Proceedings of CONF-MLA 2026 Symposium: Learning and Decision Making in Multi Agent Software Systems
DOI: 10.54254/2755-2721/2026.BA38747

Research on the Diagnosis and Dynamic Intervention of Interaction Design Efficiency Bottlenecks Based on User Experience Metrics

Jinxuan Cheng

Ulster College, Shaanxi University of Science & Technology, Xi'an, China
3612591849@qq.com

Abstract. A delayed review tells a design manager that work has stalled, but says little about why. The same difficulty arises with repeated requirement changes and clarification requests from developers after delivery. This paper uses user experience measurement to investigate such constraints during interaction design. Interview material, analyzed through grounded theory, yields four bottleneck categories: cognitive, process-related, collaborative, and tool-related. Observable indicators are assigned to these categories, and constraint theory informs the selection of responses. The resulting MLIE model repeats monitoring, identification, intervention, and evaluation as conditions change. Its application in an internet company's design team covered three intervention rounds. The procedure gives managers a way to examine evidence for a suspected constraint and then assess the response, reducing reliance on experience alone.

Keywords: Interaction Design, Design Efficiency, User Experience Measurement, Bottleneck Diagnosis, Dynamic Intervention

1. Introduction

1.1. Research background

For digital products used in everyday tasks, the effort required to complete an interaction affects competition between products. Users expect easier, quicker operation. Product teams consequently face pressure over both the form of an interaction and the time it demands.

A separate problem occurs inside the team producing the design. A decision can remain open through several reviews, or a requirement change can send completed work back for revision. Development coordination adds further delays. Experience and selected Agile practices are common managerial responses, yet they do not provide a consistent diagnosis or a rule for choosing a remedy. The concern here is how to identify the constraint while design work is still in progress.

UX measurement offers methods that may serve this purpose. In Xia Tiansheng et al.'s account, physiological activity provides evidence of emotional and cognitive processing that subjective reports alone cannot supply [1]. Cognitive neuroscience thus broadens measurement beyond product acceptance, with applications extending into design and manufacturing.

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2755-2721/2026.BA38747

The finished product remains the usual object of UX assessment. Far less attention has gone to what these measurements might reveal about the process of producing it. This paper examines that use of measurement: informing a manager's decision when design efficiency becomes a problem.

1.2. Research significance

An assessment toolkit is developed here together with intervention options for use during design activity. The practical question is which observations reveal a restriction on progress. Comparing a manager's interpretation with recorded measurements provides an evidential basis for choosing how to respond.

1.3. Current state of domestic and international research

Google's HEART framework translates product goals into metrics for pleasure, engagement, acceptance, retention rate, and task completion rate. Other work groups interaction-efficiency problems by information architecture, task flow, system dynamics, and physical interface, measuring inconsistency, complexity, and inefficiency. The breadth of UX assessment is apparent in a systematic mapping study: from 9,107 publications it selected 189 key studies, identifying 100 dimensions and 166 attributes. Efficiency, satisfaction, effectiveness, ease of learning, and memorability were the most frequent. Individual interface settings require more specific assessments. For an interface aesthetics score, Zhou Yan and Wang He combined aesthetic metrics and grey relational analysis [2]. Tao Yiwen and Chen Bingfa assessed immersive interfaces with the Fuzzy Analytic Hierarchy Process (FAHP), covering usability, spatial perception, fluency, and expressiveness [3]. Cao Tingting et al.'s intelligent cockpit touch-interaction model instead weights six dimensions with the Analytic Hierarchy Process (AHP): visual coherence, operational rationality, human-machine convenience, information readability, functional completeness, and safety [4].

Some observations can be collected without asking a user to describe the experience. EEG and functional near-infrared spectroscopy record brain activity during interaction in the methods discussed by Xia Tiansheng et al. [1]. Eye tracking was used by Tian Chenzhi et al. to evaluate human-machine interaction effectiveness in command and control systems [5]. Pupil diameter, eye jumps, and fixation points formed part of their ocular evidence.

Interpreting UX evidence also requires an account of the processes behind it. He Xiaomei and Li Jianyu examine usability and user experience through Cognitive Load Theory, drawing design strategies from their relationship with cognitive load [6]. In a qualitative study, Tang Yuanyuan used three-level grounded theory coding to derive basic functionality, cultural experience, and emotional connection as categories of user requirements. The KANO model was then used to classify those requirements [7].

Much of this literature concerns product assessment; measurements seldom act as signals of difficulty in the work that produces a design. Outcome assessment also leaves the subsequent steps underdeveloped: deciding what to change and checking whether the change has worked. Accounts of bottlenecks commonly look back on an event, although managers need a diagnosis while it is occurring. Grounded theory has been applied to user needs [8]. Its application to the classification and diagnosis of constraints in interaction design work remains limited.

DOI: 10.54254/2755-2721/2026.BA38747

1.4. Research content and technical approach

The investigation brings together qualitative accounts, measurements of design work, and repeated interventions.

Here, measurements are asked to explain where the design process is constrained. Assessing whether a completed product is satisfactory serves a different purpose. Locating the stage that restricts progress lets the team use the same measurement activity to guide improvement during ongoing work.

Interviews with 20 senior designers and managers supply the material for the classification. Grounded theory analysis proceeds through three coding levels, yielding cognitive, process-related, collaborative, and tool-related constraints. These empirically derived categories give subsequent diagnosis a basis beyond the judgment of an individual manager or a summary of a particular case.

Constraint Theory and Feedback Control Theory inform the MLIE Dynamic Intervention Model. Monitoring is followed by target identification, intervention, and evaluation; the evaluation returns the team to monitoring. A constraint can therefore be reassessed after action has been taken, with the next response adjusted to the changed design process.

The qualitative analysis establishes what kinds of bottleneck occur. Constraint theory addresses how they restrict the wider system. These complementary roles are examined in a SaaS product design team at an internet company over three intervention cycles.

2. Relevant theoretical foundations

2.1. The core concept of interaction design effectiveness

Interaction design efficiency describes what a team produces with the time, skills, tools, and coordination available to it. Clear requirements and effective coordination affect that output, as do the team's expertise and the tools it uses. The interface is part of this relationship because it mediates human–machine interaction and influences how efficiently the system can be used. An assessment of interaction effectiveness can also show where development time and cost may be reduced.

2.2. User experience measurement theory

User experience includes a person's perceptions and responses to a product, system, or service, including expectations before use. The HEART framework divides UX measurement into five dimensions. Joyfulness concerns feelings and attitudes; Engagement concerns the frequency and depth of use; Acceptance records the acquisition of new users; Retention Rate concerns continued engagement; and Task Completion Rate concerns how efficiently and effectively core tasks are completed.

Questionnaires remain the most widely used way of assessing UX, although observation, expert assessment, and physiological measurement are also available. What a questionnaire can capture depends on what users remember, are willing to report, and have consciously noticed. Recall and social desirability therefore affect the answers. Physiological measurements provide complementary evidence: Xia Tiansheng et al. describe objective and reproducible recordings of brain activity during interface use [1], while Tian Chenzhi et al. use eye movement to assess interaction efficiency [5]. Their available measures include where fixations cluster, the time to first fixation, fixation counts, eye jumps, scan path length, blink rate, and pupil diameter.

DOI: 10.54254/2755-2721/2026.BA38747

2.3. Cognitive load theory

Processing information and holding it in memory both use cognitive resources, whose capacity is limited. This is the premise of Cognitive Load Theory. He Xiaomei and Li Jianyu explain cognition and learning in terms of resource allocation and schema formation. A task that demands more resources than are available places a load beyond working memory capacity.

For users, keeping demands at an appropriate level supports efficient operation and a smooth, positive experience. The capacity limit therefore has implications for both interface usability and UX.

The same problem can arise while a designer is working on the interface. Excessive cognitive demands may slow the decision and reduce the quality of the resulting design.

2.4. Constraint theory and bottleneck identification

Constraint Theory asks which bottleneck is limiting the system as a whole. Its five steps begin with finding that constraint and deciding how to use it. The remaining processes are then organized around this decision, followed by improvement of the constraint itself. Once the constraint is resolved, the sequence begins again. Applied to design work, this means locating the stage that holds back overall output and judging proposed improvements by what they do for that stage.

2.5. Grounded theory methodology

Grounded Theory builds an explanation from systematically collected and analyzed empirical material. Analysis moves from open coding to axial coding and then selective coding, with earlier interpretations revisited along the way. Categories are refined through this repeated work; they may describe patterns that participants have not explicitly identified themselves.

In Tang Yuanyuan's study of cultural and tourism apps, the three coding levels produced the categories of basic functionality, cultural experience, and emotional connection [7]. This example provides a methodological reference for the present task of deriving categories of design-efficiency bottlenecks from qualitative accounts.

3. Types and identification of interaction design efficiency bottlenecks

3.1. Definition of design efficiency bottlenecks

In this study, a design-efficiency bottleneck is a restriction that affects output across the process rather than a single isolated event. It should recur, involve more than one task or role, and be open to measurement and improvement. A difficulty that has little effect beyond one occasion does not satisfy these conditions.

3.2. Classification of efficacy bottlenecks based on grounded theory

The interview material was organized into four categories. Twenty experienced designers and design managers provided the accounts (20 participants in total). Open coding first captured the relevant incidents, axial coding related them to one another, and selective coding brought the categories together as an explanation of constraints on design efficiency.

Cognitive bottlenecks concern how designers interpret information and make choices. A designer may repeatedly weigh proposals without settling on one; misunderstanding a requirement or having

DOI: 10.54254/2755-2721/2026.BA38747

too much information to process can also hold up the work. The common difficulty in these cases lies in interpretation, information processing, or choice.

The workflow itself accounts for the process-related category. A submitted proposal may wait a long time for approval. Requirements that keep changing can send work back for revision, yet repeated revisions may still fail to resolve the issue.

Collaborative bottlenecks arise between designers, product managers, and developers whose assumptions or ways of communicating differ. The design and its development can consequently become misaligned. Feedback presents a related difficulty when several reviewers contribute comments that cannot easily be turned into an agreed set of changes. In both situations, coordinating work across roles is the constraint.

The tool-related category covers missing tool functions, obstacles to collecting measurement data, and limited ability to interpret the data once collected. Each can restrict progress, even if the other parts of the measurement process are available.

3.3. Bottleneck identification methodology framework

A short screening exercise determines where to investigate in detail. The suspected constraints are then examined through measurement.

Designers, design managers, and product or development representatives complete the screening questionnaire independently. It asks about the four categories and their manifestations in Chapter 3. For each, a respondent rates frequency on the five-point scale Never/Very Rare/Sometimes/Often/Always and impact on the five-point scale No Impact/Small/Moderate/Large/Severe. Multiplying the two ratings gives the Bottleneck Severity Index. The team average is calculated for each category, and the two highest averages determine which categories receive detailed investigation. This selection keeps the subsequent measurement work focused.

Process logs and measurements are used to test the screening result. The mapping in Table 1 of Chapter 4 specifies which measures to examine for a suspected bottleneck, including review duration, solution modification counts, and development clarification issues. A predefined threshold or historical baseline supplies the comparison for each value. Several related indicators must be outside their normal ranges at the same time before the diagnosis is confirmed. Thus, the questionnaire selects the focus of investigation, while the measurements establish whether that initial judgment is supported.

4. Performance bottleneck-oriented user experience measurement mapping mechanism

4.1. Repositioning of measurement indicators

Measurements collected in real time or near real time can describe the condition of a design process as well as a product's user experience. For diagnosis, the question is which stage is having difficulty. The observations are interpreted here as indicators of that condition.

4.2. Mapping relationship between bottleneck types and performance metrics

Figure 1 links bottleneck categories with performance measures using the classification and measurement theory described above. The mapping also draws on the engineering design efficiency indicators in the China Internet Association's group standard, Enterprise Digital Design Engineering

DOI: 10.54254/2755-2721/2026.BA38747

Maturity Standard and Assessment Methodology (T/ISC 0071-2024), and practical indicators from the Youzan Technology Efficiency Improvement Team.



Figure 1. Mapping relationship between bottleneck types and performance metrics

Flow Efficiency is the proportion of total elapsed time for which a flow item is actively being worked on. A low proportion signals a possible process bottleneck. Style Unbinding Rate and Component Unbinding Rate describe the effective use of design assets and come from the T/ISC 0071-2024 Group Standard. The Youzan Technology Efficiency Improvement Team proposed the R&D Concentration Metric, calculated as (Project Workload Man-Days / (R&D Cycle × Number of Participating Personnel)) × 100%.

4.3. Time-series data acquisition scheme

Data are recorded at five points in the design lifecycle. Initiation records the task start time and initial conditions. Solution generation records how many proposals are produced and how long they take. Review records its duration and the number of comments. Development and delivery record compliance with delivery specifications and clarification requests. User validation supplies the user test results.

The collection toolchain draws on records already generated during work. The Figma API can retrieve design version history. Jira or TAPD logs provide task cycle information, and communication records from DingTalk or Feishu provide collaboration data.

DOI: 10.54254/2755-2721/2026.BA38747

4.4. Bottleneck diagnosis determination rules

An abnormal value in one indicator triggers a warning. Confirmation requires the combined evidence of two or three related indicators. For example, a long review cycle is classified as a process bottleneck when the time taken to obtain approval exceeds its threshold and review meetings have also increased abnormally. The extra evidence helps distinguish a persistent constraint from a single unusual measurement.

Thresholds also need to account for fluctuations caused by increasing system complexity. Historical variation provides a baseline for a dynamic threshold. Under the proposed rule, an alert is triggered when a value exceeds the historical mean by 1.5 standard deviations.

5. Construction of a dynamic intervention model based on diagnostic results

5.1. Overall framework of the MLIE model

MLIE names four linked activities: Monitoring, Locating, Intervening, and Evaluating. Together they describe how a team moves from noticing an unusual measurement to checking whether its response has helped.

The Monitoring Layer keeps the current values of the indicators in Chapter 4 available on a dashboard. Automated tools collect the values as work proceeds, so that a change in the process can be seen without waiting for a later review.

The Localization Layer compares an abnormal value with the Bottleneck-Metric Mapping Table and the diagnostic rules. For example, version changes that exceed their threshold together with a longer decision time point toward hesitation in decision making, which is treated as a possible cognitive bottleneck.

The Intervention Layer uses both the diagnosed type and its severity to select a response. Section 5.2 describes the conditions and procedures for the four categories; its options include the Design Decision Matrix, Milestone Brief Review, and Joint Design and Development Review Meeting. Selection can still be necessary within a category. When several strategies fit, the type priorities and intensity grades in Section 5.3 guide the choice. When several bottlenecks coexist, the Influence Matrix Method determines which receives priority.

After one to two iteration cycles, the Evaluation Layer collects the indicators again to check the response. A strategy that has worked is kept for later use. If it has failed, the team can reconsider the diagnosis or try a different strategy.

The findings from this check become input to monitoring. Each pass through the four layers therefore ends with an assessment of the response and begins another examination of the constraints, allowing diagnosis and intervention to adjust as those constraints change.

5.2. Differential intervention strategy repository

For cognitive bottlenecks, the available responses address the choice itself. The design decision matrix compares options by scoring them on several dimensions. A decision time limit bounds the period of deliberation, while a list of design principles gives designers a reference for making routine choices more quickly.

A process intervention may change how reviews are arranged. Replacing a large session with smaller milestone reviews reduces the length of each meeting. Where repeated requirement changes

DOI: 10.54254/2755-2721/2026.BA38747

cause the difficulty, impact assessment sets criteria for accepting a change. A further option is the design sprint, which takes the work from problem identification to a prototype within five days.

Collaboration can be addressed at the points where roles meet. Joint reviews involve development from the early definition of the design; delivery specifications and checklists standardize the handover. Designers can also rotate into development teams to understand that work and the concerns of their colleagues.

The tool-related responses depend on where measurement is difficult. Automated dashboards lessen the work of manual data collection. Training helps with interpretation, and AI-powered analytical tools can generate recommendations for diagnosis.

5.3. Matching and selection mechanism for intervention strategies

Matching a strategy to the bottleneck type leaves a second decision: how much to change. A self-check before review is a mild response suited to a small abnormality with little effect on the team. If the constraint continues across iterations and clearly restricts output, a moderate response may be appropriate, such as replacing large review sessions with milestone reviews. Structural problems call for deeper changes, including revision of delivery specifications and checklists.

Where several bottlenecks compete for attention, the Influence Matrix Method compares their reach with the cost of improving them. A bottleneck affecting more of the work at a lower improvement cost takes priority.

5.4. Re-evaluation of intervention effectiveness and model adaptation

Reassessment uses measurements taken one to two iteration cycles after the intervention. A return to the normal range counts as significant improvement and places the strategy in the best practice repository. Improvement that stops short of this point calls for adjusted intensity or combined strategies. Unchanged or worse values require another look at the diagnosis, since the bottleneck may have been identified incorrectly.

6. Empirical study: application and validation of the dynamic intervention model

6.1. Research subjects and design

The case was an internet company's team responsible for interaction design for a mid-to-large-scale SaaS product. The team included interaction designers, a visual designer, and a design manager, with members working on different business modules. During the product's growth phase, a new iteration was released on average every 4 to 6 weeks. The researcher followed the team through three action-research cycles covering diagnosis, intervention, and evaluation. The observation included three design versions and lasted approximately 18 weeks.

6.2. Baseline establishment and initial diagnosis

The baseline was assembled from records of the first two software versions: design version histories, project management lifecycle records, and communications within the team.

Review duration stood out in these records. The interval from proposal submission to approval averaged 8.3 days; the baseline account described this as 5 days beyond the predetermined timeline and an approximately 66% delay. Developers requested clarification an average of 7.2 times after delivery, compared with the threshold of 5. The mean number of proposal revisions, 4.1, was also

DOI: 10.54254/2755-2721/2026.BA38747

slightly above its threshold. Taken together, these observations indicated process-related and collaborative bottlenecks. The revision count additionally suggested a cognitive constraint.

Each team member then took part in an individual semi-structured interview lasting approximately 45 minutes. The researchers asked which processes had most affected efficiency over the previous two versions and where work had become constrained. These accounts were used to investigate the anomalies in the records.

Open coding yielded 28 initial concepts from sentence-by-sentence reading of the transcripts. The statement, "During the review meeting, everyone was talking over each other, and in the end, no one could reach a final decision," received the code Difficulty in converging opinions. High cost of post-delivery clarification captured the account, "After the design draft was delivered, the development team asked six or seven additional questions". The account, "A single requirement underwent three rounds of revision and continued to change," was coded as Repeatedly revised requirements.

Comparison in axial coding organized the initial concepts into six subcategories. Cross-Role Collaboration Barriers grouped difficulty reaching consensus, inconsistent understanding across roles, and high post-delivery clarification costs. Long reviews, repeated requirement modifications, and frequent rework were grouped as Process Structural Issues. Cognitive Processing Obstacles covered repeated trade-offs between solutions and deviations in requirement understanding.

Selective coding brought the subcategories under The Constituent Elements of Design Activity Systems as the core dimension. For the case team, this organization produced three bottleneck types.

Interview accounts of prolonged reviews, coordination difficulties, and ambiguous requirement interpretation agreed with the baseline measurements. On this combined evidence, process-related and collaborative constraints took priority over cognitive constraints. That ranking determined the sequence of the three intervention rounds.

6.3. Implementation of the three-round intervention cycle

Review delay was addressed first. Under P1, the large final-draft session became separate concept, solution, and detail reviews, with each limited to two hours. P2 added requirement change impact assessment. Implementation took approximately six weeks per iteration.

Collaboration became the focus after the first-round assessment. S1 brought one or two development engineers into each solution review through joint design-development review. S2 added a delivery checklist containing 12 inspection items. The team applied both changes over one iteration version.

For the third round, the earlier responses remained in place. C1 added a design decision matrix for the cognitive constraints that remained, scoring complex choices in a structured way over one iteration version.

6.4. Comprehensive assessment of intervention effectiveness

The before-intervention measurements and those collected after the third round appear together in Figure 2.

DOI: 10.54254/2755-2721/2026.BA38747

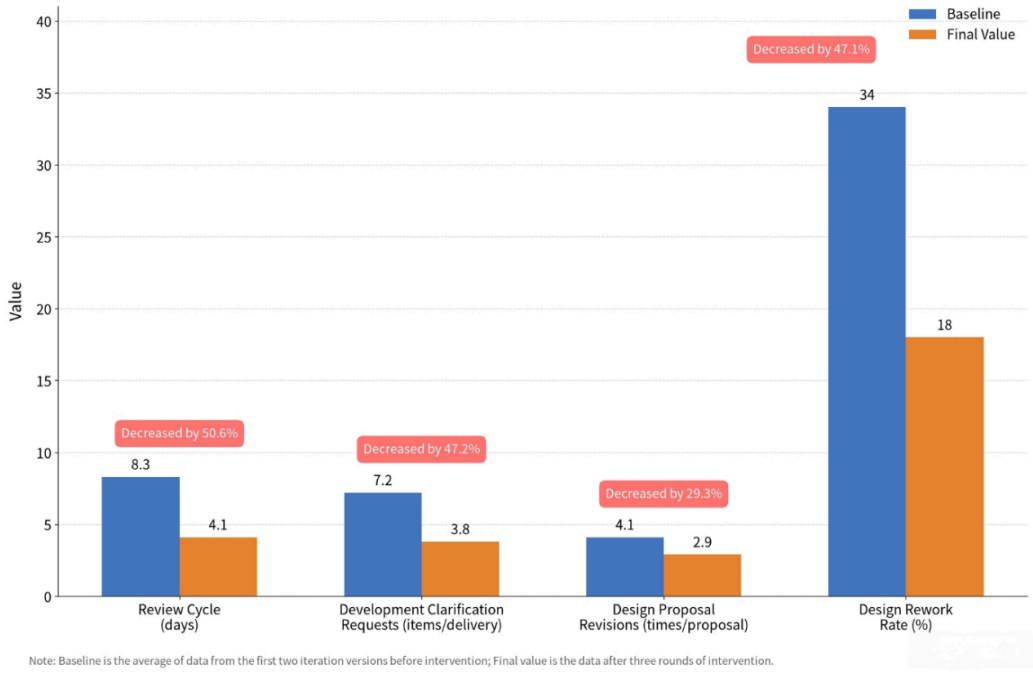


Figure 2. Comparison of key performance indicators before and after the intervention

After the three rounds, average review duration was 4.1 days, down from 8.3 days, a 50.6% reduction. Rework declined to 18% from 34%, equivalent to a reduction of 47.1%. Clarification requests from development fell by 47.2%; proposal revisions fell by 29.3%. In the team studied, these changes support the practical use of MLIE to diagnose constraints and select responses.

7. Conclusion and future outlook

The investigation produced a classification of bottlenecks and tested its use in repeated interventions. The diagnostic findings need to be read within the limits of the case evidence.

Cognitive, process-related, collaborative, and tool-related constraints form the classification, although their relative weight may differ with the team and product. Grounded theory makes it possible to derive the categories from accounts of design work. The framework's present basis is interview data from a single team; assessment of its statistical robustness requires a larger questionnaire study.

The mapping gives measurements a diagnostic role by linking indicators to plausible constraints in the work process. Efficiency problems can then be measured and traced. Data usually collected for evaluation can inform a response while design is still under way.

In the case, review duration decreased by approximately 50% and rework by approximately 47%. Monitoring, identification, intervention, and evaluation were repeated as the constraints changed. The evidence comes from one team observed for 18 weeks. More teams and a longer follow-up are required to establish whether the effects persist and apply elsewhere.

A next question concerns the constraint that remains after another has been resolved. Bottlenecks may move to a different stage or produce cascading effects. Tang Yuanyuan's combination of grounded theory and the KANO model suggests a direction for joining classification to quantitative prioritization in further diagnostic work.

DOI: 10.54254/2755-2721/2026.BA38747

Practical development is planned around a lightweight design performance assessment tool and a larger sample for comparisons across teams. Enterprise SaaS design work supplies the setting for MLIE, whose monitoring and evaluation layers automate collection and processing through Figma API and TAPD interfaces. A further possibility is to train a machine learning model on historical measurements to predict bottleneck risk. Earlier warnings could then complement diagnosis after a problem occurs, reducing manual collection and giving managers more timely, objective evidence.

References

[1] Xia, T., Li, S., & Ouyang, J. (2022). Application of cognitive neuroscience techniques in user experience measurement need. *Dress Decorat- ive Elements, 2022*(3). DOI: 10.16272/j.cnki.cn11-1392/j.2022.03.022. DOI: 10.19554/j.cnki.1001-3563.2023.06.002.

[2] Zhou, Y., & Wang, H. (2022). Research on an APP interface evaluation method based on beauty measurement. *Design Art Research, 2022*(1). DOI: 10.3963/j.issn.2095-0705.2022.01.003.

[3] Tao, Y., & Chen, B. (2017). Research on evaluation methods for immersive interface interaction design. *Mechanical Manufacturing and Automation, 2017*(5), 169–173. DOI: 10.19344/j.cnki.issn1671-5276.2017.05.044.

[4] Cao, T., Zhang, H., & Yuan, Q. (2025). An HMI touch interaction evaluation method for improving user experience. *Automotive Engineering, 47*(11) (Page numbers to be added).

[5] Tian, C., Song, M., Wang, W., et al. (2024). A review of research on evaluating human–machine interaction efficiency in digital interfaces using eye-tracking technology. *Optics & Electronics – Control, 31*(3), 8–16, 47.

[6] He, X., & Li, J. (2020). A study on the causal analysis and design strategies of cognitive load in human-computer interaction. *Packaging Engineering, 41*(10), 24–30. DOI: 10.19554/j.cnki.1001-3563.2020.10.005.

[7] Tang, Y. (2025). A design study on cultural and tourism apps based on grounded theory and the KANO model. *Design, 38*(24), 120–123.

[8] Singh, S., Bartolo, K., & Satchell, C. (2005). Grounded theory and user requirements: A challenge for qualitative research. *Australasian Journal of Information Systems, 12*(2).