AI-Powered Performance Feedback: Empirical Insights on Workforce Engagement and Labor Relations

Yun-Cheng Tsai*
Department of Technology Application and Human Resource Development
National Taiwan Normal University
No.162, He-Ping East Road Sec. 1, Taipei 10610, Taiwan

Abstract

Artificial intelligence (AI) and people analytics are rapidly reshaping how organizations evaluate and develop their workforce. While vendors and management consultancies highlight efficiency gains and more "objective" feedback, labor scholars and worker advocates warn that these systems can deepen surveillance and intensify control. This paper examines how AI-powered performance feedback systems affect workforce engagement and labor relations.

Building on literature on algorithmic management and worker technology rights, we propose a three-dimensional framework that distinguishes development-oriented, control-oriented, and co-governed AI feedback systems. We then elaborate this typology through three empirical illustrations and one normative scenario. The first is a resource-constrained health-tech startup (pseudonym "Jubo") that implemented an AI-supported 360-degree feedback system to sustain engagement in a high-burnout care environment. The second draws on a case study of digital performance management in an electronics manufacturer, highlighting how data-driven dashboards can extend managerial control. The third synthesizes evidence from call centers, where algorithmic monitoring and automated targets produce a digitally intensified form of Taylorism. The fourth is a forward-looking scenario of co-governed AI feedback, informed by recent proposals for worker technology rights.

A distinctive contribution of the paper is that it does not treat AI systems as purely conceptual. We describe and demonstrate a working prototype that combines large language models (LLMs), a multi-agent coordination layer, and automated data collection from web-based HR systems. The prototype transforms free-text feedback into structured indicators, supports qualitative and quantitative analysis, and generates visual reports for managers and worker representatives. Using this system as a methodological bridge, we show how AI-enabled feedback reconfigures voice, trust, and perceived fairness, and how these dynamics feed back into engagement, turnover, and the climate of labor relations.

We argue that AI-powered performance feedback is neither inherently emancipatory nor inherently oppressive. Its effects depend on how feedback is framed (development versus discipline), which data are collected, who can see and contest the analytics, and whether worker representatives have a role in system design and governance. The paper concludes with design principles for AI-enabled feedback systems that support both performance and decent work.

Keywords: artificial intelligence; people analytics; performance feedback; employee engagement; algorithmic management; labor relations.

1 Introduction

AI and people analytics tools are increasingly embedded in everyday HR practices, from recruitment and scheduling to performance evaluation and learning. Vendors and consultancies promote AI-powered performance feedback tools that promise continuous, data-driven insights into individual contribution and potential. At the same time, critical scholars and unions warn that algorithmic management can erode

*Corresponding author. Email: pecu@ntnu.edu.tw

autonomy, intensify work, and undermine collective bargaining ((Stefano and Taes, 2022; Giermindl et al., 2021)).

This tension is particularly salient in the domain of performance feedback ((Cianci et al., 2010)). Historically, performance appraisal has been criticized for being infrequent, biased, and disconnected from development ((Cappelli and Tavis, 2016)). AI systems are positioned as a remedy: they can aggregate behavioral and output data in real time, detect patterns that humans overlook, and deliver personalized feedback at scale ((Subhadarshini et al., 2024)). Yet the same capabilities can also turn into instruments of micro-monitoring, automated discipline, and opaque decision-making that workers cannot meaningfully challenge.

In this paper, we ask: *How do AI-powered performance feedback systems shape workforce engagement and labor relations?* More specifically, we investigate:

1. Under what conditions do AI feedback systems support employee voice, learning, and perceived fairness?

2. When and how do they intensify control, undermine trust, or trigger conflict?

3. What governance arrangements can steer AI-enabled feedback toward more balanced labor relations?

Our contribution is threefold. First, we synthesize insights from algorithmic management, people analytics, and labor relations to develop a conceptual typology of AI-powered performance feedback systems. Second, we ground this typology in three empirical illustrations and one normative scenario:

1. a health-tech startup ("Jubo") that implemented an AI-supported 360-degree feedback process;

2. a digital performance management system in a large manufacturing firm;

3. a call center case where algorithmic control shapes work pace, monitoring, and discipline; and

4. a co-governed AI feedback model that builds on recent proposals for worker technology rights.

Third, we describe a working AI prototype that combines a large language model, a multi-agent coordination layer, and automated data collection, and show how such a system can be used both as an HR tool and as a research instrument for labor and employment scholarship.

The paper is written for an audience of labor and employment researchers, HR practitioners, and worker advocates who are grappling with the practical and ethical implications of AI in performance management.

2 Conceptual Background

2.1 Algorithmic management and information asymmetry

The algorithmic management literature documents how software systems assign, monitor, and evaluate work at scale in platform work, logistics, and other sectors. Lee et al. (2015) describes how data-driven management restructures control and discretion in ride-hailing platforms. In contrast, Rosenblat and Stark (2016) shows how Uber's app combines rhetoric of autonomy with extensive data collection, opaque rating algorithms, and unilateral changes to compensation. These systems create new forms of information asymmetry: firms gain detailed visibility into worker behavior, while workers often lack insight into how their data are interpreted or used.

From a labor relations perspective, these developments challenge traditional mechanisms of worker voice and collective bargaining. Algorithmic allocation and evaluation can fragment the workforce, individualize performance problems, and undermine solidarity. At the same time, they create new points of contestation and regulation, such as rights to explanation, limits on monitoring, and joint oversight of data systems.

2.2 AI-powered feedback and people analytics

People analytics is commonly defined as the systematic use of data about workers and work to inform HR decisions and organizational strategy. Professional bodies highlight their potential to improve engagement, retention, and performance by uncovering patterns in turnover, absence, and survey data (Chartered Institute of Personnel and Development, 2024). Recent industry reports emphasize that many leaders see promise in AI-driven feedback loops that provide more granular insights into workforce experience and enable more informed decisions about talent and work design ((Vhora et al., 2024)).

In performance management specifically, AI systems are marketed as tools to:

- aggregate diverse signals (project data, communication traces, customer feedback) into dashboards;
- flag strengths and development needs for individuals and teams;
- support continuous, “check-in” style feedback rather than annual reviews;
- reduce bias by grounding evaluation in data rather than memory.

Empirical studies remain limited, but early evidence suggests that well-implemented systems can make feedback more frequent and goal-oriented. At the same time, scholars warn that algorithmic evaluations can reproduce existing inequities if training data reflect biased histories, or if models lack transparency and avenues for contestation (Bernhardt et al., 2021).

2.3 Employee engagement and the quality of labor relations

Employee engagement typically refers to a positive, fulfilling work-related state of vigor, dedication, and absorption, and is associated with outcomes such as performance, retention, and discretionary effort. HR practitioners increasingly measure engagement through surveys, pulse checks, and behavioral analytics, linking engagement metrics to people analytics dashboards (Chartered Institute of Personnel and Development, 2024).

From a labor relations standpoint, engagement cannot be treated as a purely psychological variable. It is shaped by structural factors such as job security, workload, managerial style, and the presence (or absence) of credible channels for voice and influence. AI-powered feedback systems sit at the intersection of these domains: they can clarify expectations and provide workers with faster developmental feedback, but they can also expose them to more frequent judgment, peer comparison, and automated sanctions. Whether engagement rises or falls depends on how these systems are designed and governed.

2.4 A three-dimensional framework for AI-powered feedback

Drawing on the above strands, we conceptualize AI-powered performance feedback systems along three dimensions:

1. **Feedback orientation** (development vs. discipline): whether the primary framing is growth, learning, and coaching, or monitoring, compliance, and sanction.
2. **Control intensity** (low vs. high): the degree to which data collection and analytics extend management’s capacity to observe and shape worker behavior in real time.
3. **Worker participation in system governance** (low vs. high): the extent to which workers and their representatives have a say in what data are collected, how they are interpreted, and how feedback is used in decisions.

Figure 1 summarizes this framework. An AI-powered feedback system is understood not as a neutral tool, but as a socio-technical infrastructure whose design choices shape both engagement and the climate of labor relations.

In the following sections, we use this framework to organize a comparative analysis of different types of AI-powered feedback and to connect high-level concepts to a concrete system implementation.

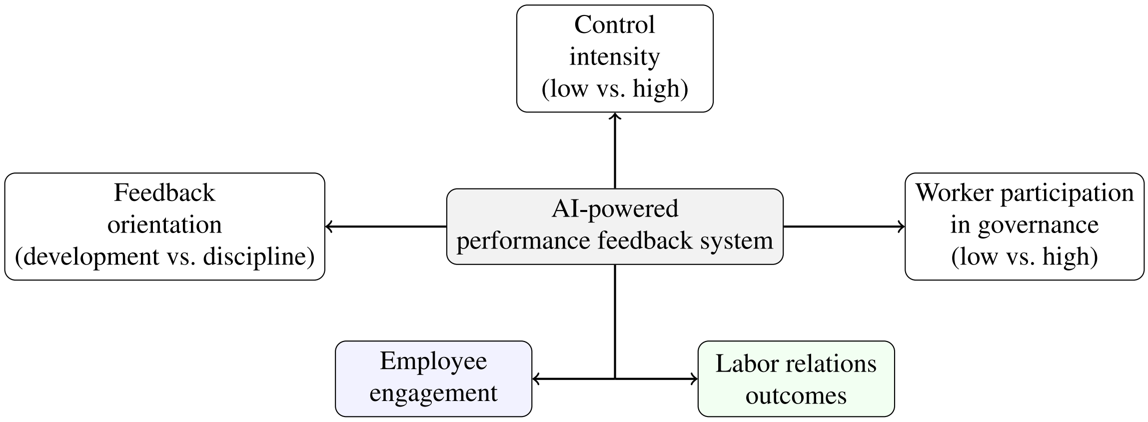

Figure 1: Conceptual framework: AI-powered performance feedback systems, their defining dimensions, and their influence on employee engagement and labor relations outcomes.

3 Research Design and Data

This study adopts a multi-method, comparative approach that combines:

- a longitudinal field study of a health-tech startup (Jubo, pseudonym) that implemented an AI-supported 360-degree feedback system in a care work context;
- secondary analysis of existing case studies on digital performance management and algorithmic control in call centers and manufacturing
- a normative scenario of a co-governed AI feedback system, inspired by emerging proposals for worker technology rights.

The Jubo case is based on internal HR records, anonymized survey data, and semi-structured interviews with managers and frontline workers collected over 18 months. The digital performance management and call center cases are drawn from published book chapters and investigative reports that provide rich qualitative descriptions of system design and worker experience. The co-governed scenario is constructed analytically, aligning with ongoing policy debates around algorithmic accountability and labor rights.

In addition, we built and deployed a working AI prototype that ingests anonymized performance feedback data, applies structured analysis, and generates visual reports. This prototype serves two purposes: it shows that the concepts in our framework can be operationalized using existing AI tooling, and it provides a concrete design space for unions and HR practitioners who wish to experiment with co-governed AI systems.

Given the heterogeneity of data sources, our aim is not to produce statistically representative findings. Instead, we use these cases and the prototype to elaborate a conceptual typology, trace mechanisms, and identify points of intervention for labor relations.

4 System Implementation: AI Modules and Data Pipeline

This section describes the architecture of the AI-powered feedback prototype used in the Jubo case. Rather than treating AI as a black box, we detail how free-text feedback is processed, how indicators are constructed, and how the system is positioned in relation to human decision-makers.

4.1 Overall architecture

The prototype is implemented in Python and combines four layers:

1. **LLM layer**: a large language model accessed via an API is responsible for summarizing free-text feedback, assigning scores on predefined dimensions, and detecting early warning signals (e.g., burnout, perceived unfairness).

2. **Multi-agent coordination layer**: built using an open-source conversational agent framework, this layer orchestrates several specialized AI agents, for summarization, typology classification, and risk screening, in a structured dialogue.

3. **Data collection and preprocessing layer**: drawing on prior automation work in other domains, a headless browser component logs into web-based HR dashboards, exports relevant tables (e.g., attendance, project completion, quality metrics), and merges them with feedback data exported as CSV.

4. **Reporting and visualization layer**: a lightweight web interface and report generator produce CSV outputs, PDF summaries, and HTML dashboards that combine numerical indicators, trend charts, and narrative explanations.

Crucially, the prototype is positioned as a support tool rather than an autonomous decision-maker. All outputs are intended as inputs to conversations between managers, workers, and, where applicable, unions or works councils; they do not directly trigger pay or discipline.

4.2 LLM-based scoring and summarization

To turn long, unstructured feedback comments into variables linked to engagement and turnover, we built an LLM-based scoring module. At a high level, the module takes each comment as input and returns a compact, structured record that captures its central themes, tone, and evaluative content.

In the Jubo implementation, each review cycle generates a corpus of peer, manager, and cross-functional feedback for each focal employee. These comments are often long, unstructured, and written in different styles. The adapted module processes this corpus in four steps:

1. **Preprocessing**: We anonymize comments where possible (removing names, team labels, and project codes) and apply basic text cleaning to normalize line breaks and punctuation. Comments are batched to respect the LLM API's token limits.

2. **Structured JSON output**: For each comment, the LLM is prompted to produce a small JSON object with fields such as:

   - `strengths`: concise bullet-style summaries of recurring strengths;
   - `development_needs`: suggested areas for growth framed in constructive language;
   - `engagement_signal`: a binary or scalar indicator of burnout, detachment, or enthusiasm;
   - `control_perception`: a 1–5 score reflecting whether the comment expresses concerns about excessive monitoring or pressure;
   - `relational_tone`: a categorical label for the emotional tone of comments about managers and colleagues.

   These fields mirror the constructs central to our conceptual framework.

3. **Multi-dimensional ratings**: Similar to multi-item rubrics in education, we define a small set of evaluation dimensions for each comment (e.g., specificity of feedback, developmental focus, solution orientation). The LLM assigns Likert-style scores (1–5) for each dimension.

4. **Aggregation**: The JSON outputs are parsed into a tabular format using a standard data frame library, and comment-level variables are aggregated to the employee level. This yields an integrated

table in which each employee has both traditional HR metrics (e.g., attendance, tenure) and LLM-derived indicators such as a development orientation score (*DevScore*), defined as the proportion of that employee's comments that are classified as developmentally framed.

This design allows the same textual data to support both qualitative interpretation (via selected quotations and narratives) and quantitative modeling (via numeric summaries). A small subset of comments was double-coded by human researchers to refine prompts and verify that LLM outputs aligned with the intended categories; we treat this as an initial plausibility check rather than a full formal validation.

4.3 Deriving a development orientation score

To connect feedback content with engagement and turnover, we define a simple development orientation score at the individual level. We first ask the LLM to classify each comment as primarily:

1. *developmental*: emphasizing growth, learning, and constructive suggestions;
2. *neutral*: primarily descriptive or informational;
3. *disciplinary*: emphasizing errors, blame, or implicit threat of sanctions.

An initial round of LLM classifications is spot-checked and corrected by human coders, and prompt refinements are applied to reduce systematic misclassifications.

For each employee i, we define:

$$\text{DevScore}_i = \frac{\#\{\text{developmental comments for employee } i\}}{\#\{\text{all comments for employee } i\}}.$$

Which ranges from 0 to 1. Higher values indicate that a larger share of feedback received by employee i is framed in developmental terms.

4.4 Multi-agent analysis and interactive feedback

The second major component is a multi-agent analysis layer. In earlier work, a similar architecture was used to analyze emotional diaries: several AI agents, including an emotion analyzer, a recommendation generator, and a quality checker, jointly generated insights and trendlines for individual users.

In the present context, we adapt this pattern to performance feedback. We define three archetypal agents:

- **Summarizer Agent**: reads all comments for a focal employee and produces a short list of strengths and development priorities, explicitly referencing underlying comments.
- **Typology Agent**: takes the summaries and raw comments and classifies the overall pattern for this cycle as primarily development-oriented, control-oriented, or relationship-repair-oriented, providing a brief justification.
- **Risk Screener Agent**: reviews summaries and classifications, scanning for signals such as acute burnout, loss of trust, or perceived unfairness; when such signals are detected, it flags the case for human follow-up.

These agents interact in a turn-taking protocol, implemented through a group chat mechanism: the Summarizer produces a draft, the Typology Agent responds with a label and critique, and the Risk Screener issues a final note that either validates or questions the previous outputs. A simplified log of this interaction is stored alongside the final summary, offering HR and worker representatives a view of the "reasoning" steps rather than only seeing final scores.

In addition, following the earlier diary-analysis system, we compute time series of engagement-related indicators (e.g., self-reported engagement scores from surveys, LLM-inferred engagement signals from comments) for each employee and team. These are plotted as simple line charts across review cycles, helping managers and worker representatives detect gradual cooling or sudden drops.

4.5 Automated data collection and report generation

The third component extends prior automation work originally developed to extract health records from a clinical dashboard. Using a headless browser framework, the system can:

1. log into a web-based HR system with dedicated credentials;
2. iterate over relevant views (e.g., departments, time windows, performance dashboards);
3. extract tabular data such as attendance, project completion rates, and quality metrics;
4. export the extracted tables as CSV files;
5. merge these structured metrics with the LLM-derived feedback indicators described above.

Once merged, the combined dataset is passed to an AI-assisted reporting module. This module generates HTML and PDF reports that include:

- basic descriptive statistics for each team (e.g., average DevScore, distribution of engagement signals);
- illustrative charts (e.g., engagement trends, distributions of control perception scores);
- narrative interpretations drafted by the LLM and reviewed by a human analyst.

By automating data extraction and report synthesis, the system allows researchers and practitioners to run repeated analyses over time with relatively low incremental cost. In the context of this paper, it also demonstrates that our conceptual variables, such as development orientation and perceived control, can be operationalized as measurable constructs in a real system.

5 Typology of AI-Powered Performance Feedback Systems

Table 1 summarizes the typology that emerged from our conceptual synthesis and empirical analysis. We differentiate three ideal types: *development-oriented*, *control-oriented*, and *co-governed* AI feedback systems. Real-world implementations often mix features from different types, but the typology helps clarify dominant logics and their implications.

Using the control intensity and worker participation dimensions, we can place concrete cases in a two-dimensional space. Figure 2 illustrates this mapping for the three empirical examples discussed in the next section and indicates the regions corresponding to the three ideal types.

6 Empirical Illustrations

6.1 Case 1: Development-oriented AI feedback in a care startup (Jubo)

Jubo is a rapidly growing health-tech startup that provides AI-assisted workflow tools for long-term care institutions. Its workforce comprises software engineers, data scientists, and a large number of care coordinators and customer success staff who interact daily with nurses and care managers in client organizations. Work is emotionally demanding: coordinators handle urgent clinical alerts, family concerns, and complex documentation across time zones.

Facing rapid scaling and signs of burnout, the leadership team wanted a performance feedback system that would:

1. recognize invisible emotional and relational labor;
2. surface strengths beyond quantitative metrics;

Table 1: Typology of AI-powered performance feedback systems

	Development-oriented	**Control-oriented**	**Co-governed**
Primary purpose	Learning, coaching, growth	Monitoring, compliance, cost control	Jointly defined performance and well-being goals
Feedback cycle	Frequent, conversational check-ins	Continuous monitoring, periodic automated reports	Frequent, with negotiated cadence and formats
Data sources	Goals, peer feedback, self-reflection, project outcomes	Detailed activity logs, quotas, error rates, time-on-task	Mix of performance, health, and climate indicators agreed in bargaining
Control intensity	Moderate	High	Moderate
Worker voice in design	Limited consultation, mostly managerial	Minimal	Formal role for unions/worker reps in design and oversight
Link to sanctions	Emphasis on development; limited direct link to pay or discipline	Strong link to targets, pay, and discipline	Explicit rules limiting use for discipline; grievance mechanisms
Typical outcomes for engagement	Can strengthen engagement if perceived as fair and supportive	Risk of stress, anxiety, and disengagement	Can support trust and sustainable engagement
Labor relations climate	Partnership rhetoric, but asymmetrical power	Heightened conflict potential; individualized resistance	Institutionalized negotiation around data and algorithms

3. identify misalignment early without resorting to punitive measures.

The HR function lacked the capacity to run traditional, administratively heavy 360-degree appraisals. Instead, the company implemented a simplified, AI-supported 360 system built on the architecture described in Section 4. The design had three key features:

- **Multi-source input**: For each review cycle, employees selected peers, cross-functional collaborators, and (where applicable) client representatives to provide structured feedback on collaboration, communication, and reliability.
- **AI-supported synthesis**: The LLM-based scoring and summarization module processed free-text feedback into thematic summaries and scores on developmental orientation, engagement signals, and perceived control. Managers and employees could inspect the raw comments, but AI summaries helped reduce cognitive load and bias in reading long narratives.
- **Developmental framing**: The system decoupled 360-degree feedback from compensation de-

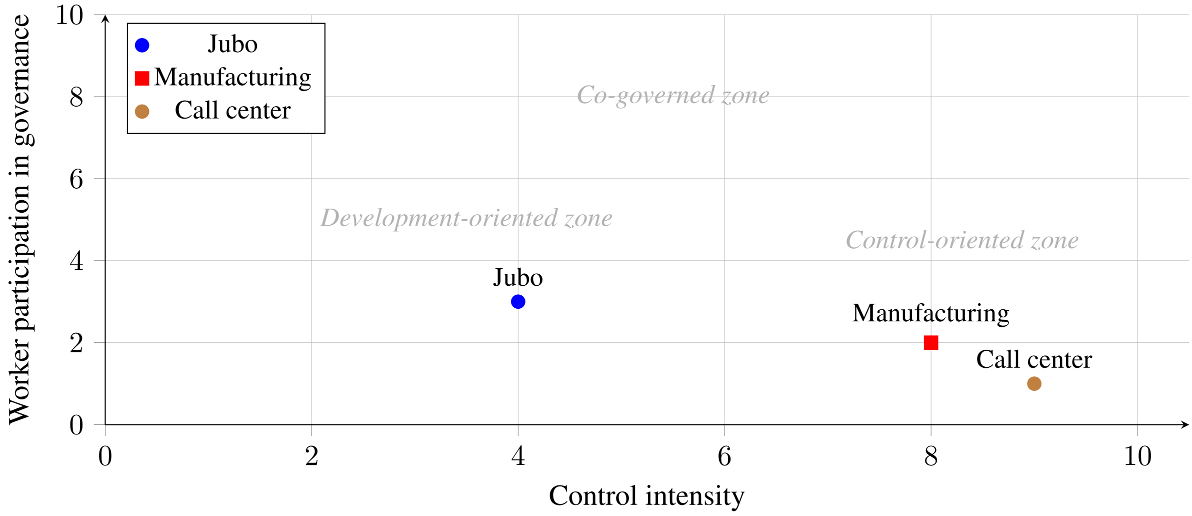

Figure 2: Locations of the three empirical cases in a space defined by control intensity and worker participation in governance. Colored markers indicate the cases; grey italic labels indicate the regions corresponding to development-oriented, control-oriented, and co-governed configurations.

> cisions. Performance-related pay was calculated based on a separate set of agreed quantitative metrics (e.g., issue resolution times, customer satisfaction), while 360-degree feedback was explicitly framed as developmental.

Over three cycles, participation rates in providing peer feedback rose steadily, and interview data suggest that employees appreciated the ability to give and receive more specific recognition. Many reported that AI-generated summaries helped them see patterns they had not noticed, such as recurring praise for mentoring or early warning signs about communication style. At the same time, some coordinators expressed concern that the system might later be repurposed for disciplinary use, indicating a latent trust issue.

Figure 3 presents an illustrative pattern observed in the turnover data: an initial uptick in departures after the system's introduction, followed by a gradual decline as expectations and role fit became clearer. Values in the figure are schematic, but the pattern matches both HR records and interview accounts.

In the typology space (Figure 2), Jubo's system sits primarily in the development-oriented zone: control intensity increased modestly due to more structured data collection, but the absence of direct links to pay or sanctions, combined with managers' emphasis on coaching conversations, helped maintain a relatively positive engagement climate. Nonetheless, the asymmetric ability of management to archive and reinterpret AI summaries raises questions about long-term governance and data use.

6.2 Case 2: Control-oriented digital performance management in manufacturing

Our second illustration draws on a case study of digital performance management in a large electronics manufacturer ("Advantech" in the original survey) (Zhou et al., 2018). In this context, the firm implemented a digital platform that aggregated real-time data from production lines, quality control, and maintenance systems into dashboards for supervisors and senior management.

The system enabled granular tracking of output, defect rates, and downtime at the team and individual worker levels. Performance indicators were highly standardized and tightly linked to production targets. Dashboards made deviations from targets immediately visible, and supervisors were expected to intervene when red indicators persisted. Over time, the system became central to performance appraisal and bonus allocation.

From a control perspective, the digital performance management platform extended management's capacity to monitor and steer work processes. It also standardized how performance was defined and

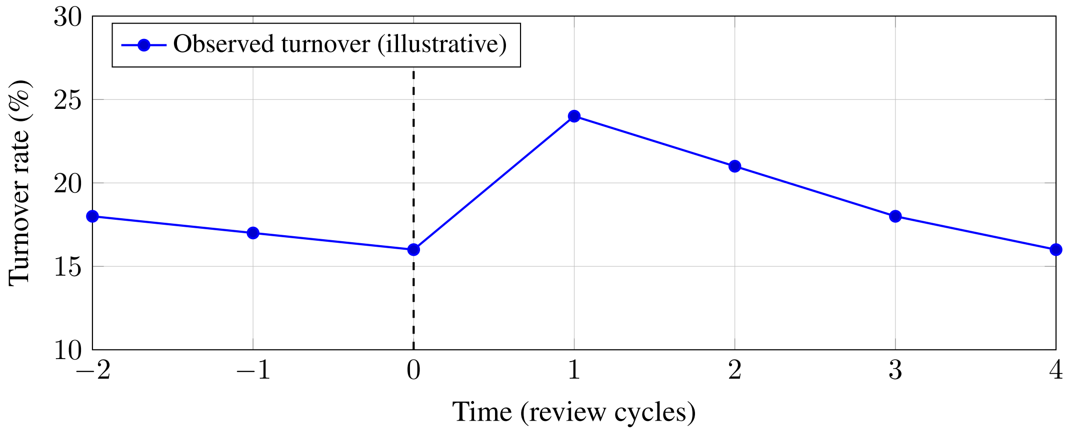

Figure 3: Illustrative turnover dynamics around the introduction of AI-supported 360-degree feedback at Jubo. Turnover initially increases as expectations are clarified and misalignments surface, then declines as cultural alignment and engagement stabilize. Values are schematic and for conceptual illustration only.

evaluated, leaving little room for local negotiation. Workers in the case study expressed ambivalent reactions: some appreciated more precise targets and quicker problem resolution, while others felt under constant pressure and worried about being "reduced to numbers."

In our typology, this case illustrates a predominantly control-oriented AI feedback system: high control intensity, strong coupling to sanctions and rewards, and limited worker participation in design. Engagement outcomes appear mixed, with potential for both increased clarity and heightened stress. For labor relations, the system may shift power toward management unless worker representatives gain access to similar data and a voice in how metrics are chosen and interpreted.

6.3 Case 3: Algorithmic control and feedback in call centers

A third illustration comes from an investigative case study of algorithmic monitoring and management in call centers (Christl, 2023). In these settings, software platforms track a wide range of metrics, including call volume, handling time, after-call work, adherence to scripts, and customer satisfaction scores. Supervisors and higher-level managers can view dashboards that rank agents, highlight deviations from expected patterns, and trigger automated interventions such as coaching prompts or disciplinary processes.

The case study documents how these systems create a dense feedback environment. Workers receive frequent notifications about their performance relative to targets and peers, and their adherence to schedules is tightly controlled. However, feedback is largely unidirectional and oriented toward compliance: agents are expected to adjust their behavior to meet predefined metrics, with limited scope to question whether the metrics are appropriate or to contextualize their performance.

Workers interviewed in the case report high levels of stress and anxiety, linked to constant monitoring and the fear of losing their jobs if they fall short of algorithmically enforced quotas. Some described strategies of resistance, such as gaming metrics or finding ways to slow down the system.

In our typology, call center systems occupy the end of the control-oriented quadrant, with very high control intensity, strong ties to sanctions, and minimal worker participation. Engagement is undermined rather than supported, and labor relations are marked by latent or overt conflict. AI-powered feedback, in this context, becomes a vector for digital Taylorism.

6.4 Scenario 4: Co-governed AI feedback and worker technology rights

While empirical examples of co-governed AI feedback systems remain rare, recent reports on worker technology rights articulate principles that could guide their design (Bernhardt et al., 2021). Building on these proposals, we sketch a scenario in which:

- unions and worker representatives are involved from the outset in defining the goals of the AI feedback system;
- joint committees negotiate which data can be collected, with privacy and proportionality safeguards;
- workers have access to their own data and analytics, not only aggregate or managerial views;
- clear rules specify when and how feedback can be used in decisions about pay, promotion, and discipline;
- grievance procedures allow workers to challenge inaccurate or unfair algorithmic evaluations.

In this scenario, the orientation of feedback is explicitly dual: supporting both organizational performance and sustainable work. Control intensity may still be moderate to high, but negotiated norms and mutual transparency bound its exercise. We classify this as a co-governed AI feedback type, with the potential to strengthen trust and engagement if implemented in good faith.

7 Quantitative Indicators and Analytic Strategy

To demonstrate how AI-powered feedback can *in principle* be studied empirically, we outline a set of quantitative indicators and models that can be implemented using the prototype described earlier. At this stage, the Jubo system has been deployed as an architectural and qualitative proof-of-concept; the indicators and models below are designed to be plugged into existing numeric HR analytics pipelines in future work. They also serve as reusable templates for other organizations that wish to combine LLM-based text analysis with standard engagement and turnover metrics.

7.1 Control intensity and worker participation indices

At the system or team level, we construct two indices on a 0–10 scale that summarize key features of AI-powered feedback deployments.

Control Intensity Index (CI). This index aggregates scores on:

- data collection frequency (annual to real-time);
- data granularity (team-level aggregates to individual click-level logs);
- strength of linkage between metrics and sanctions or rewards;
- degree of automated intervention (from informational dashboards to automated disciplinary triggers).

Each component is scored on a small ordinal scale and rescaled to produce an overall CI ranging from 0 (minimal control extension) to 10 (maximal, real-time, sanction-linked control).

Worker Participation Index (WPI). This index captures:

- whether workers or unions were consulted before deployment;
- whether a joint committee exists to oversee the system;
- whether worker representatives have review or veto rights over metrics and decision rules;
- whether workers can access and correct their own data;
- whether limits on function creep are codified.

Again, the components are scored and rescaled to yield a 0–10 index, with higher values indicating stronger participation in governance.

These indices underlie the placement of cases in Figure 2 and can be adapted for use in future surveys or comparative field studies. In our prototype, values for CI and WPI are specified by the research team in consultation with management and worker representatives, but they could also be collected through structured questionnaires.

7.2 Engagement and psychological safety

In a full quantitative deployment, we propose to measure engagement and psychological safety using standard Likert scales administered before and after the introduction of AI-supported feedback. The AI agents described in Section 4 can then be used to derive complementary indicators from free-text feedback (e.g., development orientation and perceived control), which can be linked to survey outcomes.

Basic analyses would include:

- **Paired comparisons** of average engagement and psychological safety scores for the same individuals before and after the system's introduction, using paired t-tests or non-parametric alternatives.
- **Linear models** of the form:

$$E_{it} = \alpha + \gamma\,\text{Post}_t + \delta\,\text{DevScore}_i + \lambda\,\text{ManagerSupport}_i + \epsilon_{it},$$

  where E_{it} denotes engagement for employee i at time t, Post_t is an indicator for post-introduction surveys, DevScore_i is the development orientation score derived from LLM-based coding, and ManagerSupport_i captures perceived managerial support. A positive and significant γ would indicate an overall gain beyond composition effects; a positive δ would suggest that receiving more developmentally framed feedback is associated with higher engagement.

The Jubo prototype is already capable of computing DevScore_i in near real time after each review cycle. What is currently missing—and is left for a future data collection phase—is the systematic pairing of these indicators with repeated engagement and safety surveys.

7.3 Turnover risk: logistic and survival models

For turnover, we consider both binary and time-to-event analyses. These specifications are intentionally generic so that they can be reused across firms and datasets.

Logistic regression. We model the probability that employee i leaves within a fixed window (e.g., 12 months) as:

$$\Pr(\text{Turnover}_i = 1) = \text{logit}^{-1}\big(\beta_0 + \beta_1\text{DevScore}_i + \beta_2\text{CI}_{\text{team}(i)} + \beta_3\text{Tenure}_i + \beta_4\text{Age}_i + \dots\big),$$

where $\text{CI}_{\text{team}(i)}$ denotes the control intensity index for employee i's team. A negative β_1 would indicate that higher developmental orientation is associated with lower turnover risk, while a positive β_2 would indicate higher risk in more tightly controlled environments.

Survival analysis. When finer-grained timing data are available, we can estimate a Cox proportional hazards model:

$$h(t \mid X_i) = h_0(t) \exp(\theta_1 \text{DevScore}_i + \theta_2 \text{CI}_{\text{team}(i)} + \theta_3 \text{WPI}_{\text{firm}} + \dots),$$

where $h(t \mid X_i)$ is the instantaneous hazard of leaving at time t for employee i. A negative θ_1 would mean that higher developmental orientation reduces the hazard of leaving; a positive θ_2 would suggest that stronger control increases it.

These models mirror the types of regressions already used in prior numeric-only analyses of turnover in similar settings. The added contribution of the present architecture is that LLM-derived variables, such as DevScore_i and perceived control, can be computed from text and incorporated into the same modeling framework without altering the underlying statistical machinery.

Difference-in-differences demonstration. If future deployments create quasi-experimental conditions—for example, when some units adopt AI-supported feedback earlier than others—the prototype can support difference-in-differences designs, using pre/post and treated/untreated comparisons of engagement or turnover outcomes. In that case, the agent-generated indicators would serve as mediators or moderators in standard difference-in-differences specifications.

Overall, this analytic strategy is intended as a bridge between qualitative insight and quantitative evaluation: the same AI agents that generate real-time, narrative feedback for users also produce structured indicators that can be analyzed with familiar econometric tools.

8 Discussion

8.1 AI feedback as double-edged infrastructure

Across the cases, AI-powered performance feedback emerges as a double-edged infrastructure. On the one hand, systems like Jubo's can make feedback more timely, multi-directional, and developmental, supporting engagement by helping workers see how their contributions matter. On the other hand, digital performance management and call center systems show how similar technologies can intensify control, create continuous pressure, and erode trust.

Figure 2 visualizes these contrasts in a two-dimensional space, with control intensity on the horizontal axis and worker participation in governance on the vertical axis. The development-oriented, control-oriented, and co-governed ideal types occupy distinct regions of this space, and the three empirical cases cluster in different quadrants. Jubo sits in a medium-control, low-to-moderate participation zone; the manufacturing case illustrates higher control with limited voice; and the call center example approaches the extreme of high power and minimal participation. The co-governed configuration, by contrast, is located in the upper-mid region of the figure and remains largely aspirational in our data.

A key insight from this mapping is that the same technical capabilities, data aggregation, pattern recognition, real-time dashboards, can be assembled into very different governance regimes. What matters for labor relations is less the presence of AI per se than the orientation, control intensity, and degree of worker participation in system design and oversight. The two-dimensional space offers a heuristic for situating emerging systems and identifying directions of change. For example, a developmental but weakly governed system like Jubo's could move upward toward the co-governed zone if unions or worker councils gain a formal role in governance, without necessarily increasing control intensity.

8.2 Implications for engagement and voice

Our analysis suggests several mechanisms linking AI feedback to engagement and voice:

- **Clarity of expectations**: When metrics and feedback are co-designed and communicated as developmental, they can reduce ambiguity and support self-regulation.

- **Perceived fairness**: Multi-source and transparent feedback, with access to underlying data, can enhance perceptions of procedural justice. Opaque scoring and one-sided data access have the opposite effect.
- **Psychological safety**: High-frequency feedback can either foster a learning climate or create a sense of surveillance, depending on whether mistakes are treated as learning opportunities or grounds for sanction.
- **Channels for contestation**: Engagement is more sustainable when workers can challenge and reshape AI systems, for example, through union involvement or joint committees.

In Jubo, the developmental framing and decoupling of 360-degree feedback from pay contributed to relatively positive engagement outcomes, despite underlying concerns about potential repurposing. In the manufacturing and call center cases, tight coupling of AI metrics to sanctions and limited opportunities to contest them contributed to stress and potential disengagement. Systems located in the lower-right region of Figure 2 (high control, low participation) thus face a structural disadvantage in sustaining engagement, even if they deliver efficiency gains.

8.3 Implications for labor relations practice

For unions and worker advocates, AI-powered feedback systems present both challenges and opportunities. On the one hand, they can be seen as new instruments of control that require regulatory and contractual safeguards. On the other hand, if negotiated effectively, AI systems could provide workers with better information about their conditions and support more evidence-based bargaining.

Our typology highlights several levers for intervention:

- insisting on transparency about data sources, algorithms, and decision rules;
- negotiating limits on data collection and use, particularly for discipline;
- creating joint governance structures for AI systems, with access to documentation and the ability to commission audits;
- ensuring that workers receive not only more feedback but also more influence over how feedback is used.

For employers and HR practitioners, the key message is that AI feedback systems should be treated as socio-technical infrastructures rather than plug-and-play tools. Their design should involve not only technical and HR staff, but also worker representatives and, where appropriate, regulators. The prototype described in this paper illustrates how such systems can be built to support, rather than replace, these forms of dialogue.

9 Conclusion

AI-powered performance feedback systems are rapidly moving from experimental pilots to core elements of performance management. This paper has argued that their impact on workforce engagement and labor relations depends on how they are oriented (development vs. discipline), how intensively they extend control, and whether workers and their representatives have a role in governing them.

By comparing a development-oriented startup implementation, a control-oriented digital performance management approach, a highly monitored call center environment, and a co-governed scenario grounded in worker technology rights, we have shown that AI feedback is not destiny. It can support learning and engagement, or it can deepen surveillance and conflict.

Methodologically, we have demonstrated that concepts such as development orientation, control intensity, and governance participation can be operationalized using a combination of LLM-based text analysis, multi-agent orchestration, and automated data collection. The resulting indicators can be linked to engagement and turnover outcomes using standard statistical models, opening the way for more systematic empirical research on AI in performance management.

For labor and employment scholars, the framework proposed here invites further empirical work on how AI feedback systems are negotiated in different institutional contexts, how they intersect with existing collective bargaining structures, and how regulatory regimes shape their trajectory. For practitioners, it suggests that the promise of AI-powered feedback, more timely, actionable insights into performance, will only be realized if systems are designed with engagement, voice, and fairness at their core.

References

Bernhardt, A., Kresge, L., and Suleiman, R. (2021). *Data and Algorithms at Work: The Case for Worker Technology Rights*. UC Berkeley Labor Center, Berkeley, CA.

Cappelli, P. and Tavis, A. (2016). The performance management revolution. *Harvard Business Review*, 94:58 – 67.

Chartered Institute of Personnel and Development (2024). People analytics: Factsheet. CIPD Factsheet.

Christl, W. (2023). *Surveillance and Algorithmic Control in the Call Center: A Case Study on Contact and Service Center Software, Automated Management and Outsourced Work*. Cracked Labs, Vienna.

Cianci, A. M., Klein, H. J., and Seijts, G. (2010). The effect of negative feedback on tension and subsequent performance: The main and interactive effects of goal content and conscientiousness. *Journal of Applied Psychology*, 95:618 – 630.

Giermindl, L., Strich, F., Christ, O., Leicht-Deobald, U., and Redzepi, A. (2021). The dark sides of people analytics: reviewing the perils for organisations and employees. *European Journal of Information Systems*, pages 1 – 26.

Lee, M. K., Kusbit, D., Metsky, E., and Dabbish, L. (2015). Working with machines: The impact of algorithmic and data-driven management on human workers. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, pages 1603–1612, New York, NY. ACM.

Rosenblat, A. and Stark, L. (2016). Algorithmic labor and information asymmetries: A case study of uber's drivers. *International Journal of Communication*, 10:3758–3784.

Stefano, V. D. and Taes, S. (2022). Algorithmic management and collective bargaining. *Transfer*, 29:21 – 36.

Subhadarshini, S., Nayak, A., Biswal, S. N., and Choudhury, S. (2024). The future of performance management: Leveraging ai for better feedback and coaching. *Journal of Informatics Education and Research*.

Vhora, M. A., Bhandwalkar, V., and Rege, P. M. (2024). Ai-driven hr analytics: Enhancing decision-making in workforce planning. *The Scientific Temper*, 15:3299 – 3308.

Zhou, N., Han, X., Hu, T., Zhou, Y., and Jiang, W. (2018). Digital performance management based on the perspective of organizational control: A case study of advantech. In Pan, S. L. and Sandeep, M. S., editors, *Digital Enablement: The Consumerizational and Transformational Effects of Digital Technology*, pages 83–93. World Scientific Publishing, Singapore.