User Frictions in Process Intelligence Dashboards: Implications for AI Assistants

Candan Cetin,[1,2] Teodora Lata[2]

[1] Technical University of Munich, Munich, Germany
candan.cetin@tum.de
[2] Celonis SE, Munich, Germany
candan.cetin@tum.de, t.lata@celonis.com

Process intelligence platforms increasingly support operational decisions, but their realized value is also related to the degree that users can interpret dashboards and act on insights effectively. We derived user-friction themes from the process intelligence and dashboard literature, operationalized them into a survey, and analyzed responses from 58 active and experienced dashboard users with hypothesis-based OLS regressions using robust standard errors and false discovery rate correction. The results showed that decision-related outcomes were most strongly associated with visual interpretability, question-to-query translation, trust in data credibility, and metric actionability. The most robust supported relationships indicated that clear layouts and labels, effective translation of analytical intent into filters and views, complete and up-to-date data, and decision-relevant metrics aligned with higher ease of use, usefulness, confidence, and decision effectiveness. Based on these findings, we discuss how AI assistants should complement dashboards by reducing friction in understanding, querying, verifying, and acting on process intelligence data.

DOI
https://doi.org/10.18690/um.fov.4.2026.46

ISBN
978-961-299-152-4

Keywords:
process intelligence, dashboards, AI assistants, user experience, information systems

University of Maribor Press

1 Introduction

Process Intelligence (PI) solutions have evolved from niche analytical tools into a core layer of digital operations. By extracting event logs from any source system, PI provides organizations with fact-based visibility into their actual workflows (van der Aalst, 2016; van der Aalst et al., 2012). As the PI software market experiences sustained double-digit growth (Everest Global, Inc., 2025), PI platforms become a part of the companies` digital core.

However, software investments do not automatically yield operational returns. In practice, value realization hinges on end-user adoption; humans must interact with dashboards, interpret results, and act. In this context, end-users encounter challenges related to complex layouts, opaque data origins, and the technical expertise required to extract answers.

At the same time, artificial intelligence (AI) actively reshapes the PI user landscape. Vendors are embedding conversational copilots designed to lower skill barriers and shorten the time-to-value. While current PI literature focuses heavily on backend algorithmic discovery and data integration, specifically the end-user experience aspect of dashboard consumption remains largely underexplored (Ammann et al., 2025; Ankomah, 2025, and Meduri et al., 2021). Consequently, empirical evidence is scarce regarding concrete challenges end-users face and how vendors should design AI assistants` roadmaps to alleviate those frictions.

Accordingly, this study asks: Which dashboard consumption frictions should AI assistants target in process intelligence platforms? Although this study did not test a specific AI assistant intervention, it instead identified empirically supported dashboard frictions that AI assistants may target in the future.

2 Background and Related Work

To understand the landscape of user challenges, we conducted a systematic literature review and the Gioia methodology to synthesize findings into Second-order Themes (SoTs) (Gioia, 2013).

The review revealed 24 SoTs and highlighted 10 of them, which we used as the baseline of our research. To summarize, the top 4 most frequently mentioned themes were:

- Visual Comprehension,
- Trust Issues,
- Expertise Requirement,
- Lack of Actionability.

The most consistently reported problem is the comprehension effort required to interpret highly dense, complex visual representations. Misaligned charts and poor labelling reduce the user`s ability to recognize trends or inefficiencies (Atighehchian et al., 2023; Rehse et al., 2022; Zuidema-Tempel et al., 2022). Secondly, users frequently doubt the accuracy and completeness of dashboard data. A lack of transparency regarding data processing and timestamps leads users to manually verify raw data, undermining the dashboard`s value (Alhamadi et al., 2022; Chinnathai & Alkan, 2023; Simic et al., 2023). Thirdly, dashboards often require users to possess specialized technical skills or familiarity with query languages in order to translate analytical business questions into the appropriate interface filters (Barbieri et al., 2023; Ullrich et al.; Zimmermann, 2023). Lastly, dashboards frequently display retrospective insights without supporting timely, decision-relevant follow-up, forcing users to rely on manual workflows to execute tasks (Morelli et al., 2024; Rehse et al., 2022; Simic et al., 2023; Zimmermann, 2023).

Most of the PI vendors have introduced AI and Large Language Models (LLMs) into their ecosystems (Carpenter, 2025; Everest Global, Inc., 2025; Wilson, 2024). Current AI assistant capabilities aim to democratize process data (Boareto et al., 2025). LLMs provide natural-language analytics and situation-aware explainability (Fahland et al., 2025).

3 Methodology

3.1 Conceptualization and Operationalization

To systematically translate qualitative literature into testable quantitative metrics, we used a traceable approach. To illustrate this approach, we observed the development of Hypothesis 1 (H01) (Eisenhardt, 1989). We started by extracting raw literature excerpts from articles in our sample space. H01 was traced back to the excerpts of works of Alhamadi et al. (2022)[1] and Zuidema-Tempel et al. (2022)[2]. Then, we translated these raw excerpts into First-Order Concepts (FoCs) which were problem statements such as "Some users verify data manually, indicating distrust or lack of confidence in the dashboard." and "Inaccurate or incomplete data undermines user confidence in dashboard outputs." We grouped together related FoCs under the theoretical Second-Order Themes (SoTs). In our example above, these two FoCs fall under the "Visual Comprehension" SoT. To create a falsifiable claim, we refined this theme into a specific, design-actionable sub-theme (Visual Layout & Label Clarity) and linked it to an Information Systems (IS) outcome. This generated "H01: Clear layouts and labels increase perceived ease of use." After the hypothesis formation, we measured the variables using a 6-point Likert scale in the survey by user responses. The predictor was measured using the survey item: "The visual layout, labels, and axes make data easy to understand" and the outcome was measured using the Technology Acceptance Model (TAM) item (Davis, 1989): "Learning to use this dashboard is easy for me." (PEOU).

3.2 Survey Design and Sample

We operationalized the literature-based hypotheses into a survey measuring user frictions alongside established IS outcomes, specifically the TAM constructs (Davis, 1989; Ives et al., 1983; Wixom & Todd, 2005). We tried to operationalize each

[1] When users are not sure of the accuracy of the information presented, they usually question the origin of the data. If the dashboard does not capture the depth of the data that satisfies users' needs, users tend to check the raw data sets frequently. Developers try to prevent this workaround as they believe it undermines the value of a data overview dashboard Alhamadi et al. (2022).

[2] Zuidema-Tempel et al. (2022) reported that all interviewed experts identified data quality as a major bottleneck, especially because of incorrect data formats and incomplete data entry by employees. One expert explained that "it was difficult to get the data out of the systems," which reduced trust in the availability of good-quality data. The same study further noted that responses such as "we have tried that before but it did not work" were common, and that some process mining projects were terminated because the required data were unavailable.

construct with more than one survey item without excessively extending the survey's length. All attitudinal items were measured on a 6-point Likert scale ranging from 1 = strongly disagree to 6 = strongly agree. Closed-ended survey questions captured the prevalence and perceived severity of the literature-based dashboard-related issues. We also added open-ended questions to each of the three predictor-measuring survey blocks. This gave respondents space to report issues not covered by the predefined survey items and added qualitative richness to the analysis. Specifically, 13 respondents (22.4%) gave input to optional open-ended VSC_FREE, 5 (8.6%) to TRUST_FREE, and 10 (17.2%) to ACT_MER_FREE questions shown after the predictor items blocks of the survey. These rates seem to be consistent with academic benchmarks for optional open-ended items in digital surveys, which typically range between 18%-24% for optional qualitative feedback (Asare-Marfo, 2021; Miller & Lambert, 2014).

Our sample consisted of experienced and active dashboard users with 52% of them having 3+ years experience in the PI area and 79% of them using dashboards on a weekly or daily basis. Regarding AI exposure, 34.5% reported that AI assistant functionality was available and that they had used it at least once, while 41.4% indicated that AI was available but not yet used. Respondents indicated that they need the most help with finding the right dashboard/page/view, digging into anomalies, and understanding metrics consecutively. When they were asked to rank their selected needs based on their AI assistance preference, the order that arose was first around finding the right dashboard, followed by turning insights into actions, and then digging into anomalies. Appendix Section 7.1 documents the recruitment procedure and Appendix section 7.2 summarizes the survey structure and measurement blocks.

3.3 Statistical Analysis

In the data preparation phase, before hypothesis testing, we assessed the internal consistency of multi-item survey constructs using reliability analysis (Cronbach`s alpha) (Cronbach, 1951). Constructs demonstrating acceptable internal consistency were aggregated into composite measures by averaging their items. For instance, Decision Effectiveness (outcome construct) showed satisfactory internal consistency, and, therefore, we used it as a composite of DE ("This dashboard helps me make decisions quickly when needed.") and DC ("I feel confident in the

decisions I make based on this dashboard.") items for hypothesis testing. Several constructs did not demonstrate adequate internal consistency. In particular, items within the Visual Clarity & Findability survey block (VSC1, VSC2, VSC3, VSC4) did not exhibit sufficient internal consistency and were therefore treated as conceptually singleton predictors representing their subthemes (Visual Layout & Label Clarity, Visual Time Context & Trend Readability, Visual Emphasis & Colour Encoding, Visual Glanceability & Information Hierarchy). Appendix Section 7.3 reports the reliability diagnostics and the resulting operationalization decisions for all multi-item constructs. To evaluate which dashboard features were reliably associated with user outcomes, we estimated separate OLS regression models for each of the 60 hypotheses using all completed responses ($N = 58$). We controlled for role, experience, usage frequency, and AI availability. Since the sample size is moderate, we applied the HC3 robust standard errors to obtain conservative and more reliable p-values (MacKinnon & White, 1985; White, 1980). We only considered a hypothesis supported if the three following criteria were satisfied:

- The estimated coefficient had the theoretically expected direction,
- The significance threshold of $p < .05$ was met,
- The Benjamini-Hochberg False Discovery Rate (FDR) adjustment produced $q < .10$ (Benjamini & Hochberg, 1995).

Using the conventional p-value threshold indicated that the observed effects were unlikely under the null model and were therefore less likely to reflect random sampling variation alone (Greenland et al., 2016; Wasserstein & Lazar, 2016). However, FDR adjusted values controlled for false positives arising from testing many hypotheses simultaneously (Bender & Lange, 2001, Benjamini & Hochberg, 1995).

4 Results

Before hypothesis testing, we examined mean predictor ratings to identify the lowest- and highest- rated dashboard components in our sample. Descriptively, items related to data lineage transparency, timestamp correctness, and manual effort exhibited comparatively weaker agreement. Conversely, items capturing visual layout, label clarity, and the ability to work without technical help received comparatively stronger agreement. Across all outcomes, respondents reported the

highest agreement for decision confidence and the lowest agreement for perceived usefulness for job performance. The mean differences across outcome constructs remained small in this sample. Overall, the spread of mean ratings across predictors and outcomes was modest, indicating that respondents did not report severe dissatisfaction but rather relative differences in friction levels within an otherwise positively evaluated dashboard experience.

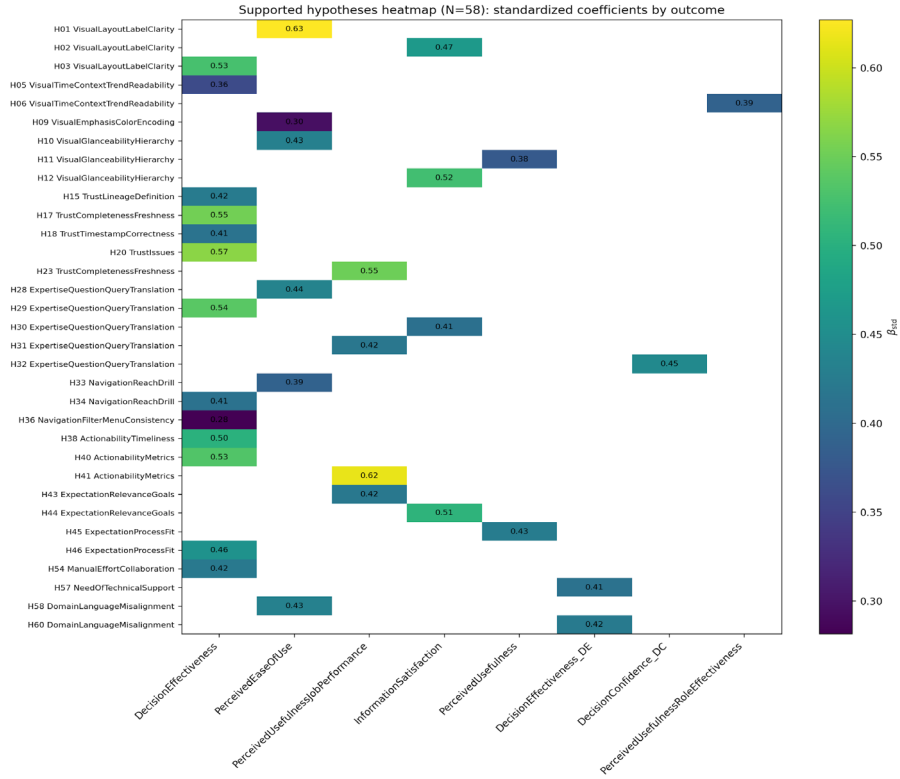

Figure 1: Accepted Hypotheses
Source: Own

Under our strict criteria (positive expected direction, $p < .05$, and FDR-adjusted $q < .10$), 33 of the 60 hypotheses were statistically supported. Figure 1 visualizes all 33 accepted hypotheses (H1, H2, H…) by mapping predictors (VisualLayoutLabelClarity,...) to outcomes (PerceivedEaseofUse, DecisionEffectiveness, PerceivedUsefulness, ...) and colouring the links according

to standardized effect size β_{std}. The results indicated that specific usability, trust, and actionability factors aligned strongly with decision-related outcomes.

Decision Effectiveness, operationalized as the average of items "This dashboard helps me make decisions quickly when needed." (DE) and "I feel confident in the decisions I make based on this dashboard." (DC), emerged as the most central outcome in our sample, with 13 supported predictors.

Table 1: Top 8 Accepted Hypotheses

ID	Hypothesis	Predictor	Outcome
H01	Clear layouts and labels increase perceived ease of use.	Visual Layout & Label Clarity	Perceived Ease of Use
H29	Translating questions into the right filters/views improves decision effectiveness.	Question-to-Query Translation	Decision Effectiveness-composite
H02	Clear layouts and labels increase information satisfaction.	Visual Layout & Label Clarity	Information Satisfaction
H23	Complete, up-to-date data increase perceived usefulness for job performance.	Data Completeness & Freshness	Perceived Usefulness (Job Performance)
H03	Clear layouts and labels improve decision effectiveness.	Visual Layout & Label Clarity	Decision Effectiveness-composite
H41	Specific and appropriately filtered metrics increase perceived usefulness for job performance.	Metric Actionability	Perceived Usefulness (Job Performance)
H20	Higher overall trust in dashboard data improves decision effectiveness.	Trust-composite (Lineage, freshness, timestamps)	Decision Effectiveness-composite
H32	Translating questions into the right filters/views increases decision confidence.	Question-to-Query Translation	Decision Confidence

Source: Own

Although we reported all 33 supported hypotheses in Figure 1, we emphasized the relationships with the smallest FDR-adjusted q-values because they remained the most robust against multiple-testing corrections and were the least likely to reflect chance findings. The eight strongest hypotheses (H01, H29, H02, H23, H03, H41, H20, H32) have the smallest FDR-adjusted q-values and therefore provided the

empirical basis for our main conclusions. Table 1 summarizes these hypotheses. Beyond statistical support (direction, $p < .05$, and $q < .10$), the top eight hypotheses also showed moderate-to-large, standardized effects and tight uncertainty bounds under HC3 robust standard errors.

The strongest result was H01, which linked Visual Layout & Label Clarity to Perceived Ease of Use ($\beta\_std = 0.627$, 95% CI [0.442, 0.813], $q = 2.09 \times 10^{-9}$). Other highly robust effects included H29, which linked Question-to-Query Translation to Decision Effectiveness ($\beta\_std = 0.539$, 95% CI [0.318, 0.759], $q = 5.3 \times 10^{-5}$), H02, which linked Visual Layout & Label Clarity to Information Satisfaction ($\beta\_std = 0.469$, 95% CI [0.270, 0.668], $q = 5.9 \times 10^{-5}$), and H23, which linked Data Completeness & Freshness to Perceived Usefulness for Job Performance ($\beta\_std = 0.551$, 95% CI [0.317, 0.785], $q = 5.9 \times 10^{-5}$). H03 further showed that Visual Layout & Label Clarity aligned with Decision Effectiveness ($\beta\_std = 0.526$, 95% CI [0.294, 0.758], $q = 1.06 \times 10^{-4}$). H41 linked Metric Actionability to Perceived Usefulness for Job Performance ($\beta\_std = 0.616$, 95% CI [0.312, 0.920], $q = 7.28 \times 10^{-4}$), H20 linked the Trust composite to Decision Effectiveness ($\beta\_std = 0.567$, 95% CI [0.283, 0.851], $q = 7.67 \times 10^{-4}$), and H32 linked Question-to-Query Translation to Decision Confidence ($\beta\_std = 0.445$, 95% CI [0.212, 0.679], $q = 1.42 \times 10^{-3}$). All eight effects had strictly positive 95% confidence intervals, which reinforced that these relationships represented robust positive associations rather than borderline estimates.

5 Discussion

5.1 Limitations

This study relies on self-reported perceptions rather than objective usage or performance measures. Respondents evaluated both predictors and outcomes within the same cross-sectional survey, which may have inflated associations because the method and response tendencies can influence all measures in a similar way. The analytical sample was moderate in size with $N = 58$ complete responses, and it skewed toward experienced users with combined builder-and-viewer roles, which may limit generalizability. Similarly, the employed recruitment approach was network-based rather than a defined sampling framework which might not produce

a holistic representative sample. Besides, several constructs were operationalized as single- item measures (singletons) such as DLM1, PRV1, and IS1 or split into single-item sub-themes after reliability checks as in the case of VSC1, VSC2, VSC3, and VSC4. These choices may have introduced measurement error or reduced coefficient stability. Finally, the regression models identified associations under controls and robust standard errors, but they did not establish causality or test the effects of a specific AI assistant intervention. Future research should validate these findings with larger, more balanced samples and with behavioural outcomes such as task accuracy, time, and interaction traces in controlled studies or field deployments.

5.2 Implications for AI Assistant Design

Under these limitations, the results still identified four friction areas that future AI assistants may target: interpretability, expertise translation, credibility, and actionability. This interpretation also aligned with the broader survey context. The sample consisted of experienced and active dashboard users, and respondents most frequently reported needing help with finding the right dashboard or view, digging into anomalies, and understanding metrics. When they ranked their preferred areas for AI support, they placed finding the right dashboard first, followed by turning insights into actions. These priorities fit well within the supported hypothesis pattern, which converged on four mechanisms: interpretability, expertise translation, credibility, and actionability.

First, the results pointed to interpretability as a foundational mechanism. H01, H02, and H03 showed that clearer layouts and labels aligned with higher perceived ease of use, higher information satisfaction, and higher decision effectiveness. These relationships suggested that respondents evaluated dashboards more positively when the interface reduced visual ambiguity at the point of consumption. This interpretation also matched the open-ended responses. Users described filters as "not seen that easily" and noted that the information hierarchy was sometimes "off," which supports the idea that visual frictions affected how efficiently they could orient themselves in a dashboard. To illustrate, we may consider an order-management dashboard as an example. In such a dashboard, users may struggle to understand whether top KPI tiles, lateness buckets, drilldown tables, and opportunity lists all reflect the same active filters, or how labels such as "Avg. Days Late" and "Open Deliveries" should be interpreted together. In this situation, an AI

assistant could alleviate the friction by explicitly summarizing the active filter context, explaining how the panels relate to one another, clarifying what a KPI or bucket represents, and answering view-specific questions such as whether the displayed average delay refers to all open deliveries or only the currently filtered subset. However, H01, H02, and H03 all originated from the dashboard's visual layout and label clarity rather than from assistant behaviour. We therefore interpreted AI support as complementary rather than substitutive. Assistants could reduce residual interpretive friction within a given view, but our findings did not support treating them as a replacement for clear layout, labelling, and information hierarchy in the dashboard itself (Ullrich et al., 2024).

Second, the results highlighted that expertise translation was a major source of friction. H29 and H32 showed that the ability to translate questions into the right filters and views aligned with Decision Effectiveness and higher Decision Confidence. These results suggested that users often knew what they wanted to ask but struggled to express that intent through dashboard interaction primitives. This interpretation also aligned with the screening questions, in which finding the right dashboard or page emerged as the most frequently selected and highest-ranked need for AI support. For instance, this friction could arise when a user wanted to identify which customers drove the highest overdue value in the accounts receivable context. It could also arise when the user wanted to determine whether the issue came mainly from disputes, credit, or collections. In other cases, the user may have wanted to understand which filtered subset of data explained the top KPIs shown in the current view. Without guidance, the user may need to infer which panel contained the relevant evidence, manually open several filter dropdowns, and test multiple slices before reaching the correct view. In this context, an AI assistant could reduce that effort by identifying the most relevant panel, opening or pre-populating the relevant filters, highlighting the chart or table that contained any necessary evidence. It could also have shared a direct link to the appropriate drilldown state and explained why a view was relevant. For example, it could have indicated that disputes and overdue days should be examined together because both indicators were already visible in the current dashboard context. Therefore, we interpreted AI assistance in this area as a translation layer between user intent and dashboard operations. Assistants could guide users toward the correct interaction state, but they still depended on the dashboard as the environment in which the evidence, filters, and drill paths remained visible and auditable.

H20 and H23 indicated that decision outcomes are stronger when respondents perceive dashboard data as credible, complete, and up to date. In our sample, higher overall trust aligned with higher decision effectiveness. Additionally, perceived completeness and freshness of dashboards aligned with higher reported usefulness for job performance. These results suggested that users needed more than an answer, they also needed fast access to credibility cues before they acted on that answer. Open-ended responses supported this interpretation by respondent expressions saying that unclear lineage or calculation logic forced them to "guess" meanings and to "cross compare with [external tools] just to make sure the numbers are accurate." This behaviour showed that verification remained part of the decision workflow when credibility remained unclear. Accordingly, we thought extending headless assistants to enable users to query KPIs without opening the dashboard on external tools should be implemented with caution to preserve data credibility. Our results suggested that this pattern risked increasing validation friction due to assistants rarely exposing structured metadata such as definitions, filter context, and refresh status in a way that supports quick auditing. We therefore positioned the dashboard as the primary representation layer for credibility and auditability, and we framed assistant design as complementary. Assistants should surface KPI definitions, active filters, freshness timestamps, and lineage on demand while staying anchored to the dashboard evidence.

Fourth, the results emphasized actionability. H41 showed that specific and appropriately filtered metrics aligned with higher perceived usefulness for job performance. This finding suggested that dashboards became more valuable when they helped users move from observation to next steps. The ranked assistance needs reinforced this point, placing insights into action immediately after finding the right dashboard. Open-ended responses pointed in the same direction, as some users reported that they needed to export results to spreadsheets or perform manual calculations before acting. For example, this friction could arise when a user recognized that overdue deliveries, disputed invoices, or high days-sales-outstanding values required follow-up queries in the current dashboard view, but still had to determine which customer segment, document type, or operational owner should be prioritized first. In this area, an AI assistant could strengthen user experience by summarizing the most decision-relevant slices, highlighting which filtered subset contributed most to the KPI, comparing the current value to target or best-in-class benchmarks, and proposing concrete next steps such as opening the affected

customer list, exporting the relevant case subset, or generating a shareable follow-up summary for operations teams. It could also reduce manual effort by performing simple comparative calculations within the current filter context and by recommending the next drilldown state needed to turn a dashboard observation into an operational action.

Taken together, these mechanisms showed that the findings suggested that AI assistants may create the greatest value when they reduce friction in understanding, querying, verifying, and acting on dashboard information. The broader survey context supported this interpretation. Respondents did not primarily ask for generic automation, but for help in locating evidence, understanding metrics, and operationalizing results. In our sample, AI assistants did not appear most useful as standalone answer engines; they appeared most useful when they remained grounded in the dashboard context and supported users at the points where decision-related friction emerged most clearly.

6 Conclusion

Even in experienced and relatively satisfied user groups, specific dashboard frictions were systematically aligned with decision-related outcomes. The results showed that decision effectiveness depended on a users` ability to interpret visual information, translate analytical intent into system interactions, verify data credibility, and act on decision-relevant metrics. These findings identify four friction areas that future AI assistants may productively target when designed to remain grounded in dashboard context.

Acknowledgment

The authors used AI-based tools to support language editing and to improve the structure and clarity of the paper. The authors reviewed and revised all AI-assisted suggestions and took full responsibility for the final content.

References

Alhamadi, M., Alghamdi, O., Clinch, S., & Vigo, M. (2022). Data quality, mismatched expectations, and moving requirements: The challenges of user-centered dashboard design. *NordiCHI '22: Nordic Human-Computer Interaction Conference*. doi:10.1145/3546155.3546708

Ammann, J., Lohoff, L., Wurm, B., & Hess, T. (2025). How do process mining users act, think, and feel? an explorative study of process mining use patterns. *Business & Information Systems Engineering*. doi:10.1007/s12599-025-00931-9

Ankomah, D. (2025). Evaluating the impact of AI-powered chatbots on user experience [Preprint]. *TechRxiv*. doi:10.36227/techrxiv.173750043.33655505/v1

Asare-Marfo, D. (2021). Why do some open-ended survey questions result in higher item nonresponse rates than others? Pew Research Center. https://www.pewresearch.org/decoded/2021/10/14/why-do-some-open-ended-survey-questions-result-in-higher-item-nonresponse-rates-than-others/

Atighehchian, A., Alidadi, T., Mohammadi, R. R., Lotfi, F., & Ajami, S. (2023). Identifying the application of process mining technique to visualize and manage in the healthcare systems. In Z. Hu, I. Dychka, & M. He (Eds.), *Advances in computer science for engineering and education vi* (Vol. 181, pp. 299-308). Springer, Cham. doi:10.1007/978-3-031-36118-0_26

Barbieri, N., Madeira, E., Stroeh, K., & van der Alst, W. (2023). A natural language querying interface for process mining. *Journal of Intelligent Information Systems*, *61*, 113-142. doi:10.1007/s10844-022-00759-9

Bender, R., & Lange, S. (2001). Adjusting for multiple testing-when and how? *Journal of Clinical Epidemiology*, *54*, 343-349. doi:10.1016/S0895-4356(00)00314-0

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, *57*(1), 289-300. doi:10.1111/j.2517-6161.1995.tb02031.x

Bernerth, J. B., & Aguinis, H. (2016). A critical review and best-practice recommendations for control variable usage. *Personnel Psychology*, *69*(1), 229-283. doi:10.1111/peps.12103

Boareto, P. A., Loures, E. d. F. R., Santos, E. A. P., & Deschamps, F. (2025). Generative assistant for digital twin simulations [12th CIRP Global Web Conference (CIRPE 2024)]. *Procedia CIRP*, *132*, 129-134. doi:10.1016/j.procir.2025.01.022

Boer, L. d. (2024, August). New process AI capabilities for SAP Signavio offer instant, tailored recommendations [Press Release]. https://news.sap.com/2024/08/new-sap-signavio-process-ai-capabilities-instant-tailored-recommendations/

Calvia, P. (2024, November). New AI and copilot capabilities open up new ways businesses interact with SAP Signavio solutions for business transformation [Press Release]. https://news.sap.com/2024/11/joule-ai-copilot-sap-signavio-business-transformation/

Carpenter, B. (2025, October). Unlocking AI success: The critical role of process intelligence in SAP transformations. Mygo Consulting Inc. Retrieved December 24, 2025, from https://mygoconsulting.com/blog/sap-garage/unlocking-ai-success-the-critical-role-of-process-intelligence-in-sap-transformations/

Chinnathai, M. K., & Alkan, B. (2023). A digital life-cycle management framework for sustainable smart manufacturing in energy intensive industries. *Journal of Cleaner Production*, *419*, 138259. doi:10.1016/j.jclepro.2023.138259

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, *16*(3), 297-334. doi:10.1007/BF02310555

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, *13*(3), 319-340. doi:10.2307/249008

Eisenhardt, K. M. (1989). Building theories from case study research. *Academy of Management Review*, *14*(4), 532-550. doi:10.5465/amr.1989.4308385

Everest Global, Inc. (2025). Process mining state of the market 2025 (State of the Market Report). Everest Group. https://www.everestgrp.com/report/egr-2025-38-r-7549/?utm

Fahland, D., Fournier, F., Limonad, L., Skarbovsky, I., & Swevels, A. J. E. (2025). How well can a large language model explain business processes as perceived by users? *Data & Knowledge Engineering*, *157*, 102416. doi:10.1016/j.datak.2025.102416

Gioia, D. A., Corley, K. G., & Hamilton, A. L. (2013). Seeking qualitative rigor in inductive research: Notes on the Gioia methodology. *Organizational Research Methods*, *16*(1), 15-31. doi:10.1177/1094428112452151

Greenland, S., Senn, S. J., Rothman, K. J., Carlin, J. B., Poole, C., Goodman, S. N., & Altman, D. G. (2016). Statistical tests, p values, confidence intervals, and power: A guide to

misinterpretations. *European Journal of Epidemiology, 31*(4), 337-350. doi:10.1007/s10654-016-0149-3

González-García, J., Estupiñán-Romero, F., Tellería-Orriols, C., González-Galindo, J., Palmieri, L., Faragalli, A., Pristás, I., Vuković, J., Misinš, J., Zile, I., & Bernal-Delgado, E. (2021). Coping with interoperability in the development of a federated research infrastructure: Achievements, challenges and recommendations from the JA-InfAct. *Archives of Public Health*, *79*, 221. doi:10.1186/s13690-021-00731-z

Harati Nik, M., van der Aalst, W. M. P., & Fani Sani, M. (2019). BIpm: Combining BI and process mining. *Proceedings of the 8th International Conference on Data Science, Technology and Applications*, 123-128. doi:10.5220/0007741901230128

Ibanez-Sanchez, G., Castillo, P. A., Juarez, J. M., Fernandez-Llatas, C., Valencia, P., Ruiz-Fernandez, D., & Rojas, I. (2024). Pmapp: An interactive process mining toolkit for building healthcare dashboards. *Explainable AI in Healthcare and Process Mining for Healthcare (XAI-Healthcare/PM4H 2023)*, 2020, 75-86. doi:10.1007/978-3-031-54303-6_8

Ives, B., Olson, M. H., & Baroudi, J. J. (1983). The measurement of user information satisfaction. *Communications of the ACM*, *26*(10), 785-793. doi:10.1145/358413.358430

Kecht, C., Egger, A., Kratsch, W., & Röglinger, M. (2023). Quantifying chatbots' ability to learn business processes. *Information Systems*, *113*, 102176. doi:10.1016/j.is.2023.102176

MacKinnon, J. G., & White, H. (1985). Some heteroskedasticity-consistent covariance matrix estimators with improved finite sample properties. *Journal of Econometrics*, *29*(3), 305-325. doi:10.1016/0304-4076(85)90158-7

Martin, N., Fischer, D. A., Kerpedzhiev, G. D., Goel, K., Leemans, S. J. J., Röglin, M., van der Aalst, W. M. P., Dumas, M., La Rosa, M., & Wynn, M. T. (2021). Opportunities and challenges for process mining in organizations: Results of a delphi study. *Business & Information Systems Engineering*, *63*(5), 511-527. doi:10.1007/s12599-021-00720-0

Martinez-Millana, A., Lizondo, A., Gatta, R., Vera, S., Traver Salcedo, V., & Fernandez-Llatas, C. (2019). Process mining dashboard in operating rooms: Analysis of staff expectations with analytic hierarchy process. *International Journal of Environmental Research and Public Health*, *16*(2), 199. doi:10.3390/ijerph16020199

Meduri, V. V., Quamar, A., Lei, C., Efthymiou, V., & Ozcan, F. (2021). BI-REC: Guided data analysis for conversational business intelligence. *arXiv preprint*. doi:10.48550/arXiv.2105.00467

Miller, A. L., & Lambert, A. D. (2014). Open-ended survey questions: Item nonresponse nightmare or qualitative data dream? *Survey Practice*, *7*(5). doi:10.29115/SP-2014-0024

Molléri, J. S., Petersen, K., & Mendes, E. (2019). An Empirically Evaluated Checklist for Survey Research in Software Engineering. *Information and Software Technology*, *106*, 20-37. doi:10.1016/j.infsof.2019.106240

Morelli, F., Sekulovska, A., & Schätter, F. (2024). Dashboard use case for supply chain resilience management and future research direction. *Communications of the ECMS, Proceedings*, *38*(1). https://www.scseurope.net/dlib/2024/ecms2024acceptedpapers/0416_simo_ecms2024_0039.pdf

Nai, R., Sulis, E., Audrito, D., Trifiletti, V. M. S., Meo, R., & Genga, L. (2025). Leveraging process mining and event log enrichment in european public procurement analysis: A case study. *Computer Law & Security Review*, *57*, 106144. doi:10.1016/j.clsr.2025.106144

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). McGraw-Hill.

O'Brien, R. M. (2007). A caution regarding rules of thumb for variance inflation factors. *Quality & Quantity*, *41*(5), 673-690. doi:10.1007/s11135-006-9018-6

Radziwill, N. M., & Benton, M. C. (2017). Evaluating quality of chatbots and intelligent conversational agents. *arXiv preprint arXiv:1704.04579*. doi:10.48550/arXiv.1704.04579

Rehse, J.-R., Pufahl, L., Grohs, M., & Klein, L.-M. (2022). Process mining meets visual analytics: The case of conformance checking. *arXiv preprint arXiv:2209.09712*. doi:10.48550/arXiv.2209.09712

Sanchez-Obando, J. W., Duque-Méndez, N. D., & Herrera, O. M. B. (2024). Construction of the invoicing process through process mining and business intelligence in the Colombian pharmaceutical sector. *Computers*, *13*(10), 245. doi:10.3390/computers13100245

SAP Signavio. (2023, November). Harnessing the power of generative AI for business process transformation. https://www.signavio.com/events/harnessing-the-power-of-generative-ai-for-business-process-transformation-apj-emea/

SAP Signavio. (2024a). AI interplays with the process world [Ebook]. https://www.signavio.com/downloads/white-papers/e-book-ai-interplays-with-the-process-world/?utm_source=chatgpt.com

SAP Signavio. (2024b). Streamline and optimize processes with sap business AI. https://www.signavio.com/business-ai/

Simic, D., Leible, S., Schmitz, D., Gücük, G., & Kučević, E. (2023). Enhancing project-based learning through data-driven analysis and visualization: A case study. *ACIS 2023 Proceedings*, 139. https://aisel.aisnet.org/acis2023/139

Ullrich, C., Teodora, L., & Benton, G. (2024). Building process mining dashboards manually or AI-assisted with Celonis views. *Proceedings of the International Conference on Process Mining (ICPM) 2024, Tool Demonstration Track*. https://ceur-ws.org/Vol-3783/paper_332.pdf

van der Aalst, W. M. P. (2016). *Process mining: Data science in action* (2nd ed.). Springer. doi:10.1007/978-3-662-49851-4

van der Aalst, W. M. P., Adriansyah, A., van Dongen, B. F., Günther, C. W., Rozinat, A., Rubin, V., Verbeek, H. M. W., Weijters, T., van der Werf, J., & Wynn, M. T. (2012). Process mining manifesto. In *Business process management workshops* (Vol. 99, pp. 169-194). Springer. doi:10.1007/978-3-642-28108-2_19

Velásquez, I., & Sepúlveda, M. (2024). Analyzing the devil's quadrangle of process instances through process mining. In J. De Weerdt & L. Pufahl (Eds.), *Business process management workshops: Bpm 2023* (Vol. 492, pp. 272-284). Springer. doi:10.1007/978-3-031-50974-2_21

Venkatesh, V., Brown, S. A., & Bala, H. (2013). Bridging the qualitative-quantitative divide: Guidelines for conducting mixed methods research in information systems. *MIS Quarterly*, *37*(1), 21-54. https://www.jstor.org/stable/43825936

Wasserstein, R. L., & Lazar, N. A. (2016). The ASA statement on p-values: Context, process, and purpose. *The American Statistician, 70*(2), 129-133. doi:10.1080/00031305.2016.1154108

White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica*, *48*(4), 817-838. doi:10.2307/1912934

Wilson, C. (2024, October). Digital transformation unleashed: The cube explores AI-powered process intelligence innovations. Retrieved December 24, 2025, from https://siliconangle.com/2024/10/25/process-intelligence-celosphere/

Wirth, R., & Hipp, J. (2000). CRISP-DM: Towards a standard process model for data mining. *Proceedings of the 4th International Conference on the Practical Applications of Knowledge Discovery and Data Mining*, 29-39. https://www.cs.unibo.it/~danilo.montesi/CBD/Beatriz/10.1.1.198.5133.pdf

Wixom, B. H., & Todd, P. A. (2005). A theoretical integration of user satisfaction and technology acceptance. *Information Systems Research*, *16*(1), 85-102. doi:10.1287/isre.1050.0042

Wooldridge, J. M. (2020). *Introductory econometrics: A modern approach* (7th ed.). Cengage Learning.

Zimmermann, L. (2023). Enhancing support for individual analysts in process mining. *Proceedings of the ICPM Doctoral Consortium and Demo Track 2023*, 3648. https://ceur-ws.org/Vol-3648/paper_7878.pdf

Zuidema-Tempel, E., Effing, R., & van Hillegersberg, J. (2022). Bridging the gap between process mining methodologies and process mining practices: Comparing existing process mining methodologies with process mining practices at local governments and consultancy firms in the Netherlands. In C. D. Ciccio et al. (Eds.), *Business process management: Bpm 2022 industry track* (Vol. 458, pp. 70-86). Springer. doi:10.1007/978-3-031-16171-1_5

7 Appendix

7.1 Recruitment Procedure

The target population involved users of dashboards on process intelligence platforms. We recruited participants through a vendor-agnostic non-probability sampling approach that combined convenience and snowball sampling. First, we approached individuals in the authors` networks who were known to use dashboards on process intelligence platforms. After confirming their eligibility and willingness to participate, we asked them to refer other eligible users in their own networks. Because this network-based recruitment did not rely on a defined sampling frame, the sample should not be interpreted as statistically representative of all PI dashboard users. We screened all respondents to ensure that they currently use a dashboard on a process intelligence platform before including them in the study. We recorded 63 eligible responses via Qualtrics, of which 58 complete responses entered the regression analyses after excluding five partial completions.

7.2 Survey Structure and Measurement Blocks

Appendix Table 2 summarizes the survey structure, measured constructs, variable codes, and survey blocks.

Table 2: Survey Structure and Measurement Blocks

Construct/ Measure	Variable codes	Short Description	Section/Block
Informed Consent	CONSENT1-CONSENT3	Eligibility (18+) and consent to participate (incl. data handling).	Informed Consent
Screeners & demographics	SCR1; PURPOSE; ROLE; EXP_YEARS; FREQ; NPS; AI_AVAIL	Dashboard use inclusion criterion followed by usage context, role/experience, frequency, advocacy, and AI availability/use.	Screeners & Demographics
Dashboard Anchor	DASH_PICK; DASH_DESC	Select one dashboard actually used and briefly describe its use case to contextualize subsequent ratings.	Dashboard Anchor
Visual Comprehension Issues	VSC1, VSC2, VSC3, VSC4	Clarity of layout, labels, time context and salience via color/emphasis.	Visual Clarity & Findability

Construct/ Measure	Variable codes	Short Description	Section/Block
Navigation difficulty	NAV1, NAV2	Speed to the right page/detail; consistency of filters/menus.	Visual Clarity & Findability
Domain language misalignment	DLM1	Labels/metrics aligned to team language; clarity of codes/IDs.	Trust & Data Credibility
Expectation misalignment	EXM1, EXM2, EXM3	Alignment of KPIs/detail/process view and responsiveness to feedback.	Trust & Data Credibility
Trust issues	TRI1, TRI2, TRI3	Lineage/definitions; completeness/freshness; correct timestamp handling.	Trust & Data Credibility
Fear of surveillance	PRV1	Context and fairness of comparisons.	Trust & Data Credibility
Expertise requirement	EXP1, EXP2	Ability to use the dashboard without specialized technical skills and to translate questions into filters/views without help.	Skills, Guidance & Getting to Action
Lack of user understanding	LKU1	Knowledge of steps/terms/KPIs needed to interpret the dashboard correctly.	Skills, Guidance & Getting to Action
Need of technical support	NTS1	In-tool help/onboarding sufficient to proceed unaided.	Skills, Guidance & Getting to Action
Lack of actionability	ACT1, ACT2	Update cadence supports timely decisions; metrics are specific/filtered enough to guide next steps.	Skills, Guidance & Getting to Action
Manual effort requirement	MER1, MER2, MER3	Thresholds/filters can be set quickly without trial-and-error; routine questions can be answered in-tool without exporting.	Skills, Guidance & Getting to Action
Decision effectiveness	DE1, DE2	Timeliness and confidence of decisions made with the dashboard.	Outcomes
Perceived usefulness	PU1, PU2	Contribution to job performance and role effectiveness.	Outcomes
Perceived ease of use	PEOU1	Ease to learn/operate the dashboard.	Outcomes
Information satisfaction	IS1	Sufficiency/fit of dashboard information for the job.	Outcomes
AI task needs	AIA_TASKS	Which user tasks most need AI help (multi-select).	AI Assistant Feasibility
AI task priorities	AIA_RANK	Rank selected tasks from most to least helpful.	AI Assistant Feasibility

Source: Own

7.3 Reliability Diagnostics

Table 3: Internal consistency (Cronbach's α) and operationalization decisions for multi-item constructs

Construct	Items	k	α	Operationalization in analysis
Visual Comprehension	VSC1, VSC2, VSC3, VSC4	4	0.675	Used as sub-themes: VSC1 → VisualLayoutLabelClarity; VSC2 → VisualTimeContextTrendReadability; VSC3 → VisualEmphasisColorEncoding; VSC4 → VisualGlanceabilityHierarchy
Navigation Difficulty	NAV1, NAV2	2	0.684	Used as sub-themes: NAV1 → NavigationReachDrill; NAV2 → NavigationFilterMenuConsistency
Expectation Misalignment	EXM1, EXM2, EXM3	3	0.652	Used as sub-themes: EXM1 → ExpectationRelevanceGoals; EXM2 → ExpectationProcessFit; EXM3 → ExpectationFeedbackIteration
Trust Issues	TRI1, TRI2, TRI3	3	0.710	Kept as composite: TrustIssues = mean(TRI1, TRI2, TRI3)
Expertise Requirement	EXP1, EXP2	2	0.589	Used as sub-themes: EXP1 → ExpertiseTechnicalIndependence; EXP2 → ExpertiseQuestionQueryTranslation
Lack of actionability	ACT1, ACT2	2	0.626	Used as sub-themes: ACT1 → ActionabilityTimeliness; ACT2 → ActionabilityMetrics
Manual effort requirement	MER1, MER2, MER3	3	0.501	Used as sub-themes: MER1 → ManualEffortConfiguration; MER2 → ManualEffortNoExport; MER3 → ManualEffortCollaboration
Decision effectiveness	DE1, DE2	2	0.811	Kept as composite: DecisionEffectiveness = mean(DE1, DE2)
Perceived usefulness	PU1, PU2	2	0.874	Kept as composite: PerceivedUsefulness = mean(PU1, PU2)

Source: Own

In general, Cronbach's Alpha values are commonly used as a minimum threshold for acceptable internal consistency (Nunnally & Bernstein, 1994). In this study, we used the common rule of thumb for interpreting alpha:

- ↑ 0.90 – excellent
- 0.80–0.89 – good
- 0.70–0.79 – acceptable

- < 0.70 – questionable or poor (items treated separately) (Nunnally & Bernstein, 1994).

Below shown all constructs measured with multiple items in the survey and their Cronbach's alpha values in the collected sample. Single item constructs like Fear of Surveillance, Need of Technical Support, PEOU, and IS; screeners and demographics; and free-text questions are not shown here because construct reliability check was not applicable to these constructs. Based on the operationalization decisions made in this step, constructs kept and referred as either singletons or composites.