Humanizing People Analytics: Designing Metrics That Empower, Not Dehumanize

Executive Summary

Drawing primarily on critiques from Harvard Business Review[1] and research from MIT Sloan School of Management[2] and MIT Center for Information Systems Research[3], this whitepaper argues that the core failure mode in many people-analytics products is not measurement itself; it is **thin measurement**. Thin measurement happens when organizations confuse what is easy to count with what is meaningful to understand. HBR has warned that workers are increasingly defined in terms of their data, that metric obsession can undermine strategy, that monitoring can backfire, and that conventional performance evaluations remain bias-prone. MIT research, by contrast, frames talent analytics as a data-driven approach that should improve decisions for both the organization and individual employees, while MIT CISR argues that organizations need ethical and values-based "acceptable data use," not just legal compliance. [4]

The two most important downstream harms are **metric fatigue** and **identity reduction**. In this report, *metric fatigue* means the cognitive, emotional, and behavioral depletion created by frequent requests to report, react to, or optimize against metrics without visible payoff. The evidence base is consistent: when survey programs are not followed by action, employees disengage, response rates drop, trust erodes, and data quality deteriorates; in longer questionnaires, careless responding rises as people progress through the instrument. *Identity reduction* is the design condition in which a person becomes legible mainly as a score, proxy, or instrument. The workplace objectification and organizational dehumanization literatures describe this as being treated like a tool, denied subjectivity, or made interchangeable; studies link it to lower belonging, worse well-being, negative job attitudes, and even mechanistic self-dehumanization. [5]

The product implication is straightforward: people analytics should shift from **scoreboards to sensemaking**, from **ranking to growth framing**, and from **one-way extraction to reciprocal participation**. Metrics should be contextual, privacy-bounded, action-linked, and narratively explained. Sensitive or high-stakes analytics should move through governance gates that test purpose, necessity, consent, worker participation, fairness, and reversibility. Research from MIT on sustainable productivity, worker voice, and dignity, together with governance guidance from the National Institute of Standards and Technology[6], the Information Commissioner's Office[7], and the OECD[8], supports a design model in which employees can understand what is collected, why it matters, how it is used, what actions follow, and how they can challenge or shape the system. [9]

The most rigorous operating principle for product and HR leaders is this: **if a metric cannot be explained, contested, or improved through human action, it should not be used to**

judge human beings. That principle is the through-line for the design patterns, governance model, KPI map, and experiment templates that follow. [10]

The Critique of Traditional People Analytics

The critique from HBR and MIT is not anti-analytics; it is anti-reductionism. MIT defines people analytics as improving people-related decisions for the success of both the organization and employees, and warns that bad variable selection, poor data quality, dependent-variable selection, and vague definitions of "high performer" lead to poor recommendations. HBR's critique is complementary: workers are increasingly defined by data traces, obsession with metrics can sink strategy, and rationality or meritocracy cannot simply be assumed because a system is data-driven. [11]

That critique matters because many people-analytics stacks are still built around three brittle assumptions. The first is that more monitoring produces more truth. The second is that a proxy for performance is equivalent to performance itself. The third is that the visibility of a metric is enough to mobilize useful action. Yet the literature on worker productivity stresses that performance is multidimensional and that poorly chosen measures distort incentives and conclusions. HBR's performance-management pieces underscore the same point from another angle: performance review systems remain vulnerable to bias even before algorithmic layers are added. [12]

A human-centered whitepaper therefore should start from a stronger proposition: **people analytics is a socio-technical sensemaking system, not a neutral mirror**. Organizational analytics does not simply "find" meaning in behavior; it produces meaning through choices about inputs, categories, transformations, and audiences. That is why Goodhart dynamics are not edge cases. When workers know what is measured, they adapt their signaling, and the metric itself changes meaning. [13]

The practical consequence is that a good people-analytics product must do four things at once. It must preserve what metrics are good at such as consistency, scale, and early detection. It must also preserve what metrics are bad at replacing such as judgment, dialogue, context, and dignity. It must avoid collapsing development into surveillance. And it must separate **measurement for learning** from **measurement for control** unless there is a clearly justified, governed reason to combine them. [14]

Metric Fatigue and Identity Reduction

In this report, **metric fatigue** is a design pathology with three components: reporting burden, interpretive burden, and optimization burden. Reporting burden is the effort required to answer surveys, pulses, check-ins, or assessments. Interpretive burden is the demand to constantly parse dashboards and trend lines. Optimization burden is the pressure to continuously self-correct against those signals, often without the authority or resources to do so. This is a broader and more workplace-relevant concept than survey fatigue alone, but the survey literature makes the mechanism visible: repeated surveying without meaningful follow-up undermines participation, trust, and response quality. [15]

Identity reduction is the human cost of thin measurement. I use the term to synthesize the objectification, dehumanization, and dignity literatures. HBR describes the shift starkly: organizations once made up of people increasingly treat workers as data. MIT CISR argues that employee data must be governed through dignity because the employer-employee relationship creates ethical complexity that privacy law alone does not resolve. The organizational dehumanization literature defines the problem even more precisely: employees experience being objectified, denied personal subjectivity, and made to feel like tools for organizational ends. [16]

The table below uses those literatures to distinguish the two harms.

Failure mode	Working definition used here	Typical causes	Observable symptoms	Product-level implication
Metric fatigue	Depletion caused by repeated measurement demands and weak follow-through	Too many surveys, unclear purpose, dashboard overload, no closure loop, action without support	Lower response rates, higher careless responding, cynicism, less trust, manager overload	Reduce cadence, collapse redundant instruments, attach every metric to an owner and action path
Identity reduction	Treating a person primarily as a proxy, score, or instrument	Output-only measurement, proxy fixation, hidden sensing, ranking language, individualized surveillance	Lower belonging, reduced autonomy, defensive behavior, self-dehumanization, gaming	Reframe metrics for growth, preserve subjectivity through narrative context, limit judgmental uses

These harms arise from several recurrent mechanisms. One is **activity-output confusion**: organizations are better at measuring outputs such as sales, inventory, or customer satisfaction than inputs such as load, friction, and well-being. MIT's sustainable-productivity work warns that employees can easily interpret activity monitoring as performance evaluation or behavior policing, which is exactly how metrics start to colonize identity. A second mechanism is **proxy fixation**: the more a measure becomes a target, the less faithfully it represents the work itself. A third is **one-way extraction**: when organizations gather data but do not close the loop with action, employees conclude that the system is for the organization, not for them. [17]

The empirical case for metric fatigue is stronger than it first appears. In the pulse-survey study from Frontiers, employees' willingness to respond depended heavily on whether their input was used meaningfully and followed by visible action; response rates of 80–90% were achievable under strong managerial follow-up, but researchers also found that weak follow-up could deteriorate the situation, leading to disengagement, survey fatigue, and erosion of trust. The same study identified ethical problems around identifiability in small groups and the burden of being constantly evaluated, especially for line managers. The broader survey-methods literature shows that longer questionnaires reduce willingness to participate and can degrade response quality, while careless responding tends to increase over the course of completion. [18]

The empirical case for identity reduction is also strong. Research on workplace objectification finds that people objectify others more in work contexts than non-work contexts and attributes that pattern to more calculative and strategic thinking. The organizational dehumanization literature links feeling treated like a tool to poorer well-being and workplace outcomes. One three-study article further shows that organizational dehumanization can lead to mechanistic self-dehumanization through increased surface acting. Put simply: once the organization treats people like instruments, people start managing themselves like instruments. [19]

Passive sensing makes both problems more acute if it is deployed carelessly. A recent survey of workplace passive sensing shows how wearables and smartphones can be used to assess well-being and productivity through physiological and contextual data, while emphasizing unobtrusiveness and worker acceptance as design constraints. MIT CISR's 5-W framework is useful here: people-analytics systems now routinely collect data about who employees are, what they do, where they are, when events occur, and why they decide or feel as they do. That richness can create more supportive systems, but it also creates more ways to collapse a person into a machine-readable object. [20]

Recent practitioner cases show how quickly systems drift into dehumanizing territory when governance lags product capability. Microsoft[21] changed its Productivity Score product after backlash, removing usernames entirely and stating that the feature should no longer be usable to monitor individual employees. This is a useful reminder that even products framed as adoption or productivity tooling are often interpreted first as surveillance systems. [22]

Barclays[23] changed a pilot that tracked how employees spent time at work after criticism, later stating that it would use anonymized data only. The important lesson is not simply "anonymize"; it is that individual-level activity visibility for managers is itself a product choice that changes the social meaning of measurement. [24]

Serco[25] was ordered by the UK regulator to stop using facial recognition and fingerprint scanning to monitor staff attendance after an investigation found unlawful biometric processing affecting more than 2,000 workers across 38 facilities. The official guidance from the ICO reinforces the broader principle: biometric systems require a lawful basis

plus a separate condition for processing special-category data, and explicit consent is often the most appropriate route, depending on context and proportionality. [26]

Amazon[27] provides the clearest example of proxy fixation becoming bodily harm. The U.S. Senate HELP Committee report describes a corporate culture obsessed with speed and productivity, argues that quotas and discipline pressures are real even when publicly obscured, and concludes that the company manipulated its workplace injury narrative to present warehouses as safer than they were. This is what happens when the organization optimizes the metric story rather than the human system. [28]

Product Design Patterns for Humanized Analytics

The strongest product translation from the literature is an inference: **every metric that could be read as judgment should be redesigned as a coaching signal unless there is a narrowly justified reason not to**. MIT's sustainable-productivity research argues that activity data should support employees, be transparent, and lead to workplace improvement. MIT worker-voice research shows that input from workers increases the likelihood that new tools improve job quality and are used effectively. HBR's employee-feedback work adds a crucial condition: the value of listening decays when there is a gap between collecting input and acting on it. [29]

That design translation leads to two headline patterns.

Growth framing means showing movement, learning, friction, and support needs rather than rank, deficiency, or suspicion. The default comparison should be self-to-self over time, then team or role benchmarks with privacy thresholds, not person-to-person leaderboards.

Narrative insights means that the interface tells a short evidence-linked story: what changed, what may be driving it, how certain the signal is, what action is possible, and what caveats apply. Narratives make subjectivity explicit instead of hiding it behind the false precision of a single score.

The table below translates those ideas into concrete features, UX patterns, data-model objects, and examples.

Design problem	Humanized feature	UX pattern	Minimum data-model objects	Example
Workers experience dashboards as rankings	**Growth trajectory card**	Self-baseline, trend band, confidence range, explicit "for development only" label	`MetricDefinition`, `Observation`, `BaselineWindow`, `ConfidenceEstimate`, `PolicyFlag`	"Your after-hours coordination load is up versus your 12-week baseline; most of the

Design problem	Humanized feature	UX pattern	Minimum data-model objects	Example
				increase comes from Tue/Thu recurring meetings.”
Metrics explain nothing	**Narrative insight card**	Brief summary, likely drivers, action suggestions, caveats, freshness timestamp	`NarrativeInsight`, `EvidenceLink`, `Driver`, `Caveat`	“The shift is associated with meeting density and not with lower task completion.”
Listening without follow-up	**Closure loop feed**	“You said / We did / What’s next” timeline with owner and due date	`IssueTheme`, `ActionPlan`, `OwnerRole`, `DueDate`, `StatusEvent`	“Focus-time concern raised in March; calendar policy changed in April; team review due in May.”
Consent buried in legal text	**Consent center**	Purpose-by-purpose controls, retention display, export/delete/request review	`ConsentRecord`, `Purpose`, `RetentionPolicy`, `NoticeVersion`, `WithdrawalEvent`	“Calendar metadata for workload balancing: enabled for 90 days, self-view + team aggregate only.”
Small groups can be identifiable	**Privacy thresholding**	Minimum group size, suppression, noise, banded results	`SuppressionRule`, `GroupSize`, `AggregationLevel`	“No subgroup view when n < 7.”
Managers use proxies as verdicts	**Decision-use badge**	Visible badge that states whether a metric may be used	`UseRestriction`, `DecisionContext`, `ReviewerRole`	“Burnout risk signal: triage and

Design problem	Humanized feature	UX pattern	Minimum data-model objects	Example
		for coaching, triage, or formal decisions		support only; cannot be used for performance rating."
People contest opaque outputs	**Challenge and appeal workflow**	Explanation, review request, second-level reviewer, audit trail	`AppealCase`, `Explanation`, `Reviewer`, `Outcome`, `TurnaroundTime`	"Request human review of this promotion-readiness inference."
High-sensitivity sensing is normalized too early	**Sensing sandbox**	Time-boxed pilots, explicit opt-in, self-view first, automatic sunset	`PilotEnrollment`, `SensorStream`, `ExitSurvey`, `SunsetDate`, `HarmIncident`	"Opt-in wearable pilot for focus and overload self-insights runs 6 weeks and auto-terminates absent reapproval."

A minimum viable human-centered data schema should make **use restrictions, consent, narrative, follow-up, and harms** first-class entities, not afterthoughts. MIT CISR's 5-W view is especially helpful for modeling this because it prevents the common trap of storing "people data" as unstructured exhaust while leaving policy and meaning outside the system. The schema below is a recommended starter design inferred from that literature and from trustworthy-AI governance guidance. [30]

```
Person:
  person_id
  team_id
  role_family
  manager_id
  demographics_ref: separate, access-controlled

MetricDefinition:
  metric_id
  name
  human_goal
  scope: self | team | org | subgroup
  decision_use_allowed: coaching | triage | formal_decision | prohibited
```

```
  min_group_size
  refresh_cadence
  owner_role
  sunset_date

Observation:
  metric_id
  subject_scope_id
  period_start
  period_end
  value
  confidence
  source_ref

ConsentRecord:
  person_id
  data_category
  purpose_id
  status: opted_in | opted_out | required_notice
  lawful_basis
  start_at
  expire_at

NarrativeInsight:
  insight_id
  metric_id
  summary
  drivers[]
  caveats[]
  recommended_actions[]

ActionPlan:
  action_id
  metric_id
  owner_role
  due_date
  status
  closure_note

AppealCase:
  case_id
  person_id
  metric_id
  decision_context
  reviewer_role
  outcome
  resolved_at

HarmIncident:
```

```
incident_id
metric_id
harm_type
severity
mitigation
resolved_at
```

The accompanying UX rule should be equally explicit: **show context before score, and action before judgment**. A poor interface says, “Bottom 10% on collaboration.” A humanized interface says, “Meeting load rose 18% above your baseline; the increase came from two recurring coordination meetings. Consider protecting two focus blocks next week and reviewing meeting ownership with your manager.” The first design collapses identity into a percentile. The second preserves agency, cause, and next step.

Participation, Governance, and Ethical Checkpoints

Participation is not a “communications” layer added after the product is built. MIT’s worker-voice research, OECD guidance on algorithmic management, and work from the International Labour Organization[31] all point in the same direction: workers should help define the problem, codesign technical and work-process features, shape education and retraining, and participate in transition planning for affected roles. In the OECD’s 2025 policy brief, the introduction of algorithmic-management tools should be guided by worker consultation to aid adoption and safeguard worker well-being. [32]

The practical question is not whether participation is desirable, but **which participation model fits which use case**.

Participation model	Appropriate uses	Consent pattern	Governance requirement
Notice and consultation	Low-sensitivity operational metrics already needed to run the business, such as aggregate meeting load or team-level workflow friction	Clear notice, purpose limitation, team consultation, no hidden reuse	Purpose review, role-based access, periodic revalidation
Developmental opt-in	Individual coaching, self-view dashboards, well-being or growth nudges	Opt-in with real non-participation alternative and no retaliation	Self-view default, no employment-decision use, easy withdrawal
High-sensitivity co-design	Biometrics, passive sensing, emotion inference, model-based decisions affecting work	Explicit opt-in where possible; if voluntariness is doubtful, do not deploy until necessity and	Worker panel or representative body, legal/privacy review, harm testing, appeals, sunset

Participation model	Appropriate uses	Consent pattern	Governance requirement
	conditions or opportunity	proportionality are established	clause

The governance model below operationalizes the literature into a product-delivery flow. It translates dignity, worker voice, trustworthiness, and data-protection requirements into stage gates that should exist before launch and after launch. [33]

```
flowchart TD
    A[Metric idea or new data source] --> B[Purpose and necessity review]
    B --> C[Worker consultation or co-design session]
    C --> D[Data suitability and privacy assessment]
    D --> E[Fairness and harm pre-mortem]
    E --> F[Pilot with opt-in or bounded notice]
    F --> G[Post-pilot review]
    G -->|approve| H[Limited deployment]
    G -->|revise| C
    G -->|reject| X[Do not launch]
    H --> I[Always-on feedback and appeal channel]
    I --> J[Quarterly audit: value, harms, drift, consent]
    J -->|continue| K[Revalidate]
    J -->|sunset| X
```

A rigorous governance checklist should answer seven questions before any metric ships. What exact decision or support action is this metric for. Why is this data necessary rather than merely available. Who helped define the problem and review the design. What are the plausible harms, especially for small groups or protected subgroups. Can a person understand and contest the output. What action follows if the metric moves. When does the metric expire if it does not demonstrate human benefit. These questions are not bureaucracy. They are product requirements for dignity-preserving systems. [34]

KPI Mapping for Human-Centered Analytics

The KPI set below is intentionally **human-goal first**. It is not a generic HR dashboard. Each KPI is meant to anchor product and operating decisions in the goals emphasized by the source base: dignity, autonomy, meaningful follow-up, fairness, sustainable productivity, and worker trust. These are recommended metrics inferred from the literature, not a claim that any single source endorses one universal scorecard. [35]

Human-centered goal	Suggested KPI	Data sources	Example formula	Potential harms	Mitigations
Preserve dignity and respect	**Dignity index**	Quarterly pulse; incident data;	Mean of validated items on respect,	Becomes another abstract score	Pair with narrative themes and mandatory

Human-centered goal	Suggested KPI	Data sources	Example formula	Potential harms	Mitigations
		appeals	subjectivity, fairness, and non-interchangeability	detached from action	owner for low scores
Increase autonomy	**Autonomy support score**	Pulse items; workflow-change logs	Avg. autonomy item score + % employees reporting influence over how work is done	Managers pressure teams to "look autonomous" without real control	Audit against actual policy changes and local decision rights
Make analytics developmental, not punitive	**Developmental-use ratio**	Audit logs; product config	Development-only metric views / total metric views	Teams may relabel punitive uses as "development"	Independent review of decision contexts and spot audits
Reduce overload without hiding output	**Sustainable productivity gap**	Calendar metadata; task systems; self-report	z(after-hours load + meeting fragmentation + overload self-report) minus z(core outcome attainment)	Hidden surveillance concerns	Use aggregation, bounded retention, visible purpose, opt-in where appropriate
Close the loop on employee voice	**Action closure rate**	Issue tracker; pulse themes; town halls	Priority actions completed within SLA / priority issues raised	Superficial "checkbox" actions	Require outcome evidence and worker validation of closure
Improve fairness in consequential decisions	**Parity gap**	Promotion, pay, ratings, hiring data	Max subgroup outcome gap after legitimate controls	Reidentification or over-interpretation in small cells	Minimum cell sizes, audited controls, human review, public caveats
Build trust in data use	**Trustworthy-use score**	Survey; consent	Mean of trust items on	Trust surveys can	Keep short, lower cadence,

Human-centered goal	Suggested KPI	Data sources	Example formula	Potential harms	Mitigations
		events; access logs	transparency, understanding, and confidence in use constraints	themselves fuel fatigue	publish “you said / we did”
Strengthen worker voice	**Voice efficacy rate**	Suggestion system; focus groups; product backlog	Accepted employee-originated changes / submitted viable proposals	Overrewards “squeaky wheels”	Track representativeness by role, shift, and location
Protect privacy and voluntariness	**Sensitive-data opt-in quality**	Consent records; withdrawal events	Opt-in rate * comprehension check pass rate * low withdrawal-friction score	Coerced opt-in	Require non-participation alternative and monitor downstream disadvantage
Make manager follow-up credible	**Manager follow-through score**	Action plans; direct-report survey; meeting notes	Weighted index of follow-up timeliness, visibility, and employee-rated usefulness	Penalizes overwhelmed managers	Pair with support resources and workload-adjusted expectations
Ensure contestability	**Appeal responsiveness**	Appeal workflow	Median days to review + % appealed cases receiving explanation	Appeal volume may be read as system failure only	Treat as learning signal; analyze root causes rather than suppress complaints
Prevent stale or purposeless metrics	**Sunset compliance rate**	Metric registry	Metrics reviewed or retired on schedule / metrics due for review	Teams keep low-value metrics alive by default	Automatic expiry unless reapproved with evidence of value

A strong applied rule is to keep **leading human indicators** and **lagging business outcomes** on the same page, but never collapse them into one composite score. The business temptation is to combine everything into a single "people health" number. The better design is a linked view where human goals remain visible in their own right and where leaders can see trade-offs instead of hiding them. That design choice directly reduces the risk that productivity gains are purchased through invisible harm. [36]

Experiment Templates and Evaluation

Experimentation in human-centered people analytics should test not only whether a feature increases engagement or prediction accuracy, but whether it does so **without increasing burden, unfairness, or coercion**. HBR's work on employee feedback shows that the value of listening collapses when action is weak. MIT's worker-voice work argues that adoption improves when workers participate in design. NIST's AI RMF suggests that trustworthy systems need to be transparent, explainable, privacy-enhanced, fair, and accountable. Those principles imply that experiment plans need explicit harm metrics, not just success metrics. [37]

The templates below are intended as production-ready starting points. Sample sizes are illustrative heuristics and should be replaced by formal power analysis using your baseline rates, intraclass correlations, and acceptable minimum detectable effects.

Template	Hypothesis	Design and sample size heuristic	Duration	Primary metrics	Analysis plan	Ethical safeguards
A/B test of growth framing vs score-only dashboard	A dashboard with self-baselines and narrative insights improves action completion and perceived dignity without hurting outcome attainme	Individual randomization; start with ~400 per arm for small continuous effects around d≈0.20, or ~800 per arm for modest binary lifts	6–8 weeks	Action completion, perceived autonomy, perceived dignity, opt-out rate, appeal rate, target business outcome	Intent-to-treat; mixed-effects regression; pre-registered subgroup checks by role, tenure, and manager	No employment decisions tied to variant; self-view default; plain-language notice; accessible appeal channel

Template	Hypothesis	Design and sample size heuristic	Duration	Primary metrics	Analysis plan	Ethical safeguards
Cluster randomized trial of manager follow-up workflow after pulse surveys	A structured follow-up workflow increases response quality and trust while reducing survey fatigue	Team-level randomization; start with ~40–60 teams per arm depending on team size and ICC	8–12 weeks	Response rate, closure rate, trust score, duplicate-issue recurrence, manager burden	Hierarchical models at team level; difference-in-differences from baseline; qualitative review of failed follow-ups	Minimum group size protections, no small-cell release, manager support resources, escalation for serious comments
Qualitative pilot of sensitive sensing or biometric/self-regulation feature	Time-boxed, opt-in, self-view-first pilots reveal usefulness and harms better than broad rollout	20–40 volunteers across roles; diaries, interviews, opt-in telemetry	4–6 weeks	Usefulness, dignity, perceived intrusion, withdrawal reasons, unexpected harms	Thematic analysis plus descriptive telemetry; triangulate with opt-out and complaint data	Explicit opt-in, no manager access to individual data, automatic deletion at pilot end absent renewed consent
Randomized encouragement design for consent language and purpose explanation	Better purpose explanations improve informed participation quality more than	Randomize onboarding copy or walkthroughs; ~500–1,000 exposures per arm if feasible	2–4 weeks	Comprehension check score, opt-in quality index, later withdrawal rate, trust score	Logistic/ordinal regression; compare quality, not just opt-in volume	Ban manipulative dark patterns; easy decline path; monitor for differential uptake by subgroup

Template	Hypothesis	Design and sample size heuristic	Duration	Primary metrics	Analysis plan	Ethical safeguards
	persuasive consent copy					
Stepped-wedge rollout of appeals and explanation tooling	Adding explanation and challenge pathways improves trust and reduces perceived arbitrariness over time	Sequential rollout across business units; each unit acts as its own control	10–16 weeks	Appeal responsiveness, trust score, manager override rate, fairness complaints	Segmented time-series or stepped-wedge mixed models	Independent reviewer pool, service-level targets, documentation retention rules

The experiment timeline below is a strong default for any higher-stakes people-analytics feature. It emphasizes baseline measurement, bounded pilot duration, qualitative debriefing, and explicit harm review rather than "ship and watch." That sequence is more defensible ethically and usually produces better product learning. [38]

```
gantt
    title Human-centered people analytics experiment timeline
    dateFormat  YYYY-MM-DD
    axisFormat  %b %d

    section Design
    Purpose, harms, preregistration        :a1, 2026-05-01, 10d
    Worker consultation and stimulus test  :a2, after a1, 10d

    section Baseline
    Baseline measurement                   :b1, 2026-05-21, 14d

    section Trial
    Randomized or staged rollout           :c1, 2026-06-04, 42d
    Ongoing safety and bias monitoring     :c2, 2026-06-04, 42d

    section Review
    Interviews and diary debriefs          :d1, 2026-07-16, 14d
    Statistical analysis                   :d2, 2026-07-16, 14d
    Governance decision                    :d3, 2026-07-30, 7d
```

One final methodological rule is worth stating plainly: **do not let the experiment inherit the product's dehumanizing assumptions**. If a trial compares "more surveillance" to "less surveillance" but does not test dignity, trust, fairness, or burden, it may optimize the wrong thing with greater confidence. That is rigorous experimentation in a narrow statistical sense and poor experimentation in a human-centered product sense. [39]

Open Questions and Limitations

Some of the most relevant HBR and MIT articles are partially paywalled, so this report uses their visible summaries and abstracts conservatively and leans more heavily on accessible primary research, official briefs, and public guidance for substantive claims. The overall pattern is strong, but the field still needs more long-run field experiments that compare humanized designs against conventional dashboards in real organizational settings rather than only in conceptual or advisory form. The pulse-survey literature itself notes that research on implementation and long-term effects remains limited. [40]

Legal requirements also vary across jurisdictions and data types. The durable guidance in this report is therefore principle-based: dignity, necessity, proportionality, worker participation, bounded purpose, explainability, appeal, and sunset review. Those principles travel well across contexts even when the exact lawful basis, labor-law obligations, or consultation requirements differ. [41]

The highest-confidence conclusion is that better people analytics is not mainly a modeling problem. It is a **product-design and governance problem**. Systems become humanizing not when they collect less data by definition, but when they collect data under conditions that preserve subjectivity, strengthen agency, make action reciprocal, and keep measurement answerable to the people being measured. [42]

---

[1] [2] [3] [6] [25] [30] [33] [34] [35] [41] https://cisr.mit.edu/publication/2021_0601_DignityDataUse_TonaLeidnerWixomSomeh

https://cisr.mit.edu/publication/2021_0601_DignityDataUse_TonaLeidnerWixomSomeh

[4] [16] [42] https://hbr.org/2022/06/are-people-analytics-dehumanizing-your-employees

https://hbr.org/2022/06/are-people-analytics-dehumanizing-your-employees

[5] [15] [18] [27] [40] https://www.frontiersin.org/journals/organizational-psychology/articles/10.3389/forgp.2025.1696769/full

https://www.frontiersin.org/journals/organizational-psychology/articles/10.3389/forgp.2025.1696769/full

[7] [26] https://www.reuters.com/world/uk/uk-watchdog-orders-serco-stop-using-facial-recognition-monitor-staff-2024-02-23/

https://www.reuters.com/world/uk/uk-watchdog-orders-serco-stop-using-facial-recognition-monitor-staff-2024-02-23/

[8] [9] [14] [17] [29] [36] https://mitsloan.mit.edu/ideas-made-to-matter/case-sustainable-productivity-and-how-to-measure-it

https://mitsloan.mit.edu/ideas-made-to-matter/case-sustainable-productivity-and-how-to-measure-it

[10] [21] [38] [39] https://nvlpubs.nist.gov/nistpubs/ai/nist.ai.100-1.pdf

https://nvlpubs.nist.gov/nistpubs/ai/nist.ai.100-1.pdf

[11] https://mitsloan.mit.edu/ideas-made-to-matter/people-analytics-explained

https://mitsloan.mit.edu/ideas-made-to-matter/people-analytics-explained

[12] https://wol.iza.org/articles/performance-measures-and-worker-productivity/long

https://wol.iza.org/articles/performance-measures-and-worker-productivity/long

[13] https://academic.oup.com/jcmc/article/28/4/zmad023/7210220

https://academic.oup.com/jcmc/article/28/4/zmad023/7210220

[19] [23] https://pubmed.ncbi.nlm.nih.gov/32658521/

https://pubmed.ncbi.nlm.nih.gov/32658521/

[20] https://arxiv.org/html/2201.03074v2

https://arxiv.org/html/2201.03074v2

[22] https://www.microsoft.com/en-us/microsoft-365/blog/2020/12/01/our-commitment-to-privacy-in-microsoft-productivity-score/

https://www.microsoft.com/en-us/microsoft-365/blog/2020/12/01/our-commitment-to-privacy-in-microsoft-productivity-score/

[24] https://www.reuters.com/article/business/barclays-backtracks-on-staff-surveillance-system-after-criticism-idUSKBN20E26L/

https://www.reuters.com/article/business/barclays-backtracks-on-staff-surveillance-system-after-criticism-idUSKBN20E26L/

[28] https://www.help.senate.gov/imo/media/doc/amazon_investigation.pdf

https://www.help.senate.gov/imo/media/doc/amazon_investigation.pdf

[31] [32] https://mitsloan.mit.edu/centers-initiatives/institute-work-and-employment-research/bringing-worker-voice-generative-ai

https://mitsloan.mit.edu/centers-initiatives/institute-work-and-employment-research/bringing-worker-voice-generative-ai

[37] https://hbr.org/2024/11/turn-employee-feedback-into-action

https://hbr.org/2024/11/turn-employee-feedback-into-action