# Sustainability Evaluation Metrics for Configuration Systems

Alexander Felfernig†, Damian Garber†, Sebastian Lubos† and Thi Ngoc Trang Tran†

*Institute of Software Engineering and AI, Graz University of Technology, Graz, Austria*

**Abstract**

Sustainability-oriented evaluation metrics offer a means to assess the quality of configuration systems beyond conventional metrics such as accuracy of personalized configurations or sales-related conversion rates. Aligned with the United Nations' Sustainable Development Goals (SDGs), these metrics enable a structured analysis of the environmental, social, and economic impacts of configuration systems. In this paper, we explore sustainability-focused evaluation metrics tailored to configurators and examine applications and implications.

**Keywords**

Configuration, Configuration Systems, Sustainability, Evaluation Metrics, Sustainable Development Goals



## 1. Introduction

Configuration can be regarded as a key technology supporting the mass customization paradigm [1, 2, 3]. Traditionally, the effectiveness of configurators is assessed on the basis of performance-related metrics such as configuration accuracy (of personalized recommendations) and sales-related conversion rates [4, 5]. However, due to societal relevance, there is a growing need to integrate evaluation criteria that also take into account sustainability aspects [6, 7, 8, 9, 10, 11, 12].

To address this need, we propose a basic set of sustainability-aware evaluation metrics. These metrics go beyond immediate system performance and focus more on embedding long-term impacts of configurations into the evaluation process. In this context, our aim is not only to optimize the utility of configurations but also to take into account global sustainability goals such as the United Nations' Sustainable Development Goals (SDGs).[1] By considering, for example, the environmental impact of selected components, fairness and inclusivity of configuration options, and economic fairness, such metrics offer a more holistic understanding of the impact of configuration systems.

In this paper, we provide an overview of basic sustainability-oriented evaluation metrics (being aware of the fact that many further variants are possible). We map these metrics to the three core sustainability areas of the United Nations' SDGs: environmental, social, and economic. We also illustrate application scenarios and discuss topics for future research.

The remainder of this paper is structured as follows: in Section 2, we present evaluation metrics related to aspects of environmental sustainability. Section 3 addresses metrics for social sustainability, followed by economic sustainability metrics which are discussed in Section 4. Sections 5 and 6 discuss cross-dimensional metrics and outline open research issues. The paper is concluded with Section 8.

## 2. Environmental Metrics

Environmental sustainability metrics extend the evaluation of configuration systems beyond traditional performance-based measures by assessing their contribution to environmental objectives.

*ConfWS'25: 27th International Workshop on Configuration, Oct 25–26, 2025, Bologna, Italy*

† These authors contributed equally.

alexander.felfernig@tugraz.at (A. Felfernig); damian.garber@tugraz.at (D. Garber); sebastian.lubos@tugraz.at (S. Lubos); trang.tran@tugraz.at (T. N. T. Tran)

0000-0003-0108-3146 (A. Felfernig); 0009-0005-0993-0911 (D. Garber); 0000-0002-5024-3786 (S. Lubos); 0000-0002-3550-8352 (T. N. T. Tran)



[1] https://sdgs.un.org/goals

## 2.1. Carbon Footprint of Configurations

The carbon footprint of a set of configurations $\mathcal{CONFS}$ proposed to users over a specific time period measures the average greenhouse gas emissions (expressed, e.g., in tons of $CO_2$ (equivalent) produced over the full configuration lifecycle) associated with the configurations $conf \in \mathcal{CONFS}$. Let CarF($conf$) denote the estimated overall carbon footprint of components in $conf$. Then, the *average carbon footprint of offered configurations* (AvgCarFConf) can be defined as follows:

$$\text{AvgCarFConf} = \frac{1}{|\mathcal{CONFS}|} \sum_{\text{conf} \in \mathcal{CONFS}} \text{CarF}(\text{conf}) \tag{1}$$

In this context, lower values of AvgCarFConf indicate that the configuration system tends to favor components with lower carbon footprints.

## 2.2. Energy Consumption of Configuration

Energy consumption of *configuration* (i.e., the generation of configurations) refers to the energy required to compute configurations ($conf \in \mathcal{CONFS}$) for users of a configuration system. Let $E_{\text{configuration}}$ denote the total energy (e.g., in kilowatt-hours) consumed by the configuration system over a defined evaluation period, and let $N_{\text{conf}}$ be the total number of configurations generated. The *Energy Consumption per Configuration* (ECConf) can be defined as follows:

$$\text{ECConf} = \frac{E_{\text{Configuration}}}{N_{\text{conf}}} \tag{2}$$

This metric is particularly relevant for large-scale configuration systems that involve complex reasoning and consistency management, where an optimization for energy efficiency is important.

## 2.3. Energy Consumption of Model Building

Energy consumption of *configuration model building* refers to the total energy consumed during the construction of a configuration knowledge base. Let $EC_{\text{dev}}$ represent the cumulative energy consumed throughout the entire configuration model development process, and let $N_{\text{versions}}$ denote the number of developed configuration model versions over a specific time period. The *Energy Consumption per Model Version* (ECModVer) can be defined as follows:

$$\text{ECModVersion} = \frac{EC_{\text{dev}}}{N_{\text{versions}}} \tag{3}$$

Such metrics are particularly relevant for evaluating the environmental and computational efficiency of different configuration environments.

## 2.4. Energy Savings Through Configuration

*Energy Savings through Configuration* (ESTConf) refers to the reduction in energy consumption or resource usage achieved as a result of applying a configuration system [13]. Configuration systems can enhance efficiency by guiding users toward environmentally friendly component selections, reducing over-dimensioning, or optimizing system designs. Let $EC_{\text{baseline}}$ denote, for example, the energy consumption of a system (e.g., annual energy usage) designed without the support of a configuration system, and let $EC_{\text{withconf}}$ represent the energy consumption observed when a configuration system has been used (e.g., a configuration tool supporting energy-efficient component selection). Related energy savings can be expressed as follows:

$$\text{ESTConf} = \frac{EC_{\text{baseline}} - EC_{\text{withconf}}}{EC_{\text{baseline}}} \tag{4}$$

Such metrics are particularly relevant in domains where configuration decisions can significantly influence consumption patterns [10, 13].

Environmental sustainability metrics reflect a paradigm shift from short-term optimization goals—such as maximizing user engagement—toward long-term ecological considerations. However, implementing these metrics in practice presents challenges such as limited access to reliable carbon footprint data and the lack of standardized definitions for the sustainability of components.

## 3. Social Metrics

Social sustainability in configuration systems emphasizes the fair and inclusive generation of configurations. These metrics go beyond traditional performance measures to ensure that the system's design and resulting configurations support social equity, accessibility, and community well-being. In this context, configuration processes should avoid bias, promote inclusive component ($comp \in conf$) selection, and ensure that all users can effectively engage with and benefit from the configuration system.

### 3.1. Fairness and Bias

Related metrics can be used to assess whether configuration outcomes (components) are equitably distributed across different demographic groups. A commonly discussed fairness criterion is *demographic parity*, which requires that the share of selected components ($comp \in conf$) in generated configurations is similar across sensitive attributes (e.g., gender, age group). Let $G$ denote a set of demographic groups, and let $P_g(comp)$ represent the probability that component $comp$ is included in a configuration for users belonging to group $g \in G$. In this context, demographic parity is satisfied if the following condition holds:

$$P_g(\text{comp}) \approx P_{g'}(\text{comp}) \quad \forall\, g, g' \in G \quad (g \neq g') \tag{5}$$

### 3.2. Diversity

To promote exposure to diverse perspectives, diversity in a configuration list ($\mathcal{C}_u$) presented to a user $u$ is crucial. *Configuration diversity* $\text{ConfDiv}_u$ for a user $u$ can be measured, for example, based on the average pairwise similarity (sim) among configurations ($\{conf_i, conf_j\} \subseteq C_u$, $i \neq j$) presented to user $u$ where similarity values (between two configurations) are assumed to be in the interval $(0, 1)$:

$$\text{ConfDiv}_u = 1 - \frac{\sum_{\text{conf}_i \in \mathcal{C}_u} \sum_{\text{conf}_j \in \mathcal{C}_u} \text{sim}(\text{conf}_i, \text{conf}_j)}{|\mathcal{C}_u| \times (|\mathcal{C}_u| - 1)} \tag{6}$$

In this context, $\text{sim}(\text{conf}_i, \text{conf}_j)$ denotes the similarity between configurations $\text{conf}_i$ and $\text{conf}_j$ (e.g., based on component equality). Higher values of $\text{ConfDiv}_u$ indicate greater diversity in $C_u$ which reflects a lower average similarity among configurations. The metric *ConfDiv* would then represent the average calculated over all user-specific values ($\text{ConfDiv}_u$).

### 3.3. Accessibility and Inclusivity

Accessibility aims to ensure that both, configurators and configurations can be effectively used by users with diverse abilities and backgrounds, represented by different groups $g \in G$. Let $\mathcal{AC}$ denote a set of *accessibility criteria* (e.g., understandability of configuration steps, clarity of component descriptions, or usability for users with visual or cognitive impairments), and let $sat$ be a function that measures the extent to which the components and/or user interface elements in a set $\mathcal{Q}$ satisfy these criteria for a group $g$ on a scale from 0 to 1. Then, the *accessibility score* ($ACC_g$) for a group $g \in G$ can be defined as follows:

$$\text{ACC}_g = \frac{\sum_{q \in \mathcal{Q}} sat(q, \mathcal{AC}, g)}{|\mathcal{Q}|} \tag{7}$$

A higher value of $\text{ACC}_g$ indicates better accessibility for group $g$. *Inclusivity* is considered to be achieved when accessibility is fulfilled equitably across different demographic or ability-based groups $g_i \in G$:

$$\text{ACC}(g_i) \approx \text{ACC}(g_j) \quad \forall\, g_i, g_j \in G \quad (i \neq j) \tag{8}$$

### 3.4. Health Improvement through Configuration

*Health Improvement through Configuration* (HIConf) refers to the enhancement of individual or population health outcomes achieved through the use of configuration systems. Configuration systems can support healthier decisions by guiding users toward appropriate selections of components related, for example, to diet plans and training plans.

Let $\mathcal{M}_{\text{with}}$ denote a health outcome metric (e.g., average activity level or body mass index) for users of a configuration system, and let $\mathcal{M}_{\text{without}}$ represent the same metric for users not using such a system. The health improvement across all users can be defined as:

$$\text{HIConf} = \frac{\mathcal{M}_{\text{with}} - \mathcal{M}_{\text{without}}}{\mathcal{M}_{\text{without}}} \tag{9}$$

This metric is particularly relevant in application domains such as digital health platforms, wellness configurators, and preventive healthcare services, where personalized configurations can positively impact well-being [14, 15].

## 4. Economic Metrics

Economic sustainability metrics assess the role of configuration systems in fostering inclusive, resilient, and locally grounded economic ecosystems. These metrics extend traditional evaluation dimensions by considering how configuration systems influence market fairness and the visibility of small and/or local component suppliers. Such metrics help evaluate whether configurations promote equitable access to market opportunities and support economically sustainable choices.

### 4.1. Support for Local Businesses

This metric quantifies the proportion of components in the configuration knowledge base that originate from small or local businesses. Let $\mathcal{C}_u \subseteq \mathcal{COMPS}$ denote the components from local or small-scale providers available to user $u$. The *Local Business Promotion Rate* (LBPR) can be defined as:

$$\text{LBPR} = \frac{\sum_{u \in \mathcal{U}} |\{\text{comp} \in \mathcal{COMPS} : \text{comp} \in \mathcal{C}_u\}|}{|\mathcal{U}|} \tag{10}$$

Higher LBPR values indicate that the configurator supports community-level economic development by promoting components supplied by small and/or local businesses.

### 4.2. Fairness in Exposure

Configuration systems can inadvertently concentrate exposure and revenue on a smaller subset of producers supplying specific types of components (i.e., we regard fairness in exposure as a context-dependent metric). The reasons behind could be, for example, specific variable (value) orderings specified in the underlying constraint solver. To foster economic fairness, we define fairness in the context of *component producer exposure* as *Component Producer Exposure Fairness* (CPEF):

$$\text{CPEF} = \frac{avgdist_2(\mathcal{P})}{maxdist_2(\mathcal{P})} \tag{11}$$

In this context, $\mathcal{P}$ denotes producers associated with components appearing in user configurations. The term $avgdist_2(\mathcal{P})$ represents the average pairwise distance in exposure counts between two producers, while $maxdist_2(\mathcal{P})$ is the maximum observed distance between any two producers in $\mathcal{P}$. Both are calculated based on how often a producer's components are presented to users across all configurations.

The discussed economic sustainability metrics can offer valuable insights into how configuration systems influence the distribution of economic value. These metrics help to evaluate whether a system promotes equitable market exposure and supports small or local businesses.

## 5. Cross-cutting Metrics

Cross-cutting sustainability metrics capture and assess the multifaceted effects of configurators spanning environmental, social, and economic aspects.

### 5.1. Sustainable User Behavior

This metric evaluates sustainability-related user interaction behavior. Let $B_u$ represent the set (more precisely, the bag) of user behaviors over a specific time period (of user $u$) when interacting with the configurator (e.g., inspecting component details, reading explanations, or selecting a component). Furthermore, let $\mathcal{S}$ be a set of sustainable behaviors (e.g., selecting eco-friendly components). The *Sustainable Configuration Behavior Score* (SCBS) can be defined as:

$$\text{SCBS} = \frac{\sum_{u \in \mathcal{U}} |\{b \in B_u : b \in \mathcal{S}\}|}{\sum_{u \in \mathcal{U}} |B_u|} \tag{12}$$

Higher SCBS values indicate a higher degree of sustainability-related user interaction behaviors.

### 5.2. Interpretability of Configurations

Interpretability (*IntCS*) of configuration support is essential for enabling informed user decision-making. Let $\mathcal{E}_u$ denote the set of explanations $e$ provided to user $u$ over a specific time period (e.g., justifications for selecting a particular component *comp*), and let $\text{interpret}(e)$ quantify the interpretability of explanation $e$. The *Average Explanation Interpretability* (IntCS) across all users can be defined as:

$$\text{IntCS} = \frac{1}{|\mathcal{U}|} \sum_{u \in \mathcal{U}} \frac{1}{|\mathcal{E}_u|} \sum_{e \in \mathcal{E}_u} \text{interpret}(e) \tag{13}$$

Interpretability may be estimated on the basis of explicit user feedback, information complexity scores, or automated assessments (e.g., using large language models).

### 5.3. Life Cycle Impact of Configurations

Life cycle impact analysis considers both, upstream and downstream effects in the production, distribution, usage, and disposal of configurations. Let LCIC(*comp*) denote the total estimated life cycle impact score for component *comp* (including aspects such as carbon footprint or the potential for reuse and recycling). The *Average Life Cycle Impact of Configurations* (AvgLCIC) can be defined as:

$$\text{AvgLCIC} = \frac{\sum_{\text{conf} \in \mathcal{CONFS}} \sum_{\text{comp} \in \text{conf}} \text{LCIC(comp)}}{\sum_{\text{conf} \in \mathcal{CONFS}} |\text{conf}|} \tag{14}$$

Lower values of AvgLCIC indicate that the configuration system favors components with lower ecological and social burdens throughout their life cycles.

The deployment of such cross-cutting sustainability metrics also depends on the availability and reliability of life cycle metadata for the involved components.

## 6. Challenges and Research Directions

Despite growing interest in sustainability-aware configuration [6, 10, 11, 12], several challenges hinder a widespread adoption and evaluation.

### 6.1. Multi-objective Optimization

The incorporation of sustainability goals into configuration systems often introduces trade-offs between traditional performance metrics (e.g., e.g., accuracy with regard to recommended/included components of a configuration) and sustainability-related outcomes. Formally, this leads to the optimization of a vector-valued objective function:

$$\max_{\theta} \quad \mathbf{FOPT}(\theta) = [\text{Accuracy}(\theta), \text{Sustainability}(\theta)] \tag{15}$$

where $\theta$ denotes the configuration parameters. This requires the definition of specific multi-objective optimization problems, typically resulting in Pareto-efficient solutions that aim to balance competing goals.

### 6.2. Data Availability and Labeling

Most sustainability metrics rely on fine-grained metadata, such as the carbon footprint of a component, ethical sourcing labels, or the classification of vendors (e.g., local or small-scale). Let $\mathcal{COMPS}$ be the set of components available in the configuration catalog, and let $s_{\text{comp}}$ be a binary sustainability label for a component comp $\in \mathcal{COMPS}$. The share of labeled components is defined as:

$$\text{CLabelCov} = \frac{|\{\text{comp} \in \mathcal{COMPS} : s_{\text{comp}} \text{ is known}\}|}{|\mathcal{COMPS}|} \tag{16}$$

Low CLabelCov values limit the applicability and accuracy of sustainability-related evaluations.

## 7. Productive Usage of Metrics

Developers have to integrate sustainability indicators such as carbon footprint or component origin into evaluation workflows of the configuration environment. This includes activities such as extending existing logging frameworks with the goal to capture relevant data such as energy usage, component source information, and demographic data of users. In addition, configuration algorithms can be enhanced to prioritize sustainable choices, for example, by applying "green component variable value ordering" that favor environmentally friendly or ethically produced components. Beyond implementation, companies have the opportunity to promote transparency by reporting the sustainability performance of their configuration engines.

## 8. Conclusions

Sustainability-oriented evaluation metrics are essential for advancing configuration systems beyond conventional performance criteria. By embedding environmental, social, and economic considerations into system assessment, these metrics help to align the development of configuration systems with global sustainability goals, particularly those outlined in the United Nations Sustainable Development Goals (SDGs). Configuration systems have the potential to promote eco-friendly component selection, ensure equitable access to configurable solutions, and support local and responsible value chains. With this, these systems can contribute to more sustainable forms of product customization. A central focus of our future work will be to provide software components that will support the application of our proposed metrics in real-world contexts.

## Declaration on Generative AI

While preparing this work, the author(s) used ChatGPT-4 (GPT-4-turbo) and Grammarly to check grammar and spelling and improve formulations. 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] A. Felfernig, L. Hotz, C. Bagley, J. Tiihonen, Knowledge-based Configuration: From Research to Business Cases, 1 ed., Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 2014.

[2] D. Sabin, R. Weigel, Product configuration frameworks-a survey, IEEE Int. Sys. and their Applications 13 (1998) 42–49. doi:10.1109/5254.708432.

[3] M. Stumptner, An overview of knowledge-based configuration, AI Commun. 10 (1997) 111–125.

[4] A. Falkner, A. Felfernig, A. Haag, Recommendation technologies for configurable products, AI Mag. 32 (2011) 99–108. doi:10.1609/aimag.v32i3.2369.

[5] S. Lubos, V.-M. Le, A. Felfernig, T. N. T. Tran, Analysis operations for constraint-based recommender systems, in: RecSys'23, RecSys '23, ACM, New York, NY, USA, 2023, pp. 709–714. doi:10.1145/3604915.3608819.

[6] C. Fagarasan, C. Cristea, M. Cristea, O. Popa, A. Pisla, Integrating sustainability metrics into project and portfolio performance assessment in agile software development: A data-driven scoring model, Sustainability 15 (2023) 13139. doi:10.3390/su151713139.

[7] A. Felfernig, M. Wundara, T. N. T. Tran, S. Polat-Erdeniz, S. Lubos, M. El Mansi, D. Garber, V.-M. Le, Recommender systems for sustainability: overview and research issues, Frontiers in Big Data 63 (2023). doi:10.3389/fdata.2023.1284511.

[8] A. Konys, Knowledge based approach to sustainability assessment, Procedia Computer Science 176 (2020) 3626–3635. doi:10.1016/j.procs.2020.09.023, 24th Int. Conf. on Knowledge-Based and Intelligent Information & Engineering Systems.

[9] H. Koziolek, Sustainability evaluation of software architectures: a systematic review, QoSA-ISARCS '11, ACM, New York, NY, USA, 2011, pp. 3–12. doi:10.1145/2000259.2000263.

[10] S. Lubos, A. Felfernig, L. Hotz, T. Tran, S. Polat-Erdeniz, V.-M. Le, D. Garber, M. E. Mansi, Responsible configuration using llm-based sustainability-aware explanations, in: ConfWS2024, CEUR, 2024, pp. 68–73.

[11] K. Medini, T. Wuest, D. Romero, V. Laforest, Integrating sustainability considerations into product variety and portfolio management, Procedia CIRP 93 (2020) 605–609. doi:10.1016/j.procir.2020.04.147, 53rd CIRP Conf. on Manufacturing Sys.

[12] R. Wiezorek, N. Christensen, Integrating Sustainability Information in Configurators, in: M. Aldanondo, A. Falkner, A. Felfernig, M. Stettinger (Eds.), ConfWS'21, volume 2945, CEUR, 2021, pp. 65–72.

[13] G. Leitner, M. Stettinger, Configuration of domotic systems based on constraint solving, in: 26th ACM International Systems and Software Product Line Conference, SPLC '22, ACM, New York, NY, USA, 2022, pp. 145–149. doi:10.1145/3503229.3547045.

[14] A. Felfernig, M. Wundara, T. N. T. Tran, V.-M. Le, S. Lubos, S. Polat-Erdeniz, Sports recommender systems: overview and research directions, J. Intell. Inf. Syst. 62 (2024) 1125–1164. doi:10.1007/s10844-024-00857-w.

[15] T. N. Trang Tran, M. Atas, A. Felfernig, M. Stettinger, An overview of recommender systems in the healthy food domain, J. Intell. Inf. Syst. 50 (2018) 501–526. doi:10.1007/s10844-017-0469-0.