

TrustmeWatcher: An Application for Workplace Micro-Sensing and Explainable Well-Being Feedback

Chengyu Yu
Università della Svizzera italiana (USI)
Lugano, Switzerland

Leon Jacopo Costa
Università della Svizzera italiana (USI)
Lugano, Switzerland

Zoja Anžur*
Jožef Stefan Institute (JSI)
Ljubljana, Slovenia

Mohan Li
Università della Svizzera italiana (USI)
Lugano, Switzerland

Gašper Slapničar*
Jožef Stefan Institute (JSI)
Ljubljana, Slovenia

Daniil Kirilenko
Università della Svizzera italiana (USI)
Lugano, Switzerland

Martin Gjoreski
Università della Svizzera italiana (USI)
Lugano, Switzerland

Mitja Luštrek*
Jožef Stefan Institute (JSI)
Ljubljana, Slovenia

Marc Langheinrich
Università della Svizzera italiana (USI)
Lugano, Switzerland

Abstract

Workplace sensing studies combine long-running behaviour traces with self-reports, yet the tools that collect those data often sit apart from the interface that returns results. We present TrustmeWatcher, the application built for the TRUST-ME project to connect this work. TrustmeWatcher reuses ActivityWatch’s OS-level watchers for computer-activity collection and adds its own application layer. It turns the collected traces into an interactive screen-time dashboard, synchronizes responses from short questionnaires completed on the StreamDeck, and presents questionnaires alongside video highlights. Activity records and self-reports are aligned into labelled records for model development. The scope of this paper is limited to ActivityWatch data as model input. Artificial intelligence (AI) uses these activity records to predict six normalized state scores and an overall well-being score. The trained model runs locally, and the dashboard presents its predictions in semantic bands. Explainable artificial intelligence (XAI) helps users understand how recorded activity contributed to a prediction. Privacy Control lets users pause or resume the camera and eye tracker used by the study. We describe the workflow, its user-device and sensing-setup boundaries, and its use with records from 17 participants. The result is a deployed application and study workflow that integrates activity review, study data collection, privacy control, local prediction, and a participant-facing interface for XAI evaluation.

CCS Concepts

• **Human-centered computing** → **Ubiquitous and mobile computing**; • **Computing methodologies** → *Machine learning*.

Keywords

explainable AI, workplace well-being, micro-sensing, self-report, privacy control, ubiquitous computing

1 Introduction

Digital traces can reveal work patterns that repeated surveys alone miss, including changes in engagement, fatigue, stress, physical comfort, and social interaction [1]. The same traces can become instruments of workplace surveillance. TRUST-ME (TRUSTworthy

enhancement of job Satisfaction and productivity using Micro-sensing in work Environments) brings together Università della Svizzera italiana and the Jožef Stefan Institute to study this conflict [11]. Its aim is to examine how sensing and AI can support knowledge workers while preserving privacy, autonomy and providing understandable explanations for its predictions.

Computer activity and self-reports describe different perspectives of a workday. Activity logs record how the computer was used, but they do not say whether a person felt focused, tired, stressed, or productive. Self-reports capture that experience, although they are subjective and burdensome if requested too often. Automatically measured and self-reported productivity can diverge, so neither should stand in for the other [2]. The study collects two forms of self-report: periodic nine-question check-ins using the Stream Deck device seen in Figure 1 during work, and video questionnaires completed after reviewing recorded highlights. TRUST-ME also uses data collected through ActivityWatch namely application usage, mouse and keyboard input. A camera, an eye tracker, audio, and recorded highlights are also used to provide study context. Because the camera and eye tracker capture sensitive contextual data, participants need a visible and reversible way to pause or resume sensing. Such control is particularly important in workplace settings, where power asymmetries can make consent less freely given [5]. It may also help reduce concerns about surveillance, which is associated with lower autonomy and poorer worker well-being [7].

The TRUST-ME study has two phases. The first collected computer activity and repeated self-reports from 17 participants for model development and evaluation. The second, currently being prepared, will use TrustmeWatcher as its participant-facing application. For computer-activity collection, TrustmeWatcher reuses ActivityWatch’s open-source watchers. These background processes locally record application usage and keyboard and mouse activity [3]. TrustmeWatcher’s own service and interface turn the resulting traces into an interactive screen-time dashboard. They also synchronize responses from both questionnaires, prepare labeled records for model development, and run the trained model locally. Daily feedback and ambient-sensor controls use the same interface. Figure 1 shows the boundaries between the user device, the sensing setup, and the modeling pipeline.

This paper contributes:

*Also with Jožef Stefan International Postgraduate School.

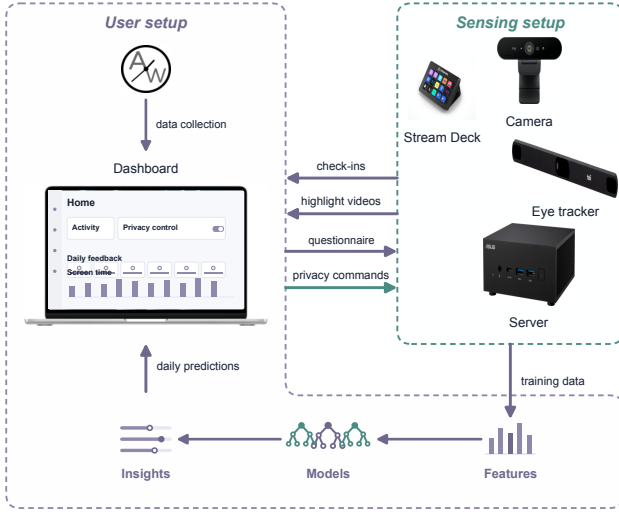


Figure 1: TrustmeWatcher system overview and data flow.

- an integrated application with an interactive screen-time dashboard that lets participants explore computer activity from hours to years, alongside self-report collection, privacy controls, and daily well-being and productivity predictions;
- a longitudinal field-study record that connects repeated self-reports from 17 participants with the computer activity surrounding them; and
- an implemented participant-facing counterfactual interface for examining explanation use across repeated encounters.

2 Related Work

Ubiquitous-computing platforms often combine passive sensing with active reports. StudentLife [12] paired continuous phone sensing with ecological momentary assessment over a university term, while deliberately withholding feedback to avoid changing participant behaviour. RADAR-base supports passive phone and wearable sensing together with active questionnaires, study management, storage, and visualization [9]. In the second phase, TrustmeWatcher gives the participant device a different role. The person who produces the data can review the recorded activity, control ambient sensing, and receive model estimates and explanations in the same application. This makes those interactions observable parts of the participant’s study experience.

Device interaction, physiological sensing, environmental data, and self-reports have different validity and privacy costs [1]. Beller et al. paired desktop telemetry with daily productivity ratings and found that activity measures cannot replace a person’s account of the day [2]. TrustmeWatcher aligns both forms of self-report with the surrounding computer activity.

The model layer generates counterfactual results, while TrustmeWatcher presents them through its participant-facing interface. BayCon produces model-agnostic counterfactual instances over

mixed features [10]. Cheruvalath et al. combine attribution, counterfactuals, and accessible text for non-expert users [4]. The generated explanation still has to fit the context and questions that matter to its user [8]. Repeated use adds questions about comprehension, trust calibration, and changing mental models that a single encounter cannot answer.

3 TrustmeWatcher Application

3.1 System Roles and Data Flow

TrustmeWatcher separates reused OS-level ActivityWatch watchers from its own local backend and dashboard (Figure 2). ActivityWatch watchers record computer activity on the user device. The local backend turns those events into a unified activity history and supports the study-data synchronization, Privacy Control, and local inference paths described below.

The sensing setup provides check-ins, video highlights, and ambient sensing. TrustmeWatcher brings these inputs into the same user workflow, returns video questionnaire responses to the study server, and carries Privacy Control commands back to the camera and eye tracker.

Model development sits outside both runtime boundaries. The study records combine activity context with both forms of self-report for training and evaluation. The trained model is then deployed to the user device. TrustmeWatcher prepares features from the local activity history, runs the model locally, and stores one daily report for the interface. Training data and model predictions follow different routes, as shown in Figure 1.

3.2 A Unified Workplace Study Workflow

TrustmeWatcher presents computer activity through a screen-time interface for immediate review and longer-range exploration (Figure 3). Users can move from hourly activity to weekly, monthly, yearly, and custom periods without changing how applications and categories are defined. The summary combines a timeline, an application distribution, and ranked category and application lists. This gives users a readable account of the same activity history that supports the study.

The application also presents check-ins and video highlights. Participants complete check-ins during work. The application groups the available highlights for a day, allowing the user to review several

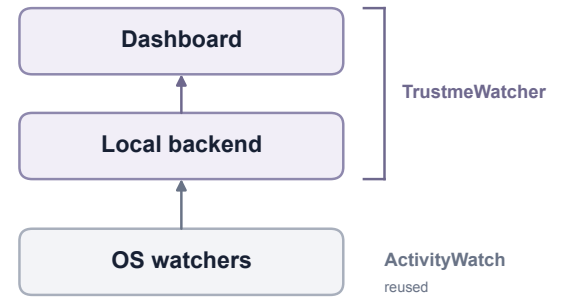


Figure 2: Architecture view.

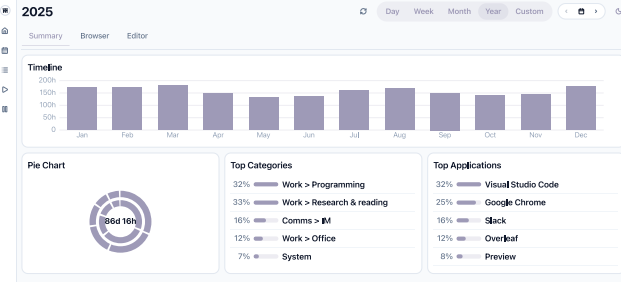


Figure 3: Screen-time summary using synthetic data.

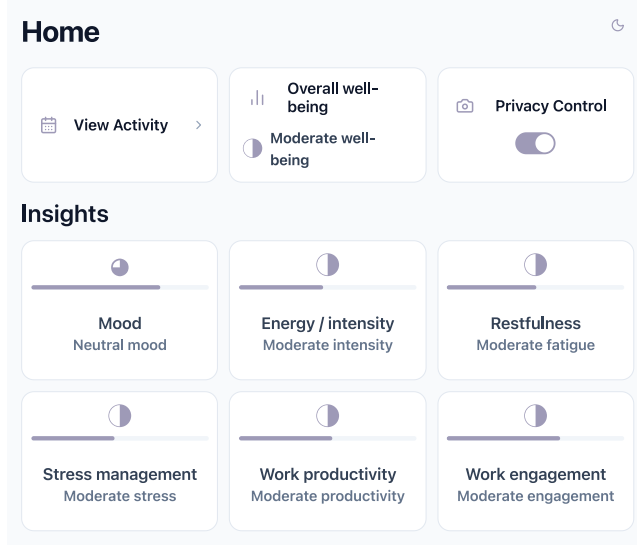


Figure 4: Daily feedback and sensing control using synthetic data.

relevant moments. The two channels provide different temporal views of the same work experience.

TrustmeWatcher integrates Privacy Control into the application (Figure 4). Users can see whether camera and eye-tracker sensing is active and pause or resume it. Pausing Privacy Control turns off the camera and eye tracker. Placing this control beside returned feedback makes consent an ongoing choice instead of an installation-time setting [5].

3.3 From Study Signals to Model Estimates

The study collects nine responses on a Likert scale: mood (q_1), $[-3, +3]$; emotion intensity (q_2), $[0, 6]$; tiredness (q_3), $[0, 6]$; work enthusiasm (q_4), $[0, 6]$; work immersion (q_5), $[0, 6]$; physical comfort (q_6), $[-3, +3]$; social experience (q_7), $[-3, +3]$; work stress (q_8), $[0, 6]$; and productivity (q_9), $[0, 6]$. In the first phase, participants answered the same questions in check-ins during work and in video questionnaires after reviewing recorded highlights. Responses from both sources were combined for model training.

The prediction pipeline uses seven regressors to estimate six state scores and overall well-being. The training targets are calculated

from the questionnaire responses as mood valence $V = q_1 + 3$, arousal $A = q_2$, restfulness $R = 6 - q_3$, stress management $S = 6 - q_8$, productivity $P = q_9$, engagement $E = (q_4 + q_5)/2$, and overall well-being $W = (V + E + S)/3$. Tiredness and stress are reversed so higher scores always indicate a better state. Physical comfort (q_6) and social experience (q_7) are collected but are not prediction targets.

The target construction follows the questionnaire design: distinct states are modeled separately, while complementary items are averaged into composite scores. Cronbach's α was used to assess which state scores were consistent enough to combine for overall well-being. Based on this, valence, engagement, and stress management were included, while fatigue and arousal were excluded. For each target multiple regressors use the same 25-feature vector. This vector summarizes the preceding 60 minutes together with 24-hour and seven-day context. The features describe learned activity-state shares, application-category ratios, focus fragmentation, keyboard and pointer activity, and changes from recent history. Features are standardized separately for each participant using statistics fitted on training data only. A participant not present during training uses global training-set statistics. For each target, the regressor with the lowest validation MSE is selected.

During inference, the same feature pipeline constructs the input vector locally from ActivityWatch data. The application presents the seven estimates on the common 0–6 scale.

3.4 From Model Estimates to Explanations

TrustmeWatcher presents the six state scores as semantic cards in Figure 4. Scores below 2 map to the low band, scores from 2 to below 5 to moderate, and scores from 5 to 6 to high. Labels follow each construct: a high stress-management score appears as low stress, and a high restfulness score as low fatigue. Selecting a card opens the counterfactual view in Figure 5.

The user chooses a target score on the shared 0–6 scale. Starting from the current 60-minute computer-activity allocation, the search swaps time between personal distraction, media, communication, development, writing and research in five-minute units. A swap removes five minutes from one category and adds those five minutes to another category. The system evaluates every valid allocation and returns the one whose model estimate is closest to the selected target. This zero-sum search adopts the framing of counterfactual explanation used in prior work [4, 10].

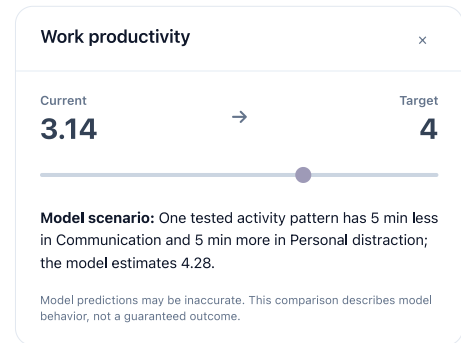


Figure 5: Counterfactual example.

4 Field Study and Model Evaluation

4.1 Data-Collection Study

Participants and setting. The TRUST-ME data-collection study followed knowledge workers in their everyday work setting rather than in a single laboratory session. The current archive contains pseudonymized records from 17 participants and spans July 2025 to March 2026. The amount of recorded self-report data varies with each participant’s work schedule: individual records cover between 10 and 28 distinct workdays, with a median of 18.

Procedure. Computer activity was recorded in the background while participants worked. At intervals during the day, participants completed check-ins on the StreamDeck to describe their current state. The sensing setup also produced video highlights for retrospective reflection. For each available highlight, participants rated the same dimensions in the video questionnaire while viewing the recorded moment. This produced two kinds of self-report: an immediate response close to the experience and a later response grounded in a concrete episode from the day.

Study record. All 17 participants contributed computer-activity context and check-ins; nine also contributed video questionnaire responses. TrustmeWatcher preserves the response source and timestamp when it aligns each self-report with the preceding activity context. This allows the two channels to be analysed together without treating them as interchangeable labels. The aligned records feed the external model-development pipeline, and the deployed model later uses recent and longer-range activity features to produce daily estimates on the user’s device.

4.2 Prediction Performance

Questionnaire days were ordered chronologically within each participant and divided approximately 80/10/10 into training, validation, and test sets. One complete day at each boundary was excluded to reduce temporal overlap. Instances were not shuffled to avoid data-leakage. Models were selected using validation MSE and evaluated once on the held-out test set.

We compared the model with random, global-mean, and personal-mean baselines. The random baseline draws scores uniformly from the 0–6 response scale. The global mean is calculated separately for each target using the training and validation responses. The personal mean is calculated from each participant’s own training and validation responses.

Table 1: Within-user prediction performance

Model	Validation MSE ↓	Test MSE ↓
Random baseline	5.9	5.8
Global-mean baseline	1.4	1.3
Personal-mean baseline	1.0	0.85
Our model	0.85	0.78

The global mean tests improvement over a population-level constant. The personal mean provides a stricter comparison because it tests whether the activity features add information beyond each participant’s typical response.

5 Discussion

The reported evaluation covers prediction performance and data-collection coverage; explanation comprehension, trust calibration, and interpretation of model scenarios require analysis of repeated participant interactions [6]. Computer activity omits meetings and off-screen work, and application categories cannot establish intent, so estimates reflect recorded context rather than complete accounts of work or well-being.

6 Conclusion

TrustmeWatcher integrates a model developed from records collected from 17 participants with computer-activity review, self-report, privacy control, local prediction, and counterfactual explanations for the second study phase. It provides the infrastructure for longitudinal evaluation of explanation use, trust, and privacy choices.

References

- [1] Zoja Anžur, Klara Žinkovič, Junoš Lukan, Pietro Barbiero, Gašper Slapničar, Mohan Li, Martin Gjoreski, Maike E. Debus, Sebastijan Trojer, Mitja Luštrek, and Marc Langheinrich. 2025. A Review of Methods for Unobtrusive Measurement of Work-Related Well-Being. *Machine Learning and Knowledge Extraction* 7, 3 (2025), 62. <https://doi.org/10.3390/make7030062>
- [2] Moritz Beller, Vince R. Orgovan, Spencer Buja, and Thomas Zimmermann. 2021. Mind the Gap: On the Relationship Between Automatically Measured and Self-Reported Productivity. *IEEE Software* 38, 5 (2021), 24–31. <https://doi.org/10.1109/MS.2020.3048200>
- [3] Erik Bjäreholt and Johan Bjäreholt. 2024. *ActivityWatch*. ActivityWatch. <https://doi.org/10.5281/zenodo.13893531>
- [4] Swathy Sathesnan Cheruvalath, Matias Laporte, Francesco Bombassei De Bona, Teena Hassan, and Martin Gjoreski. 2025. Generating Explanations for Models Predicting Student Exam Performance. In *Companion of the 2025 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. Association for Computing Machinery, New York, NY, USA, 1679–1684. <https://doi.org/10.1145/3714394.3756320>
- [5] Shreya Chowdhary, Anna Kawakami, Mary L. Gray, Jina Suh, Alexandra Olteanu, and Koustuv Saha. 2023. Can Workers Meaningfully Consent to Workplace Wellbeing Technologies?. In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency*. Association for Computing Machinery, New York, NY, USA, 569–582. <https://doi.org/10.1145/3593013.3594023>
- [6] Martin Gjoreski, Matias Laporte, Marc Langheinrich, and Tim Miller. 2024. How to Validate XAI in Longitudinal Studies?. In *Companion of the 2024 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. Association for Computing Machinery, New York, NY, USA, 866–869. <https://doi.org/10.1145/3675094.3678997>
- [7] Paul Glavin, Alex Bierman, and Scott Schieman. 2024. Private Eyes, They See Your Every Move: Workplace Surveillance and Worker Well-Being. *Social Currents* 11, 4 (2024), 327–345. <https://doi.org/10.1177/23294965241228874>
- [8] Tim Miller. 2019. Explanation in Artificial Intelligence: Insights from the Social Sciences. *Artificial Intelligence* 267 (2019), 1–38. <https://doi.org/10.1016/j.artint.2018.07.007>
- [9] Yatharth Ranjan, Maximilian Kerz, Zulqarnain Rashid, Sebastian Böttcher, Richard J. B. Dobson, and Amos A. Folarin. 2018. RADAR-base: A Novel Open Source m-Health Platform. In *Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers*. Association for Computing Machinery, New York, NY, USA, 223–226. <https://doi.org/10.1145/3267305.3267579>
- [10] Piotr Romashov, Martin Gjoreski, Kacper Sokol, Maria Vanina Martinez, and Marc Langheinrich. 2022. BayCon: Model-Agnostic Bayesian Counterfactual Generator. In *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence*. International Joint Conferences on Artificial Intelligence Organization, California, USA, 740–746. <https://doi.org/10.24963/ijcai.2022/104>
- [11] Università della Svizzera italiana. 2023. TRUST-ME: TRUSTworthy Enhancement of Job Satisfaction and Productivity Using Micro-Sensing in Work Environments. Swiss National Science Foundation project.
- [12] Rui Wang, Fanglin Chen, Zhenyu Chen, Tianxing Li, Gabriella Harari, Stefanie Tignor, Xia Zhou, Dror Ben-Zeev, and Andrew T. Campbell. 2014. StudentLife: Assessing Mental Health, Academic Performance and Behavioral Trends of College Students Using Smartphones. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. Association for Computing Machinery, New York, NY, USA, 3–14. <https://doi.org/10.1145/2632048.2632054>