

Article

From Stakeholder Feedback to Product–Service System Design: A Sentiment-Based Decision-Support Framework Validated in Industrial Cases

Rui Neves-Silva ^{1,*}  and Paulo Pina ²

¹ School of Science and Technology and Uninova CTS, NOVA University of Lisbon, Campus de Caparica, 2829-516 Caparica, Portugal

² Uninova CTS, 2829-516 Caparica, Portugal

* Correspondence: rns@fct.unl.pt

Highlights

Please indicate how your work links to systems science via your contributions to systems practice, theory, and or methodology.

- This study proposes a systems-oriented feedback-to-design framework linking stakeholder feedback, sentiment indicators, reach/influence, and PSS design decisions.
- It integrates sentiment analysis, lifecycle-oriented PSS design, and decision support into a structured framework for systems practice.

What are the main findings and/or the implications of the main findings?

- The framework was assessed through polarity checking, simulation testing, industrial cases, and prototype-level software evidence.
- Results show practical value under TRL-6 conditions, while prediction accuracy and large-scale benchmarking remain areas for future research.

Abstract

Product–Service Systems (PSS) require feedback loops that connect stakeholder experience with design decisions, but distributed feedback is often difficult to transform into actionable design knowledge. This paper synthesises previously distributed results from the DIVERSITY project into an integrated decision-support framework for feedback-driven PSS design. The framework combines social-sentiment representation, reach/influence weighting, opinion modelling, monitoring, extraction and prediction modules, and a hierarchical similarity mechanism for exploring likely sentiment toward new PSS configurations. It was implemented in the DIVERSITY engineering environment and assessed through a preliminary polarity plausibility check, controlled testing with synthetic data generated by OpinionSim, and three industrial business cases. The results indicate that the framework can support segmentation-aware interpretation of stakeholder feedback, shorten the path from market signals to design action, and provide exploratory evidence for early PSS concept generation and redesign. The predictive component is presented as an exploratory design-support mechanism, with forecasting accuracy remaining a topic for future empirical validation. The paper contributes an integrated feedback-to-design framework for PSS and clarifies the conditions under which sentiment-based indicators are most useful in B2C and B2B settings.



Academic Editor: Ying Fan

Received: 11 May 2026

Revised: 8 July 2026

Accepted: 27 July 2026

Published: 1 August 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article

distributed under the terms and

conditions of the [Creative Commons](https://creativecommons.org/licenses/by/4.0/)

[Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

Keywords: product–service systems; sentiment analysis; stakeholder feedback; design decision support; social media; prediction; lifecycle engineering

1. Introduction

Product–Service Systems (PSS) have been promoted as a way of moving from isolated product delivery toward integrated value propositions that combine products, services, supporting networks, and lifecycle relationships [1,2]. This shift is attractive because it can strengthen differentiation, extend lifecycle value, and support more sustainable offerings. At the same time, it increases design complexity. Design teams must reason about the physical artefact, the service layer, their interaction during use, and the implications of design changes for multiple stakeholders over time.

In that setting, stakeholder feedback becomes a strategic resource. In conventional product development, feedback often arrives late, is poorly structured, and passes through several organisational layers before it reaches the people making design choices. In PSS contexts the problem is broader, because feedback may concern the product, the service, their integration, or the way the offering behaves along the lifecycle [3,4]. What designers need, therefore, is more than a way to collect comments. They need a way to turn distributed and uneven feedback into evidence that helps them judge where intervention is needed, what kind of intervention is reasonable, and which alternatives deserve further development.

During the last decade, sentiment analysis has matured as a field for extracting opinions from text and supporting monitoring, summarisation, and classification tasks [5,6]. More recent research has shown its relevance for market-driven product design [7], product review analytics [8], online-review-based product evaluation [9], aspect-opinion extraction [10], warranty and service improvement [11], and AI-enhanced design support [12–14]. This body of work is useful, but much of it remains product-centred or service-centred. In many cases, the analytics layer is also kept separate from the engineering environment in which actual design decisions are made and, for PSS, this separation is limiting. Design teams need support that stays connected to the combination of product elements, service elements, and lifecycle context that defines the PSS, and to the practical task of deciding what to change, what to combine, and what to prioritise next.

The research reported here builds on results from the DIVERSITY project [15], the Horizon 2020 project “Cloud Manufacturing and Social Software Based Context Sensitive Product–Service Engineering Environment for Globally Distributed Enterprise”. The project developed the DIVERSITY engineering environment, an integrated software environment for context-sensitive collaborative PSS design. Within that environment, sentiment analysis was used as a decision-support layer linking stakeholder feedback to concept generation, redesign, and early evaluation of alternatives. In the remainder of this manuscript, DIVERSITY project refers to the funded research project, DIVERSITY engineering environment refers to the implemented software environment, and sentiment-analysis framework or proposed framework refers specifically to the feedback-to-design decision-support framework analysed in this paper.

The research phenomenon examined in this paper is the feedback-to-design discontinuity in PSS development: stakeholder expressions exist in distributed, uneven, and weakly structured forms, while design teams require structured, weighted, and interpretable evidence to support concept generation and redesign. The study is therefore framed as a framework-building and assessment contribution. Its research object is an artefact: an integrated decision-support framework that connects feedback representation, opinion weighting, extraction, prediction, and PSS concept support. The evaluation focuses on

the utility, internal consistency, applicability, and boundary conditions of this artefact in realistic industrial settings.

Several technical elements of this work were previously reported in project-related publications: the DIVERSITY engineering environment [3,4], the social-sentiment model and reach/influence formulation [16], the sentiment-analysis architecture [17], Opinion-Sim [18], and the Product-Service Concept Tree method [19]. The present manuscript does not claim these elements as isolated new modules. Its contribution lies in integrating them into a coherent PSS design-support framework, clarifying their joint role in feedback-driven decision-making, and reassessing the resulting framework across validation evidence, industrial cases, limitations, and boundary conditions.

Within this design-oriented framing, the paper makes four main contributions at the level of synthesis, integration, and design-support interpretation. First, it provides a unified formulation of the feedback-to-design problem in PSS contexts, connecting posts, sentiments, opinions, global sentiment, reach, influence, strength, intensity, stability, and conflict into a single decision-support logic. Second, it integrates the previously distributed DIVERSITY components into one operational architecture linking opinion modelling, monitoring, extraction, prediction, and PSS concept selection. Third, it clarifies how hierarchical similarity can be used as an exploratory mechanism for early reasoning about new PSS configurations, while explicitly delimiting the evidence currently available for prediction accuracy. Fourth, it consolidates validation evidence from preliminary polarity plausibility checking, controlled simulation, industrial cases, and software-quality assessment, and uses this evidence to identify the conditions under which the framework is most and least applicable.

In PSS settings, sentiment analysis has greater value when it is embedded in a design decision-support framework. Its role is to connect distributed stakeholder expressions to engineering choices in a form that can be monitored, segmented, interpreted, and projected forward with an explicit view of uncertainty. This is especially relevant in early design stages, where teams need grounded support without yet having full operational evidence.

The remainder of the paper is organised as follows. Section 2 positions the work against the state of the art and clarifies the research gap. Section 3 presents the proposed framework. Section 4 describes its implementation in the DIVERSITY engineering environment and its connection to PSS design activities. Section 5 reports validation evidence. Section 6 discusses contributions, limitations, and implications in light of the more recent literature. Section 7 concludes the paper and provides a reflection on the obtained results.

2. Background and Research Gap

PSS design methods have long emphasised the importance of understanding customer needs and integrating product and service design activities [1,2]. The recent PSS literature has also emphasised that PSS development involves not only products and services, but also actors, stakeholder relationships, and enabling technologies [20]. In parallel, digital PSS and servitisation studies have highlighted the role of data and digital technologies in supporting service-oriented business models, circular-economy transitions, and value creation in manufacturing contexts [21–23]. For PSS design, these contributions are relevant because they frame design as the configuration of product elements, service elements, actors, data flows, and lifecycle relationships. In practice, however, the operational mechanisms for feeding large volumes of stakeholder feedback into design remain underdeveloped. The DIVERSITY project itself started from the observation that manufacturing companies, especially SMEs, had limited means to collect, contextualise and exploit feedback from both business customers and final consumers in a structured PSS design process. The gaps identified in the project's state-of-the-art analysis included the lack of tools to analyse

customer and consumer feedback, weak integration between feedback and PSS design workflows, and insufficient support for context-sensitive knowledge reuse.

In parallel, sentiment analysis research has provided increasingly powerful ways to detect polarity, extract opinions and summarise trends in text [5,6]. Yet classical sentiment analysis was not originally developed for PSS design. Many studies focus on document, sentence or aspect-level polarity for products, brands or generic service settings. More recent studies have extended this line toward market-driven product design [7], sentiment analysis of online product reviews [8], product evaluation and improvement from review corpora [9], and more flexible aspect-opinion analysis [10]. Studies on service operations and warranty processes have also shown that feedback mined from digital sources can support operational improvement [11]. The latest AI-based work pushes further toward BERT-based aspect extraction and fine-grained sentiment classification [13,14].

In a contemporary implementation, established baselines for the polarity layer would include lexicon-based methods, classical supervised classifiers, transformer-based models such as BERT, RoBERTa or DeBERTa, aspect-based sentiment models, and LLM-assisted classification. These methods are especially relevant for improving the extraction and classification layer. Recent aspect-based sentiment research also shows the value of manually annotated review data and transformer-based models, such as RoBERTa, for deriving actionable customer-satisfaction insights from user-generated content [24]. The research gap addressed here is different: how sentiment outputs, once available, can be represented, weighted, connected to PSS structures, and reused in support of concept generation and redesign decisions.

The academic debate can therefore be read across three complementary methodological traditions. PSS design methods provide structures for combining product and service elements, but they often rely on feedback already interpreted by the organisation [1,2,20]. Product review and sentiment-analysis approaches provide stronger mechanisms for extracting customer opinions from large volumes of text, but they usually remain closer to product evaluation, feature analysis, or service-quality monitoring [7–14,24]. Decision-support approaches can help compare alternatives, but they require indicators that are already structured and meaningful for the design problem. The methodological relevance of the present framework lies in connecting these traditions: it transforms distributed stakeholder expressions into weighted and interpretable indicators, links them to PSS product–service structures, and makes them usable in early concept generation and redesign decisions.

This evolution is valuable, but it still leaves an important gap for PSS. First, much of the literature assumes that the analytic target is either a product or a service, whereas PSS design concerns a coupled system whose product and service elements interact over time. Second, many approaches terminate at the level of classification, scoring or topic discovery, without an explicit connection to the design environment where concept generation, trade-offs and configuration decisions are made. Third, prediction of the likely reception of a new configuration remains less explored than the monitoring of existing offerings. Fourth, the relative importance of opinions in networked environments is often reduced to polarity alone, although the actual impact of an opinion depends also on who expresses it and how far it propagates.

The work reported in the DIVERSITY project addresses these issues from a design-support perspective. It does not attempt to compete with state-of-the-art transformer models in raw classification performance. Instead, it proposes an integrated engineering framework in which sentiment analysis is one layer in a broader PSS lifecycle environment [4]. The focus is on how feedback is collected, weighted, visualised, connected to design context, and used for early decision support. The practical issue is on how to make

that feedback usable at the point where design decisions are actually taken. In this sense, the framework is complementary to more recent advances in AI-based sentiment analysis: the underlying polarity engine may evolve, but the architectural and methodological problem of connecting feedback to PSS design remains.

From this perspective, the central issue is the continuity of the feedback-to-design chain. In PSS design, stakeholder expressions must be transformed into structured feedback, weighted according to their diffusion potential, interpreted in relation to product and service elements, and reused in concept generation or redesign. The contribution of the present framework lies in organising these steps as a connected decision-support process, where sentiment analysis is one component of a wider PSS design activity. This remains relevant for current research because contemporary sentiment-analysis engines can improve the extraction layer, while the design-support problem remains: feedback still has to be linked to PSS structures, weighted according to relevance and diffusion, and used in early concept decisions (as depicted in Figure 1).

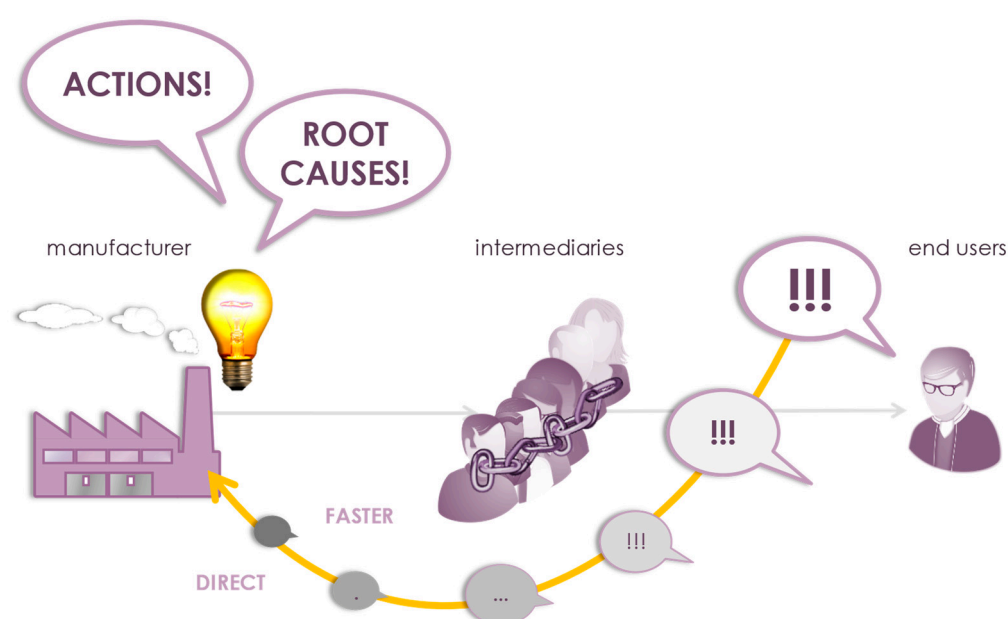


Figure 1. Obtaining first-hand feedback related to PSS design in DIVERSITY.

Taken together, these gaps lead to the following research questions:

1. How can stakeholder feedback relevant to a PSS be represented and aggregated so that designers can interpret both current reception and its temporal evolution?
2. How can the relative impact of opinions be modulated through variables such as reach and influence rather than polarity alone?
3. How can sentiment knowledge associated with existing products and services be used to estimate the likely sentiment of a new PSS configuration?
4. What evidence supports the practical usefulness of this framework in realistic industrial settings?

These research questions are addressed through framework construction and multi-level assessment. The study does not formulate causal hypotheses; it develops and assesses an artefact intended to address a practical and theoretical gap in PSS design support. The research procedure followed five steps. First, the literature on PSS design, sentiment analysis, product review analytics, and decision support was analysed to identify the feedback-to-design gap. Second, previously distributed DIVERSITY components were integrated into a single framework connecting feedback representation, reach/influence

lishes a structured representation of feedback entities, distinguishing posts, sentiments, opinions, and global sentiment. The second introduces network-aware weighting through reach and influence, so that the relative impact of opinions reflects their actual diffusion potential rather than treating all contributions equally. The third organises these capabilities into an operational architecture of four modules—opinion modelling, monitoring, extraction, and prediction—each addressing a distinct phase of the feedback-to-design cycle. The fourth extends the framework beyond monitoring existing offers by enabling sentiment estimation for new PSS configurations not yet released to the market.

Together, these layers give design teams a continuous and structured connection between what stakeholders express and what designers need to decide. The following subsections describe each layer in detail.

3.1. Positioning Inside the DIVERSITY Engineering Environment

The DIVERSITY engineering environment integrated product/service data management, KPI assessment, social knowledge acquisition, sentiment analysis, context extraction, and intelligent search as services for collaborative PSS design. In this architecture, the sentiment-analysis layer converted external and internal stakeholder feedback into indicators usable during concept generation, redesign, and evaluation.

The framework analysed here therefore sits between feedback acquisition and design action. Upstream, it receives social-media, collaboration-space, and repository data. Downstream, it provides indicators and visualisations that can be used with concept-selection tools such as the Product–Service Concept Tree. In DIVERSITY scenarios, feedback could reveal a new PSS opportunity or support the update of an existing one; predicted customer opinion could then be considered together with impact and implementation difficulty when selecting candidate solutions [19].

3.2. Core Representation: From Post to Global Sentiment

A central step in the work reported in the DIVERSITY project was the clarification of the levels at which feedback is represented [16]. Four entities were distinguished (as shown in Figure 3).

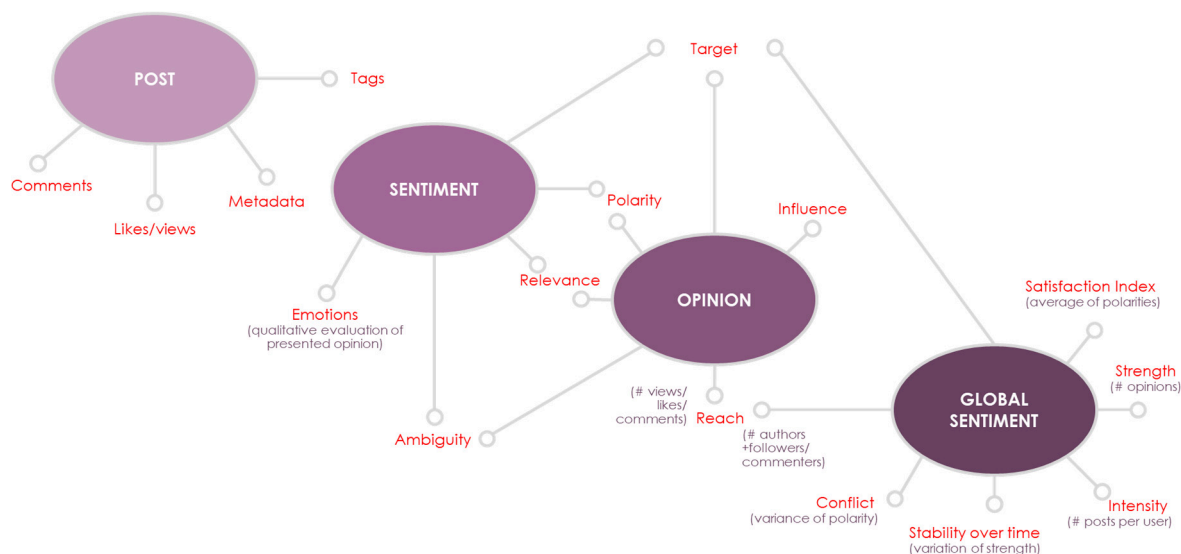


Figure 3. Representation of posts, sentiments, opinions, and global sentiment.

A *post* is a piece of content published by a user in a social or collaborative medium, expressing a position on a target. A post may be followed by comments, approvals or other interactions.

A *sentiment* is the attitudinal content directed toward a target within an individual post or comment. In the social-sentiment model used here, it comprises, at minimum, the target, polarity, ambiguity, and emotions.

An *opinion* is the global intrinsic sentiment of a posted content. It aggregates the targeted sentiments expressed in the post and in related comments. In this sense, the opinion is not merely the polarity of a sentence; it is the summarised stance associated with a specific conversation or exchange.

A *global sentiment* emerges from the aggregation of opinions over a time interval. It represents the overall weighted polarity toward a target and includes not only polarity but also auxiliary descriptors: *strength* (number of opinions in the sampled interval), *intensity* (number of posts per user regarding the target), *reach* (related to audience), *stability* (changes in strength over time) and *conflict* (variance in the polarity of opinions).

This representation matters because it separates different analytical questions. Polarity alone answers whether a message is favourable or unfavourable. The opinion level answers what a local discussion means in aggregate. The global sentiment level answers how the target is being received over time, by whom, and with what degree of dispersion or volatility.

The representation is intended to be layered and configurable, not necessarily fully instantiated in every deployment. A minimal implementation can operate with targets, polarity, time windows, sources, and opinion aggregation. Additional descriptors such as reach, influence, intensity, stability, and conflict can be activated when the corresponding data are available and relevant for the design decision. This reduces the initial setup effort and allows the framework to support both simple monitoring scenarios and richer decision-support scenarios.

3.3. Weighting Opinions Through Reach and Influence

One of the differentiating elements of the DIVERSITY framework is the explicit modelling of *reach* and *influence* [18]. In networked environments, not all opinions should contribute equally to the global sentiment. An opinion that reaches many viewers, comments and approvals has higher diffusion potential than one that remains marginal. Likewise, an author whose contributions systematically obtain high reach may influence the overall sentiment more than a peripheral participant.

The theoretical rationale for the weighting structure is that diffusion potential combines exposure and engagement. Views are treated as a proxy for potential audience size, while likes and comments represent different levels of interaction with the content. Comments are distinguished from likes because they require more active participation and can extend the conversation around the target, whereas likes indicate a lighter form of approval or attention. The coefficients w_c , w_l , and w_v therefore do not represent universal constants; they encode how the monitored source and the design team interpret the relative importance of comments, likes, and views for a given analysis.

In the implemented model, reach is computed from normalised counts of comments, likes and views, or their analogues in the monitored platform. These three interaction variables are combined through non-negative weighting coefficients, w_c , w_l , and w_v , corresponding respectively to comments, likes and views. The coefficients are constrained so that $w_c + w_l + w_v = 1$. The reach value is therefore calculated as a weighted sum of normalised interaction indicators, as formulated in the following equation:

$$\text{reach}(p_i) = w_c \frac{\#comments(p_i)}{\phi_{\text{comments}}} + w_l \frac{\#likes(p_i)}{\phi_{\text{likes}}} + w_v \frac{\#views(p_i)}{\phi_{\text{views}}} \quad (1)$$

where ϕ denotes global or contextual averages used for normalisation [18].

The coefficients are treated as configuration parameters of the opinion model. They are defined before running the monitoring process and should remain fixed during each analysis run. Their selection depends on the monitored source and on the intended interpretation of diffusion potential. For example, comments may receive higher weight when the objective is to capture active engagement and discussion, while views may receive higher weight when exposure is considered more important than interaction. When no source-specific or domain-specific reason exists to prioritise one interaction type, an equal-weight setting can be used as a neutral baseline.

For reproducibility, an application of the framework should therefore report: (i) the monitored source; (ii) the interaction variables available in that source; (iii) the coefficient vector used for reach calculation; (iv) the normalisation basis used for comments, likes and views; and (v) whether the same coefficients were used across all monitored sources or adjusted by source type. In the present implementation, the coefficients were used to support transparent weighting of opinions, not to estimate causal influence or to optimise predictive accuracy.

Influence is computed analogously at author level from the author's average interaction history, using the same logic of normalised interaction indicators and pre-defined coefficients. If a different coefficient vector is used for influence, this should also be specified as part of the opinion model configuration. The distinction is important: a post may have high reach even if its author is not generally influential, and an influential author may occasionally publish a post with limited reach.

These quantities are then used as weights in the aggregation of opinions. Conceptually, the framework assumes that polarity should be interpreted together with diffusion potential and network status. This is especially relevant in domains where electronic word-of-mouth can amplify or dampen perceptions around specific products or services [16].

The procedure implemented for reach- and influence-weighted aggregation can be summarised as follows.

From a computational perspective, the reach and influence indicators have limited complexity once the interaction data have been retrieved. For a monitoring window containing n relevant posts or opinion structures, the normalisation of comments, likes, and views and the computation of reach values are linear in the number of monitored items. The main scalability constraints therefore do not arise from the weighting formula itself, but from data acquisition, storage, sentiment extraction, metadata availability, and repeated updating over large source streams.

3.4. Operational Architecture

The implemented framework is organised into four interacting modules [17]. Together, they define how stakeholder feedback is scoped, collected, interpreted, and reused in support of PSS design decisions.

Opinion modelling is the user-facing stage in which the monitoring context is defined. An opinion model specifies the target PSS, the relevant sources, whether final products should be included, and the update frequency. Multiple models can coexist for the same PSS to support different monitoring scopes or segmentation needs.

Opinion monitoring is the background process that periodically retrieves new posts, reconstructs their conversational structure, computes polarity, and derives reach and influence. It is the central runtime module of the framework.

Opinion extraction provides the visual analytics used by designers. It combines filtering and segmentation with graphical summaries and direct inspection of posts. This is where the framework becomes actionable for design practice.

Opinion prediction estimates the expected sentiment toward a new PSS based on the sentiment histories of similar products and services. This module supports concept design rather than monitoring only existing offerings.

This architecture, shown in Figure 4, is deliberately modular. It allows the polarity engine to be implemented either through a commercial service or an internal dictionary-based approach. It also separates acquisition, computation, visualisation and estimation concerns, which proved useful both for engineering integration and for validation.

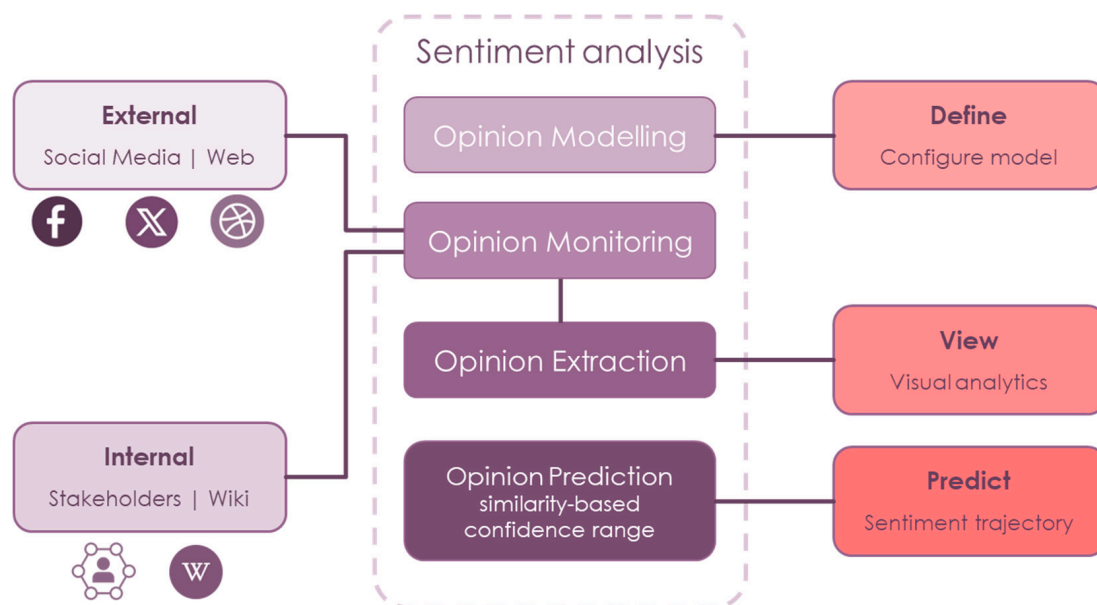


Figure 4. Functional architecture of the sentiment-analysis framework within the DIVERSITY engineering environment.

Figure 4 complements the decision-support flow shown in Figure 2 by detailing the functional modules used for opinion modelling, monitoring, extraction, and prediction inside the DIVERSITY engineering environment.

3.5. From Monitoring to Prediction: Hierarchical Similarity for New PSS Configurations

A key limitation of conventional feedback analytics is that they explain reactions to offerings already in use. PSS design, however, also needs forward-looking evidence. The DIVERSITY prediction module addresses this by estimating the likely sentiment toward a new configuration from the sentiment associated with similar products and services.

The underlying assumption is that PSS elements can be represented inside hierarchical trees, where products and services are described by related categories, versions and features. When a new PSS is being designed, the system identifies similar existing elements and reuses their sentiment histories.

This assumption has three implications. First, the product and service hierarchies must encode meaningful functional or market proximity between elements. Second, proximity in the hierarchy is treated as a proxy for the transferability of stakeholder reactions from existing offerings to new configurations. Third, the sentiment associated with similar product and service elements is assumed to provide useful early evidence for concept comparison, especially when direct market feedback on the new PSS does not yet exist. These assumptions are strongest when the hierarchy is curated by domain experts and when the compared elements share use contexts, customer groups, or service conditions. They are weaker for radical innovations, poorly maintained taxonomies, or cases where

sentiment is driven mainly by brand, price, regulation, or contextual factors outside the product–service hierarchy.

Similarity is computed using a hierarchical extension of the cosine similarity idea, following the approach of Ganesan et al. [25]. Let U be a rooted hierarchy and let l_1 and l_2 be two leaves. If $LCA_U(l_1, l_2)$ is their lowest common ancestor and $depth(\cdot)$ is the depth in the hierarchy, then similarity is computed as

$$\text{sim}(l_1, l_2) = \frac{2 \cdot \text{depth}(LCA_U(l_1, l_2))}{\text{depth}(l_1) + \text{depth}(l_2)} \quad (2)$$

This value lies between 0 and 1 and equals 1 only when the two elements are identical [25]. Other similarity measures could be used in this step, including exact matching, attribute-vector similarity, set-based overlap, or embedding-based similarity. The hierarchical measure was adopted because it is interpretable, can use the product and service taxonomies already present in the engineering environment, and does not require a large, labelled dataset. Its limitation is that it depends on the quality and granularity of the hierarchy. A comparative evaluation against alternative similarity measures was not part of the present validation and is therefore identified as future work. In operational terms, the prediction module performs five steps: it receives the products and services selected for the new PSS; it searches the product and service trees for elements above a similarity threshold; it retrieves the sentiment histories of PSSs containing similar elements; it computes weighted sentiment distributions; and it estimates the predicted sentiment for the new PSS by combining the product-side and service-side results.

The prediction is accompanied by an explicit confidence interpretation. Two mechanisms are used. First, similarity itself acts as a relevance filter: low similarity reduces confidence in the projection. Second, the implementation aligns the temporal evolutions of relevant PSSs and computes weighted mean and standard deviation over time using inverse-variance weighting, generating a confidence envelope around the predicted trajectory. This is important because a design team should not only receive a forecasted value; it should also understand how reliable that forecast is, given the available evidence.

In this form, the prediction module should be understood as an exploratory decision-support mechanism instead of a validated forecasting model. Its purpose is to structure early reasoning about likely reception by reusing relevant historical sentiment evidence, not to provide outcome-level prediction accuracy comparable to a supervised benchmark trained and tested on held-out industrial observations.

4. Prototype Implementation and Connection to Design Practice

4.1. Feedback Acquisition from External and Internal Stakeholders

The framework distinguishes external feedback from public or web-visible sources and internal feedback from company personnel, business partners, customers, collaboration systems, claims, questionnaires, and wiki-based tools. This distinction is central in PSS design because public sentiment is useful mainly when the PSS is connected to consumer-visible products or services. In more technical B2B settings, structured internal or partner-generated feedback may be the main source of design evidence. The framework therefore includes both public feedback acquisition and a wiki-based knowledge-acquisition layer for technical comments, milestone feedback, and issue histories.

The prototype allowed the user to create an opinion model by selecting the target PSS, the monitored sources, whether final products were included, and the update frequency. This configuration stage matters because PSS monitoring is never context-free: different offerings may require different temporal resolutions, segmentation choices or source portfolios. In practice, the configuration effort is concentrated at the creation of

the opinion model. Once the model is defined, collection, polarity processing, reach calculation, and visualisation are executed by the monitoring cycle, reducing the burden on designers during routine use. Once created, an opinion model enters the monitoring cycle. New posts are periodically gathered from the configured sources, grouped into threads or conversations, and enriched with interaction indicators such as comments, likes and views. Polarity was represented on a 0 to 100 scale, with values below 50 interpreted as negative. At this layer, the framework was agnostic regarding the polarity engine and could use either a commercial sentiment-analysis service or an internal dictionary-based method. The same interface could also be connected to a contemporary BERT-, RoBERTa-, DeBERTa-, or LLM-assisted classifier. At integration level, the required condition is that the classifier output can be mapped to the framework representation, namely a target or aspect when available, a polarity value or class, and optionally a confidence score. The downstream components for reach, influence, aggregation, visualisation, and PSS concept support are independent of the specific classifier used to generate the polarity value.

4.2. Opinion Extraction and Visual Analytics

The opinion extraction module supports design interpretation rather than simple data display. Designers can filter the results by gender, location, age, or final product when the necessary metadata exist. Three visual devices are central.

First, the module provides an integrated chart page containing:

- The total number of opinions;
- A polarity distribution chart;
- A gauge for current sentiment;
- The average reach value;
- Reach evolution over time;
- Sentiment evolution over time.

Together, these charts support trend reading, anomaly detection and time-based inspection. They also support routine design decisions by helping the team judge whether a signal is broad or segment-specific, recent or persistent, and strong enough to justify intervention.

Second, a *tag cloud* summarises the most frequent words extracted from high-reach posts. The size of the terms is proportional to occurrence, allowing rapid recognition of dominant positive or negative vocabulary. The prototype linked the tag cloud to the time charts so that the user could restrict the cloud to a selected time point, avoiding misleading mixes of past negative and recent positive language.

Third, a *Top 5 table* exposes the posts with highest reach and their comments, together with metadata such as author, date, polarity, reach, location, gender and age when available. This direct inspection layer is important because designers often need to read the actual statements behind aggregate curves.

Taken together, these visualisations operationalise the feedback-to-design link. They allow the designer to move from a global reading of performance to specific comments, time periods and segments requiring intervention.

4.3. Prediction Interface

The prediction page allows the design team to select related products and services for the PSS under development and obtain an estimated sentiment trajectory with an associated confidence envelope. This is not a black-box prediction of generic market success; it is a structured reuse of sentiment evidence from similar offerings and configurations. The value of this module lies in early design stages, when the team needs an evidence-informed estimate of likely acceptance before investing in full development.

The interface is therefore intended to support early comparison of alternatives under uncertainty. It provides indicative, evidence-informed estimates to guide discussion and concept screening, but it is not presented as a substitute for post-deployment validation or market observation.

4.4. *OpinionSim: Controlled Feedback Generation for Testing and Segmentation*

A recurrent problem in validating feedback analytics is the lack of large, controlled datasets with known sentiment structure. *OpinionSim* was developed precisely to address this issue [18]. The simulator generates large numbers of posts about products or PSSs while controlling population characteristics, opinion distributions and influence effects.

The simulator models three core object groups: PSSs, actors and posts. Actors are characterised by age, gender and location. The user can define population distributions over these variables and specify how sentiment toward a product varies as a function of them. A temporal baseline over a 360-day period can also be superimposed.

OpinionSim explicitly incorporates the same logic of reach and influence used in the sentiment framework. It simulates special authors responsible for triggering conversations and assigns them differentiated influence levels, from which variable numbers of comments, likes and views are generated. Posts are then stored in a database and retrieved by the sentiment-analysis component as if they had been acquired from the outside world.

This capability is important for two reasons. First, it supports functional validation under known conditions. Second, it makes segmentation testing possible even when public data sources do not provide all the metadata required for gender-, age- or location-based filtering.

4.5. *From Sentiment Indicators to Concept Selection*

The present paper focuses on the sentiment-analysis and prediction framework, while the Product–Service Concept Tree (PSCT) method provides the concept-selection context [19]. In the concept-design phase of the DIVERSITY engineering environment, candidate PSS solutions are evaluated using criteria such as perceived impact, implementation difficulty, and predicted customer opinion expressed through global sentiment. The sentiment framework therefore acts as an evidence layer inside a broader decision-support activity.

This decision-support activity is explicitly human-in-the-loop. Sentiment indicators, reach and influence values, segment-specific concerns, and predicted opinion trajectories are presented to designers as inputs for interpretation, discussion, and concept screening. Designers then combine these indicators with feasibility, expected impact, implementation difficulty, and business relevance. In this sense, the framework is compatible with multicriteria decision-making procedures: sentiment-derived indicators can be used as criteria or criterion inputs together with technical, economic, and implementation criteria. The framework does not select the final PSS concept automatically; it supports the team in comparing alternatives and justifying the selected direction.

5. Validation and Assessment

5.1. *Validation Strategy*

The project assessment methodology distinguishes between evaluation in the *use phase* and evaluation in the *development phase*. In the use phase, industrial partners execute structured validation scenarios corresponding to their business cases and provide both qualitative and quantitative feedback. In the development phase, software quality, correctness and performance are monitored through code-quality and testing metrics.

For the sentiment-analysis work, validation was therefore assembled from four complementary sources of evidence:

1. Preliminary polarity plausibility checking against human judgement.
2. Controlled functional testing with *OpinionSim*-generated data.
3. Industrial validation in three industrial scenarios following structured workflows.
4. Software-quality and integration evidence for the full prototype.

Figure 5 summarises the validation strategy adopted in this study. It shows how the four evidence sources contribute to different aspects of framework assessment, from polarity plausibility and controlled functional behaviour to industrial usefulness, boundary conditions, and implementation maturity. This distinction is important because the proposed framework is not validated through a single benchmark, but through the convergence of complementary forms of evidence.

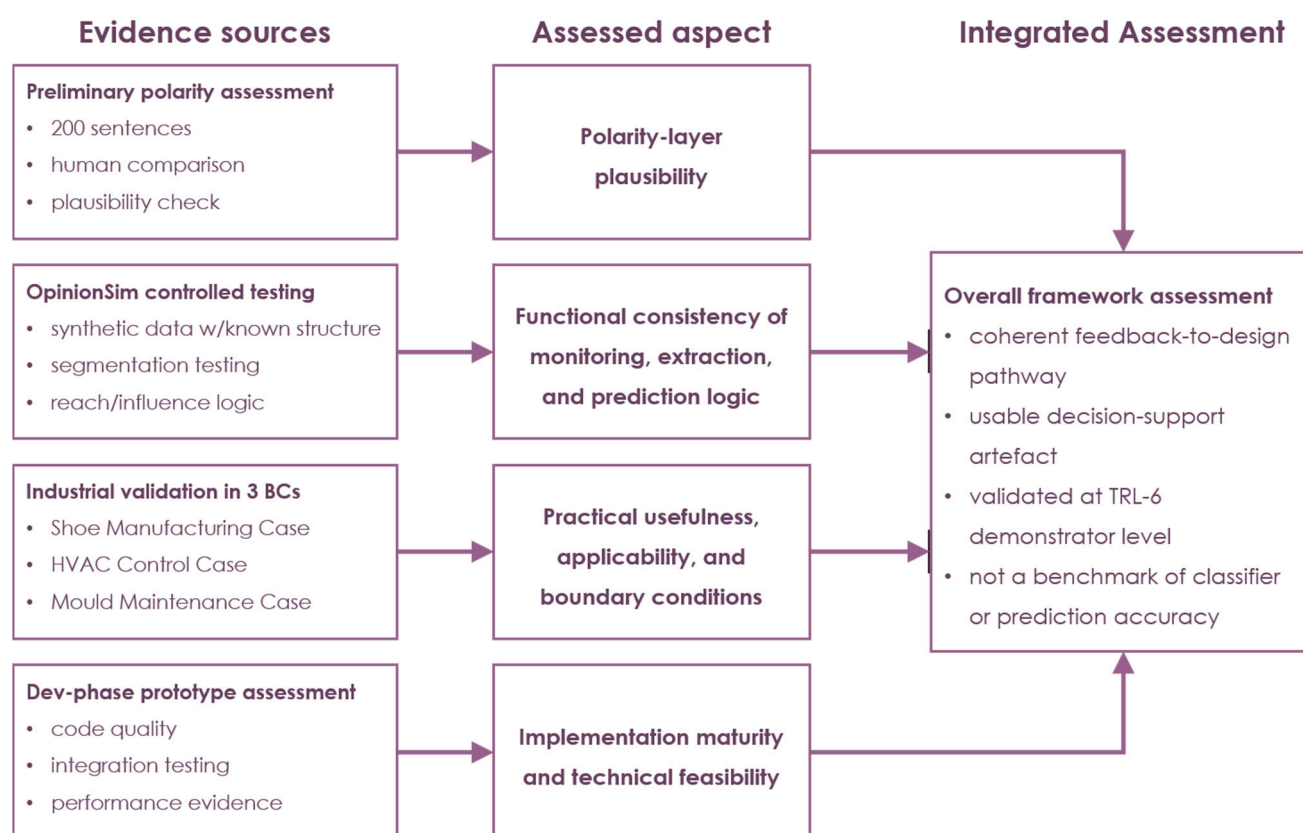


Figure 5. Multi-level validation strategy of the proposed framework.

These sources of evidence should be read as the assessment basis for a TRL-6 engineering demonstrator, not as a controlled statistical benchmark of the polarity or prediction modules. The validation focuses on functional correctness, integration with the PSS engineering environment, usefulness in structured industrial scenarios, implementation maturity, and boundary conditions.

This layered approach is appropriate because no single validation exercise can cover all aspects of such a framework. Classification quality, functional coverage, industrial usefulness and software robustness are distinct concerns. The weighting coefficients used in reach and influence calculation are therefore treated as documented configuration choices of the framework, not as fitted parameters validated independently in the industrial cases.

The goal of this validation is therefore not to establish statistical generalisability, classifier-level performance, or prediction accuracy. It is to provide converging evidence

of functional correctness, practical usefulness, implementation maturity, and boundary conditions across different industrial settings.

This evaluation logic is consistent with the design-oriented framing of the study. The relevant question is not only whether each technical component performs as an isolated algorithm, but whether the constructed artefact provides a usable and coherent pathway from stakeholder feedback to PSS design support. The polarity check, OpinionSim testing, industrial cases, and software-quality evidence therefore address different aspects of artefact assessment.

5.2. Preliminary Polarity Assessment

A first assessment of polarity accuracy was conducted on a set of 200 sentences by comparing the output of a sentiment-analysis tool with human classification. The sentences were manually reviewed by members of the project research team with expertise in the sentiment-analysis component. Each sentence was assigned a polarity judgement using a simple interpretation protocol: positive when the expressed attitude toward the target was favourable, negative when it was unfavourable, and neutral or ambiguous when no clear polarity could be assigned. The comparison indicated an overall agreement of 80%, while 19% of classifications presented a deviation above 30% from human judgement [16].

This assessment had a limited calibration purpose. It was used to check whether the polarity layer produced plausible results before integration with reach, influence, segmentation, and design-support functions. It was not designed as a full annotation campaign. No independent inter-rater reliability measure, such as Cohen's kappa or Krippendorff's alpha, was calculated. For this reason, the 80% agreement value is treated here as preliminary evidence of plausibility, not as a complete validation of sentiment-classification performance. Precision, recall, F1-score, confusion matrices, ROC analysis, and statistical significance tests were not calculated in the original assessment. Consequently, the polarity result is used only to document initial plausibility of the sentiment layer and not to claim classifier-level performance.

The result nevertheless supports one design choice of the framework: polarity output should remain interpretable by human users and should be combined with contextual indicators before it informs design decisions. In the proposed framework, the polarity layer is only one input to a wider decision-support process that also includes reach, influence, segmentation, temporal evolution, and designer interpretation.

5.3. Controlled Validation with OpinionSim

OpinionSim was used to test whether the sentiment-analysis component could recover the biases introduced in a controlled dataset. The simulator makes it possible to vary sentiment distributions by age, gender and location, to superimpose temporal sentiment trends, and to inject differentiated influence levels through privileged authors. Because the generated posts are stored in a database and retrieved by the platform, the sentiment-analysis module can be tested under conditions resembling real operation while the ground-truth structure remains known.

This controlled setting is particularly relevant for segmentation and visual analytics. The filtering mechanisms in opinion extraction, the interpretation of reach and influence, and the temporal charts can all be verified without relying on noisy real-world social data. *OpinionSim* therefore provides a bridge between conceptual modelling and industrial deployment. No held-out industrial dataset was available to compare predicted sentiment trajectories with later observed stakeholder sentiment.

Although *OpinionSim* was designed primarily to validate the monitoring and extraction modules, it also provided indirect evidence for the prediction module. The prediction

module relies on weighted aggregation of sentiment histories from similar products and services, i.e., the same histories that *OpinionSim* generates under controlled conditions. Because the ground-truth sentiment structure was known, it was possible to verify that the weighting and aggregation logic recovered the expected distributions before the module was exposed to noisier industrial data. This verification followed the aggregation and prediction procedures summarised in Algorithms 1 and 2. This does not constitute a direct benchmark of prediction accuracy, but it confirms that the computational core of the module behaves as specified.

Algorithm 1. Reach- and influence-weighted opinion aggregation

Input: Target PSS or product/service element; monitoring time window; collected posts and comments; polarity scores; available interaction indicators; reach coefficients; aggregation settings.

Step 1: Select the posts and comments referring to the target within the chosen time window.

Step 2: Identify the sentiment target and polarity associated with each relevant post or comment.

Step 3: Group posts, comments, and related interactions into opinion structures.

Step 4: Normalise available interaction indicators, such as comments, likes, and views, according to the selected source.

Step 5: Compute reach for each opinion using the coefficient configuration defined in the opinion model.

Step 6: Compute author influence from historical interaction behaviour when such information is available.

Step 7: Combine polarity, reach, and influence according to the configured aggregation settings.

Step 8: Compute the global sentiment for the selected target and time window.

Output: global sentiment value, opinion-level reach and influence values, and time-window indicators for design interpretation.

Algorithm 2. Similarity-based sentiment estimation for a new PSS configuration

Input: new PSS configuration; selected product and service elements; product hierarchy; service hierarchy; historical PSS configurations; sentiment histories; similarity threshold.

Step 1: Identify the product and service elements included in the new PSS configuration.

Step 2: Search the product and service hierarchies for existing elements comparable to those selected for the new configuration.

Step 3: Compute hierarchical similarity between the selected elements and the existing elements using Equation (2).

Step 4: Retain only historical configurations whose similarity is above the defined threshold.

Step 5: Retrieve sentiment histories associated with the retained configurations.

Step 6: Align the retrieved sentiment histories over the relevant temporal reference.

Step 7: Combine the retained histories according to similarity and available confidence information.

Step 8: Generate the estimated sentiment trajectory and associated confidence envelope.

Output: Estimated sentiment trajectory, retained historical configurations, similarity basis, and confidence indication for design interpretation.

Accordingly, the evidence reported here supports internal consistency of the prediction logic, not predictive accuracy in the strict empirical sense. The study validates that the module combines similarity-weighted histories as intended and produces directionally coherent outputs under controlled conditions, but it does not establish how closely those outputs would match future observed sentiment in a real deployment.

5.4. Industrial Validation in Three Business Cases

The quantitative metrics reported below are project-time estimates produced during structured TRL-6 validation activities. They capture perceived and projected value at the prototype stage, not operational outcomes measured after project completion.

In the following case descriptions and in Table 1, baseline values refer to the initial situation reported by the industrial partners, full-prototype values refer to estimates produced during the TRL-6 validation exercises, and post-deployment values refer to project targets or expected benefits. They are not treated as measured longitudinal industrial performance. The three cases correspond to the industrial validation scenarios available in the DIVERSITY full-prototype assessment and are used here as instrumental cases. Together, they cover three relevant conditions for the framework: a consumer-visible setting where public feedback can be connected to PSS redesign, a B2B setting where structured internal and partner feedback substitutes for public social media, and a boundary case where the sentiment-analysis layer is less applicable but the wider feedback-to-design environment remains useful.

Table 1. Business metrics reported during full-prototype validation.

Business Case	Metric	Baseline	Full-Prototype Estimate	Expected Post-Deployment Value
Shoe Manufacturing Case	Time until customer problem is solved	≥ 0.5 days	30–60 min	Minutes
	Time for PSS development or improvement	≥ 12 months	<6 months	1–6 months
HVAC Control Case	Number of new needs identified	Not available	5	20
	Percentage of redesign required	5%	4%	2%
Mould Maintenance Case	Capacity for managing PSS	0	3	5
	Data availability within PSS	Hours to weeks	Seconds to minutes	Seconds to minutes

5.4.1. Shoe Manufacturing Case

The Shoe Manufacturing Case, centred on a German supplier of shoe-production machinery and associated services, is the setting in which the feedback-to-design logic is most directly visible. Because the machines are used to produce consumer-facing shoes, social media feedback on shoes can be related back to the machinery and services involved. In the full-prototype validation, users worked with sentiment analysis, social knowledge acquisition, KPI assessment, intelligent search, PSCT and related tools inside a structured scenario.

The qualitative feedback on sentiment analysis was positive, especially regarding the ability to obtain feedback from social sources such as Facebook and Amazon and to feed this information into the wider engineering environment. At the same time, the case also revealed a limitation that is methodologically important: segmentation filters based on gender, location and age could not always be used because the retrieved public data did not

provide those metadata. This is a real-world reminder that segmentation-aware analytics depends on the quality and richness of source metadata.

From the business-metrics perspective, the full-prototype assessment estimated substantial benefits. In this case, the validation exercise produced an estimate that the number of manageable PSSs could increase from fewer than 30 before DIVERSITY to roughly 100 to 200 at prototype level, while values above 1000 were treated as a long-term project target for future deployment. The prototype-stage assessment indicated a reduction in the estimated time from at least 0.5 days to approximately 30 to 60 min at prototype level, while the time from identification of a new PSS opportunity to market release was estimated to drop from at least 12 months to less than six months. The authors note, however, that these figures are extrapolated estimates consistent with a TRL-6 demonstrator rather than operational production data.

5.4.2. HVAC Control Case

An Italian supplier of HVAC Control Systems used the DIVERSITY engineering environment to support the development of new service-enhanced offerings around a humidification system. This case is especially relevant because it shows how the sentiment-analysis logic can be adapted to a B2B context where public social media feedback is limited. Instead of relying primarily on public posts, the case gathered feedback through claims and technician questionnaires, which were stored and analysed inside the platform.

The validation used a dataset comprising 153 claims and 15 questionnaires related to the special humidifier unit. The analysis showed that 72% of the requests concerned installation, maintenance and commissioning phases, with installation emerging as the most problematic phase. These results were used to identify two main needs—long lifecycle and continuous functioning—and to derive candidate PSS solutions such as augmented reality, automatic installation support and maintenance interventions. In this case, the selected concept was an augmented-reality-based application supporting installation. In practical terms, the framework supported an early decision about where the new service concept should concentrate before detailed development began.

This outcome represents the closest industrial instantiation of the prediction module's intended use. The concept selection was informed by the sentiment estimate derived from similar offerings, and the eventual focus on AR-based installation support was consistent with the framework's identification of installation as the most salient problem area. This should be read as illustrative industrial evidence of decision relevance, not as a direct test of predictive accuracy.

The quantitative business metrics reported in the full-prototype validation were more moderate than in the Shoe Manufacturing Case, but still meaningful. In this case, the prototype validation recorded 5 newly identified needs against no formal baseline, with 20 needs defined as a future project target. The percentage of redesign required was estimated at 4% at full-prototype level against a 5% baseline, while 2% was reported as a post-deployment target. The reported reduction in PSS development or improvement time from at least 24 months to 18 months should also be read as an estimated project benefit, not as a measured operational result.

The HVAC Control Case demonstrates an important generalisation of the framework: the same feedback-to-design logic can operate over structured internal or partner questionnaires rather than public social streams, provided that textual or semi-structured stakeholder feedback is available.

5.4.3. Mould Maintenance Case

The Mould Maintenance Case, centred on a mould manufacturer in Greece, offers the clearest view of the framework's boundary conditions. In the full-prototype validation, the Mould Maintenance Case successfully used wiki-based knowledge acquisition, KPI assessment, PSCT, service-delivery modelling and related tools to design new maintenance-oriented PSS offerings. However, the sentiment-analysis component was not substantively tested, because the case operates in a strictly B2B context with limited and highly technical feedback, and no relevant end-consumer social-media stream.

The Mould Maintenance Case shows that the broader DIVERSITY engineering environment remains useful even when public sentiment analysis is not applicable, but it also makes clear that the sentiment-analysis layer is most valuable when one of the following conditions holds: (i) the PSS is connected to consumer-visible products with external digital traces; or (ii) internal stakeholder feedback is sufficiently rich and textual to support sentiment-like analysis. Interestingly, even in this case, sentiment analysis was recognised as an innovative concept with potential for SMEs, especially when end-customer feedback becomes available in more consumer-oriented applications.

5.4.4. Cross-Case Synthesis of Industrial Validation

Table 1 brings together the main business metrics reported across the three industrial cases. The metrics were selected during the project validation activities with the industrial partners as business-level indicators of expected usefulness in each case. They were not generated automatically by the sentiment-analysis framework. For this reason, they should be distinguished from the technical indicators defined in the framework, such as polarity, reach, influence, similarity, and temporal sentiment descriptors. Baseline values describe the initial situation reported by the partners; full-prototype values describe estimates obtained during structured TRL-6 validation scenarios; and post-deployment values describe project targets or expected benefits. The table is therefore used as complementary evidence of perceived and projected industrial value, not as a standalone proof of model validity.

These business metrics should be interpreted together with the qualitative case evidence and the technical validation reported in Sections 5.2, 5.3 and 5.5. They indicate whether the industrial partners considered the framework useful for improving their PSS-related processes. They do not, by themselves, validate the computational model. The suitability of the proposed framework is assessed through the convergence of several forms of evidence: preliminary polarity plausibility checking, controlled functional testing with OpinionSim, structured industrial use cases, and software-quality assessment.

Table 2 complements these business metrics by summarising the qualitative industrial validation evidence for each case. It identifies the main feedback source, the role played by sentiment analysis, the salient evidence obtained during validation, and the main constraint observed in each industrial setting. This distinction is important because Table 1 reports business-level indicators, while Table 2 summarises case-level evidence about applicability, usefulness, and boundary conditions.

Table 2. Summary of the industrial validation evidence.

Case	Main Feedback Source	Role of Sentiment Analysis	Salient Evidence	Main Constraint
Shoe Manufacturing Case	Public social media and wiki-based internal knowledge	Direct support to redesign decisions via market feedback on consumer-visible shoes	Prototype-level estimate: problem-solving time from ≥ 0.5 days to 30–60 min; PSS development time from ≥ 12 months to < 6 months	Limited user metadata reduced practical segmentation

Table 2. *Cont.*

Case	Main Feedback Source	Role of Sentiment Analysis	Salient Evidence	Main Constraint
HVAC Control Case	Claims and technician questionnaires	Structured stakeholder-feedback analysis to identify needs and support new service concepts	153 claims and 15 questionnaires analysed; 72% of requests concerned installation, maintenance and commissioning	Public social-media feedback scarce in the business context
Mould Maintenance Case	Wiki-based technical history and KPI snapshots	Boundary-case demonstration: sentiment layer not central, but wider feedback-to-design environment remains useful	Useful for maintenance-oriented PSS design and knowledge digitalisation	Strictly B2B setting with little relevant end-consumer social feedback

5.5. Development-Phase Evidence for the Full Prototype

The full-prototype validation of the DIVERSITY engineering environment also evaluated code quality, correctness, and performance. For the sentiment-analysis component, the first release comprised approximately 5068 lines of code and passed the quality gate, with critical bugs and vulnerabilities reported as corrected in the final assessed version. Integration testing was run across the platform, and performance tests showed acceptable average response times, with intelligent search being the most demanding function because it had to aggregate heterogeneous sources.

These results suggest that the computation of reach-related indicators was not the dominant scalability bottleneck in the assessed prototype. The more demanding operations were those involving heterogeneous data retrieval, integration across sources, and search over distributed information. Nevertheless, the validation was performed at demonstrator scale and does not constitute a stress test for very large-scale social-media streams.

These software metrics are included as complementary evidence of implementation maturity. They show that the framework was not limited to conceptual design or isolated prototypes, but reached an integrated demonstrator assessed in realistic use scenarios.

6. Discussion

6.1. Main Contributions

The contribution should be interpreted as a framework-building and assessment contribution centred on an integrated decision-support artefact. The paper does not claim to test a causal theory of PSS design behaviour. It contributes by structuring a relevant design problem, integrating previously distributed technical components into a coherent framework, and assessing the usefulness and limits of that framework through complementary forms of evidence.

The first contribution of this paper is conceptual. It clarifies a multi-level representation of feedback for PSS design in which posts, sentiments, opinions and global sentiment are distinguished, and where global sentiment is enriched by auxiliary indicators such as strength, intensity, stability and conflict. This structure remains useful even in the presence of more powerful NLP models, because it organises the path from local expressions to design-relevant aggregate evidence. This contribution is relevant because it connects a debate usually split between PSS design methods, sentiment-analysis techniques, and decision-support models.

The second contribution is methodological. The framework does not stop at polarity classification. It explicitly integrates network-aware weighting through reach and influence, operational modules for modelling, monitoring, extraction and prediction, and a role inside a broader engineering environment. The algorithmic descriptions added to Sections 3.3 and 3.5 make the aggregation and prediction procedures clearer and easier to

reproduce. In doing so, it addresses a gap still visible in the more recent literature: the connection between feedback analytics and actual design workflows.

The third contribution is methodological and forward-looking. The hierarchical similarity mechanism provides a tractable way of reusing sentiment knowledge associated with existing products and services when reasoning about new PSS configurations. While this does not amount to a fully validated forecasting method, it gives early design teams a more grounded basis than intuition alone when comparing alternatives before market launch.

The fourth contribution is empirical. The paper brings together and integrates several forms of evidence: preliminary polarity assessment, simulation, industrial validation, and software assessment. The cross-case results also make visible a more nuanced conclusion than a single successful case would allow: the framework shows practical value under specific conditions, but its value depends strongly on feedback availability, metadata richness, and the degree to which a PSS is connected to user-visible outcomes.

6.2. Relation to the More Recent Literature

Recent work has advanced sentiment analysis through richer aspect extraction, graph-based representations, deep learning and transformer models [10–14,24]. These developments are directly relevant to the polarity and aspect-extraction layer of the proposed framework. BERT-based and fine-grained aspect-level models could improve the identification of sub-component issues, reduce manual interpretation effort, and support more precise feedback mapping when designers need information about specific product or service features.

The present manuscript does not report a comparative benchmark against BERT-based or other contemporary sentiment-classification baselines. This is a limitation of the study. The industrial validation was conducted in the original DIVERSITY engineering environment, and the available evidence was produced for framework validation and industrial assessment, not for a controlled NLP benchmark with labelled training and test data.

This absence limits the claims that can be made about polarity-classification performance. The system-level contribution concerns the representation of stakeholder feedback, reach and influence weighting, connection to PSS structures, and reuse of sentiment histories in early design decision support. Compatibility with contemporary models is architectural: a transformer-based or LLM-assisted classifier can replace the original polarity component if its outputs are normalised into the sentiment representation used by the framework. Future work should test whether such replacement improves the quality of downstream design-support outputs, not only sentence-level polarity classification.

6.3. Limitations and Boundary Conditions

Several limitations deserve explicit acknowledgement.

First, the polarity layer assessed in the project reflects the state of sentiment-analysis tools available at the time and was evaluated through a limited preliminary calibration exercise. The 200-sentence comparison did not include an independent inter-rater reliability analysis, so the reported 80% agreement should be interpreted as plausibility evidence only. This constrains the reliability of downstream analytics and reinforces the need for human interpretation. Recent developments in sentiment analysis provide promising outlooks [26], even if relevant limitations remain [27].

Second, segmentation depends on source metadata. Validation in the Shoe Manufacturing Case showed that the use of public social data depends on the availability of age, location or gender information, reducing the practical value of segmentation filters. The weighting coefficients used for comments, likes and views are configurable parameters.

Although this makes the framework adaptable to different sources and business contexts, it also means that coefficient selection requires explicit reporting and, ideally, sensitivity analysis in future applications. The present validation did not include an independent optimisation or robustness analysis of these coefficients.

Third, implementation effort is itself a boundary condition. The full framework requires the definition of targets, sources, update frequency, available metadata, and weighting settings. This effort is justified mainly when the organisation expects recurrent PSS monitoring or repeated use in concept generation and redesign. For isolated or one-off design decisions, a simplified configuration using only the core representation and basic aggregation may be more appropriate.

Fourth, scalability is also a boundary condition. The computation of interaction-based indicators such as comments, likes, and views is conceptually simple and grows linearly with the number of monitored items, but practical scalability depends on data-source access, update frequency, storage volume, sentiment-processing cost, and dashboard responsiveness. The present validation supports feasibility at demonstrator level, but it does not establish performance under continuous high-volume social-media streams.

Fifth, the prediction module depends on the availability of sufficiently similar products and services in the repositories. The confidence envelope addresses this only partially; if the historical basis is too weak, prediction becomes speculative.

Sixth, the business metrics reported in the industrial cases are prototype-level estimates and project targets associated with a TRL-6 demonstrator, not long-term operational measurements. Baseline values, full-prototype estimates, and expected post-deployment targets are therefore distinguished explicitly in the revised manuscript. These values are useful as evidence of perceived and estimated value, but they should not be read as definitive productivity claims or experimentally verified industrial improvements.

Seventh, public social sentiment is not universally applicable in B2B environments. The Mould Maintenance Case makes clear that purely technical or sparse partner feedback may be better handled by knowledge-management and KPI tools than by classic social sentiment analysis.

Eighth, the prediction module was validated architecturally through simulation and informed one industrial concept-selection decision, but it was not benchmarked against held-out outcomes in any of the three cases. Prediction accuracy—in the sense of comparing estimated sentiment trajectories against subsequently observed ones—remains an open empirical question that future deployments should address.

Ninth, the study does not include a comparative experiment with contemporary BERT-based, transformer-based or LLM-assisted sentiment models. Such experiments would be needed to assess the classification performance of an updated polarity layer. The present results should therefore not be read as evidence that the original polarity engine performs competitively against current sentiment-classification methods. The contribution lies in the framework that connects sentiment outputs to PSS design support.

For that reason, the practical value of the module in its current form lies in supporting comparative early-stage reasoning rather than in delivering forecast-grade accuracy. Its outputs should be interpreted as structured evidence for concept selection and redesign discussions, not as validated predictions of market behaviour.

6.4. Implications for Future Development

The validation results point toward a clear roadmap. The first improvement lies in stronger aspect-level analytics, particularly for sub-component characteristics and their connections to the end-product. A direct benchmark comparing the original polarity layer with BERT-based, transformer-based and LLM-assisted models on domain-specific PSS feedback

should therefore be a priority in future work, including both classification performance and the effect of those models on downstream design-support indicators. Such future work should also include a larger labelled corpus, inter-rater reliability analysis, standard classification metrics, and longitudinal comparison between predicted and subsequently observed sentiment. For the prediction module, future work should also compare the hierarchical similarity measure with alternative similarity approaches, such as attribute-based, set-based, or embedding-based measures, and assess their effect on downstream sentiment estimation for new PSS configurations. The Shoe Manufacturing Case explicitly pointed to the need for better analysis of consumer-product sub-components and more structured sentiment snapshots. The second lies in tighter integration with company systems and internal tools. A further development path is the formalisation of the decision-support stage through multicriteria decision-making procedures and structured human-in-the-loop validation, so that sentiment-derived indicators can be combined more systematically with feasibility, impact, implementation difficulty, and business criteria. The third lies in richer multilingual and context-aware analytics, including seasonal or situational context variables already identified in the project's future-work discussion.

More broadly, the framework suggests that future research should connect modern LLM- or transformer-based opinion extraction to engineering representations such as product hierarchies, service blueprints, and PSS lifecycle models. That combination is still relatively rare in the literature and remains a promising research direction.

7. Conclusions

This paper integrated previously distributed DIVERSITY project results into a unified journal contribution on feedback-driven PSS design support. The proposed framework organises stakeholder feedback through a multi-level representation of posts, sentiments, opinions and global sentiment; enriches aggregation with reach and influence; implements operational modules for modelling, monitoring, extraction and prediction; and connects feedback analytics to PSS design activities through an engineering environment.

The work contributes both conceptually and practically. Conceptually, it clarifies how sentiment indicators can be embedded in a broader lifecycle-oriented PSS design framework. Practically, it demonstrates that such a framework can be implemented, validated with simulation, and assessed in realistic industrial scenarios. The main value of the framework lies in bringing structured decision support into the early and uncertain part of PSS design, where teams need to identify needs, interpret stakeholder reactions, compare alternatives, and prioritise redesign effort.

The conclusions should nevertheless be read within the limits of the validation setting. The three industrial cases provide evidence of usefulness across different conditions, but they do not guarantee generalisation to all sectors or all types of PSS. The framework appears strongest where rich textual feedback exists, where product or service elements can be connected to stakeholder expressions, and where meaningful metadata or internal feedback sources are available. Its applicability is weaker in settings with sparse technical feedback, limited access to user traces, poorly structured product–service taxonomies, or highly context-dependent reactions. The business metrics reported in the industrial cases are also prototype-level estimates and project targets, not operational measurements.

Future research should therefore extend validation across additional sectors, larger datasets, and longer deployment periods. From a methodological perspective, future work should include comparative benchmarking of contemporary sentiment-analysis engines, sensitivity analysis of reach and influence coefficients, comparison of alternative similarity measures for prediction, and longitudinal evaluation of predicted versus observed stakeholder sentiment. From a decision-support perspective, further work should formalise

the human-in-the-loop and multicriteria decision-making stage, so that sentiment-derived indicators can be combined more systematically with feasibility, impact, implementation difficulty, and business criteria.

Beyond the specific framework, the paper also points to a broader research agenda at the intersection of PSS design and sentiment analysis. Future theoretical work should examine how stakeholder sentiment evolves across the PSS lifecycle, how product–service coupling changes the interpretation of feedback, and how sentiment indicators can be connected to engineering representations such as service blueprints, product hierarchies, lifecycle models, and concept-selection methods. In this sense, the main contribution of the paper is not merely the use of sentiment analysis for PSS, but the positioning of sentiment analysis as one component of a wider feedback-to-design decision-support system.

Author Contributions: Conceptualisation, R.N.-S.; methodology, R.N.-S. and P.P.; software, P.P.; validation, R.N.-S. and P.P.; formal analysis, P.P.; writing—original draft preparation, R.N.-S.; writing—review and editing, P.P. and R.N.-S.; project administration, R.N.-S.; funding acquisition, R.N.-S. All authors have read and agreed to the published version of the manuscript.

Funding: The presented work was partly supported by the European Commission through Horizon 2020 Project DIVERSITY: Cloud Manufacturing and Social Software Based Context Sensitive Product-Service Engineering Environment for Globally Distributed Enterprise (FoF-05-2014-636692).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author because the data is included in the confidential and public deliveries of the DIVERSITY project.

Acknowledgments: The authors express their acknowledgement to the DIVERSITY project.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Abbreviations

The following abbreviations are used in this manuscript:

AI	Artificial Intelligence
B2B	Business to Business
B2C	Business to Consumer
BERT	Bidirectional Encoder Representations from Transformers
DSS	Decision-Support System
HVAC	Heating, Ventilation and Air-Conditioning
KPI	Key Performance Indicator
LLM	Large Language Model
NLP	Natural Language Processing
PSCT	Product–Service Concept Tree
PSS	Product–Service Systems
SME	Small and Medium-sized Enterprise
TRL	Technology Readiness Level

References

1. Morelli, N. Developing new product service systems (PSS): Methodologies and operational tools. *J. Clean. Prod.* **2006**, *14*, 1495–1501. [[CrossRef](#)]

2. Müller, P.; Kebir, N.; Stark, R.; Blessing, L. PSS Layer Method—Application to Microenergy Systems. In *Introduction to Product/Service-System Design*; Springer: London, UK, 2009; pp. 3–30.
3. Neves-Silva, R.; Pina, P.; Spindler, P.; Pezzotta, G.; Mourtzis, D.; Lazoi, M.; Ntalaperas, D.; Campos, A.R. Supporting context sensitive lean product service engineering. *Procedia CIRP* **2016**, *47*, 138–143. [\[CrossRef\]](#)
4. Mourtzis, D.; Fotia, S.; Gamito, M.; Neves-Silva, R.; Correia, A.; Spindler, P.; Pezzotta, G.; Rossi, M. PSS design considering feedback from the entire product–service lifecycle and social media. *Procedia CIRP* **2016**, *47*, 156–161. [\[CrossRef\]](#)
5. Montoyo, A.; Martínez-Barco, P.; Balahur, A. Subjectivity and sentiment analysis: An overview of the current state of the area and envisaged developments. *Decis. Support Syst.* **2012**, *53*, 675–679. [\[CrossRef\]](#)
6. Serrano-Guerrero, J.; Olivas, J.; Romero, F.; Herrera-Viedma, E. Sentiment analysis: A review and comparative analysis of web services. *Inf. Sci.* **2015**, *311*, 18–38. [\[CrossRef\]](#)
7. Jin, J.; Liu, Y.; Ji, P.; Liu, H. Understanding big consumer opinion data for market-driven product design. *Int. J. Prod. Res.* **2016**, *54*, 3019–3041. [\[CrossRef\]](#)
8. Ireland, R.; Liu, A. Application of data analytics for product design: Sentiment analysis of online product reviews. *CIRP J. Manuf. Sci. Technol.* **2018**, *23*, 128–144. [\[CrossRef\]](#)
9. Yang, C.; Wu, L.; Tan, K.; Yu, C.; Zhou, Y.; Tao, Y.; Song, Y. Online User Review Analysis for Product Evaluation and Improvement. *J. Theor. Appl. Electron. Commer. Res.* **2021**, *16*, 1598–1611. [\[CrossRef\]](#)
10. Kpiebaareh, M.Y.; Wu, W.-P.; Agyemang, B.; Haruna, C.R.; Lawrence, T. A Generic Graph-Based Method for Flexible Aspect-Opinion Analysis of Complex Product Customer Feedback. *Information* **2022**, *13*, 118. [\[CrossRef\]](#)
11. Sarmast, Z.; Shokouhyar, S.; Ghanadpour, S.; Shokoohyar, S. Unravelling the potential of social media data analysis to improve the warranty service operation. *Ind. Manag. Data Syst.* **2023**, *123*, 1281–1309. [\[CrossRef\]](#)
12. Li, Y.-H.; Yue, S.-T.; Zheng, J.; Wang, W. Customer-oriented product design: An integrated decision framework with sentiment analysis and optimisation model. *J. Control Decis.* **2024**, *11*, 165–179. [\[CrossRef\]](#)
13. Vadla, M.K.S.; Suresh, M.A.; Viswanathan, V.K. Enhancing Product Design through AI-Driven Sentiment Analysis of Amazon Reviews Using BERT. *Algorithms* **2024**, *17*, 59. [\[CrossRef\]](#)
14. Zou, H.; Wang, Y. A novel automated framework for fine-grained sentiment analysis of application reviews using deep neural networks. *Autom. Softw. Eng.* **2024**, *31*, 43. [\[CrossRef\]](#)
15. DIVERSITY Project, Horizon 2020 Grant Agreement No. 636692. 2016. Available online: <https://cordis.europa.eu/project/id/636692> (accessed on 11 May 2026).
16. Neves-Silva, R.; Gamito, M.; Pina, P.; Campos, A.R. Modelling influence and reach in sentiment analysis. *Procedia CIRP* **2016**, *47*, 48–53. [\[CrossRef\]](#)
17. Pina, P.; Góis, G.; Mano, F.; Silva, F.; Neves-Silva, R. A sentiment analysis solution for the re-design of product–service systems. In *2017 International Conference on Engineering, Technology and Innovation (ICE/ITMC)*; IEEE: New York, NY, USA, 2017.
18. Neves-Silva, R.; Pina, P. OpinionSim: Modelling and simulating stakeholders’ feedback to product–service systems. In *2017 International Conference on Engineering, Technology and Innovation (ICE/ITMC)*; IEEE: New York, NY, USA, 2017.
19. Pezzotta, G.; Pirola, F.; Sala, R.; Margarito, A.; Pina, P.; Neves-Silva, R. Identifying New PSS Concepts: The Product-Service Concept Tree. In *Enterprise Interoperability: Smart Services and Business Impact of Enterprise Interoperability*; Zelm, M., Jaekel, F.-W., Doumeingts, G., Wollschlaeger, M., Eds.; ISTE/Wiley: Hoboken, NJ, USA, 2018.
20. Wahyudi, R.D.; Singgih, M.L.; Suef, M. Investigation of Product–Service System Components as Control Points for Value Creation and Development Process. *Sustainability* **2022**, *14*, 16216. [\[CrossRef\]](#)
21. Langley, D.J. Digital Product-Service Systems: The Role of Data in the Transition to Servitization Business Models. *Sustainability* **2022**, *14*, 1303. [\[CrossRef\]](#)
22. Wu, D.; Pi, Y. Digital technologies and product–service systems: A synergistic approach for manufacturing firms under a circular economy. *J. Digit. Econ.* **2023**, *2*, 37–49. [\[CrossRef\]](#)
23. Li, H.; Zhao, J.; Cao, Y.; Su, L.; Zhao, Z.; Zhang, Y. Servitization and product service system: A literature review on value creation. *Technol. Forecast. Soc. Change* **2024**, *208*, 123724. [\[CrossRef\]](#)
24. Davoodi, L.; Mezei, J.; Heikkilä, M. Aspect-based sentiment classification of user reviews to understand customer satisfaction of e-commerce platforms. *Electron. Commer. Res.* **2026**, *26*, 1417–1459. [\[CrossRef\]](#)
25. Ganesan, P.; Garcia-Molina, H.; Widom, J. Exploiting hierarchical domain structure to compute similarity. *ACM Trans. Inf. Syst.* **2003**, *21*, 64–93. [\[CrossRef\]](#)
26. Krugmann, J.O.; Hartmann, J. Sentiment Analysis in the Age of Generative AI. *Cust. Needs Solut.* **2024**, *11*, 3. [\[CrossRef\]](#)
27. Harper, G. AI and the navigation of irony. *Creat. Ind. J.* **2025**, *18*, 125–126. [\[CrossRef\]](#)

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.