are components of larger systems with real-world consequences. Improving accuracy remains important, but it is only one dimension of system performance. Efficiency, long-term impact, and system-level effects need to be considered alongside it.

This also raises broader questions about responsibility. If personalization systems shape behavior at scale, their design choices have consequences beyond individual interactions. Addressing these questions likely requires approaches that extend beyond traditional optimization, including perspectives from economics, sustainability, and policy.

6. Conclusion

This paper has argued that the tension between personalization and sustainability is not incidental, but follows from how these systems are conceived, optimized, and evaluated. By structuring this tension into a set of recurring conflicts and grounding it in concrete examples, we have aimed to make this pattern more explicit.

The implication is not that personalization systems are inherently unsustainable. It is that current objectives and evaluation practices systematically favor outcomes that are difficult to reconcile with sustainability goals. As long as engagement remains the dominant metric, and system-level effects remain outside the scope of evaluation, these tensions are likely to persist.

Addressing this does not require a single new method. It requires a shift in perspective. Instead of optimizing isolated metrics, personalization systems need to be understood as components of larger systems with real-world consequences.

Future work should focus on making these system-level interactions visible, both in how systems are evaluated and how they are designed. This includes developing methods that account for environmental and downstream effects, as well as exploring alternative objectives that better reflect long-term and system-level considerations.

Declaration on Generative AI

During the preparation of this work, the author(s) used ChatGPT and Grammarly for Grammar and spelling check and rephrasing. After using these tool(s)/service(s), the author(s) reviewed and edited the content as needed and take(s) full responsibility for the publication's content.

References

[1] D. Jannach, G. Adomavicius, Recommendations with a Purpose, in: Proceedings of the 10th ACM Conference on Recommender Systems, RecSys '16, Association for Computing Machinery, New York, NY, USA, 2016, pp. 7–10. URL: https://dl.acm.org/doi/10.1145/2959100.2959186. doi:10.1145/2959100.2959186.

[2] European Environment Agency, Europe's environment and climate: knowledge for resilience, prosperity and sustainability, 2025. URL: https://www.eea.europa.eu. doi:10.2800/3817344, europe's Environment 2025 – Main report.

[3] E. Strubell, A. Ganesh, A. McCallum, Energy and Policy Considerations for Modern Deep Learning Research, Proceedings of the AAAI Conference on Artificial Intelligence 34 (2020) 13693–13696. URL: https://ojs.aaai.org/index.php/AAAI/article/view/7123. doi:10.1609/aaai.v34i09.7123, number: 09.

[4] P. Henderson, J. Hu, J. Romoff, E. Brunskill, D. Jurafsky, J. Pineau, Towards the Systematic Reporting of the Energy and Carbon Footprints of Machine Learning, Journal of Machine Learning Research 21 (2020) 1–43. URL: http://jmlr.org/papers/v21/20-312.html.

[5] Y. Xu, G. Hua, T. C. E. Cheng, T.-M. Choi, Y. Li, S. Liu, Retailing and ordering strategies for online apparel retailers facing bracketing purchase behaviour, International Journal of Production

Research 61 (2023) 2841–2853. URL: https://doi.org/10.1080/00207543.2022.2070045. doi:10.1080/00207543.2022.2070045, _eprint: https://doi.org/10.1080/00207543.2022.2070045.

[6] T. Vente, L. Wegmeth, A. Said, J. Beel, From Clicks to Carbon: The Environmental Toll of Recommender Systems, in: Proceedings of the 18th ACM Conference on Recommender Systems, RecSys '24, Association for Computing Machinery, New York, NY, USA, 2024, pp. 580–590. URL: https://dl.acm.org/doi/10.1145/3640457.3688074. doi:10.1145/3640457.3688074.

[7] L. Wegmeth, T. Vente, A. Said, J. Beel, Green Recommender Systems: Understanding and Minimizing the Carbon Footprint of AI-Powered Personalization, ACM Trans. Recomm. Syst. (2025). URL: https://dl.acm.org/doi/10.1145/3768626. doi:10.1145/3768626, just Accepted.

[8] G. Spillo, A. De Filippo, C. Musto, M. Milano, G. Semeraro, Towards Sustainability-aware Recommender Systems: Analyzing the Trade-off Between Algorithms Performance and Carbon Footprint, in: Proceedings of the 17th ACM Conference on Recommender Systems, RecSys '23, Association for Computing Machinery, New York, NY, USA, 2023, pp. 856–862. URL: https://doi.org/10.1145/3604915.3608840. doi:10.1145/3604915.3608840.

[9] A. De Filippo, G. Spillo, L. Boratto, M. Milano, C. Musto, G. Semeraro, Recommender systems and sustainability: a dual perspective, Computer Science Review 60 (2026) 100912. URL: https://www.sciencedirect.com/science/article/pii/S1574013726000213. doi:10.1016/j.cosrev.2026.100912.

[10] M. Niederlaender, A. Lodi, S. Gry, R. Biswas, D. Werth, Garment Returns Prediction for AI-Based Processing and Waste Reduction in E-Commerce:, in: Proceedings of the 16th International Conference on Agents and Artificial Intelligence, SCITEPRESS - Science and Technology Publications, Rome, Italy, 2024, pp. 156–164. URL: https://www.scitepress.org/DigitalLibrary/Link.aspx?doi=10.5220/0012321300003636. doi:10.5220/0012321300003636.

[11] M. Niederlaender, S. Driouech, S. Gry, A. N. Lodi, R. Biswas, D. Werth, Towards Waste Reduction in E-Commerce: A Comparative Analysis of Machine Learning Algorithms and Optimisation Techniques for Garment Returns Prediction with Feature Importance Evaluation, SN Computer Science 6 (2025) 436. URL: https://doi.org/10.1007/s42979-025-03944-z. doi:10.1007/s42979-025-03944-z.

[12] E. Zangerle, C. Bauer, Evaluating Recommender Systems: Survey and Framework, ACM Comput. Surv. 55 (2022) 170:1–170:38. URL: https://dl.acm.org/doi/10.1145/3556536. doi:10.1145/3556536.

[13] D. Jannach, A. Said, M. Tkalcic, M. Zanker, Recommender Systems for Good (RS4Good): Survey of Use Cases and a Call to Action for Research that Matters, ACM Trans. Recomm. Syst. (2025). URL: https://dl.acm.org/doi/10.1145/3746648. doi:10.1145/3746648, just Accepted.

[14] M. D. Ekstrand, M. C. Willemsen, Behaviorism is not enough: Better recommendations through listening to users, RecSys 2016 - Proceedings of the 10th ACM Conference on Recommender Systems (2016) 221–224. doi:10.1145/2959100.2959179.

[15] A. Said, M. S. Pera, M. D. Ekstrand, We're Still Doing It (All) Wrong: Recommender Systems, Fifteen Years Later, 2025. URL: https://arxiv.org/abs/2509.09414. doi:10.48550/ARXIV.2509.09414, version Number: 1.

[16] B. Pathak, R. Garfinkel, R. D. Gopal, R. Venkatesan, F. Yin, Empirical Analysis of the Impact of Recommender Systems on Sales, Journal of Management Information Systems 27 (2010) 159–188. URL: https://doi.org/10.2753/MIS0742-1222270205. doi:10.2753/MIS0742-1222270205, _eprint: https://doi.org/10.2753/MIS0742-1222270205.

[17] M. Yocabè, Guida ai resi nel mondo dell'e-commerce, 2023. URL: https://www.yocabe.com/resi_nel_mondo_e-commerce/.

[18] R. Velazquez, S. M. Chankov, Environmental Impact of Last Mile Deliveries and Returns in Fashion E-Commerce: A Cross-Case Analysis of Six Retailers, in: 2019 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM), 2019, pp. 1099–1103. URL: https://ieeexplore.ieee.org/abstract/document/8978705. doi:10.1109/IEEM44572.2019.8978705, iSSN: 2157-362X.

[19] R. F. Bertram, T. Chi, A study of companies' business responses to fashion e-commerce's environmental impact, International Journal of Fashion Design, Technology and Education 11 (2018) 254–264. URL: https://doi.org/10.1080/17543266.2017.1406541. doi:10.1080/17543266.2017.1406541, _eprint: https://doi.org/10.1080/17543266.2017.1406541.

[20] C. Bauer, E. Zangerle, A. Said, Exploring the Landscape of Recommender Systems Evaluation: Practices and Perspectives, ACM Trans. Recomm. Syst. 2 (2024) 11:1–11:31. URL: https://dl.acm.org/doi/10.1145/3629170. doi:10.1145/3629170.

[21] T. Silveira, M. Zhang, X. Lin, Y. Liu, S. Ma, How good your recommender system is? A survey on evaluations in recommendation, International Journal of Machine Learning and Cybernetics 10 (2019) 813–831. URL: https://doi.org/10.1007/s13042-017-0762-9. doi:10.1007/s13042-017-0762-9.

[22] A. Said, Recommender Systems for Social Good:The Role of Accountability and Sustainability, volume Proceeding of the 2024 workshop on Recommender Systems for Social Good of *RecSoGood*, Springer, 2024.