Original Research Article

Maintaining Psychological Safety in the Toyota Production System: The Role of Transparency and Decision-Making in the Age of Artificial Intelligence

Cayden Bunjamin

Design Tech High School, 275 Oracle Parkway, Redwood City, CA, 94065, United States; Lumiere Education

ABSTRACT

The Toyota Production System (TPS) is recognized for fostering continuous improvement and its emphasis on team-based problem-solving within a culture that values open communication and learning. At the same time, artificial intelligence (AI) is rapidly transforming manufacturing by introducing opportunities for augmentation and automation. This study models how AI integration could influence psychological safety in environments inspired by TPS principles, using synthetic data generated by large language models (LLMs). Drawing from literature on organization learning, AI integration, and workplace psychology, a conceptual model is developed and tested across four simulated organizational scenarios: TPS with transparency and participative decision-making (TPS+), TPS without these moderators (TPS−), traditional manufacturing environment with automation-based AI transparency and participative decision-making (Non-TPS+), and a traditional manufacturing environment with automation-based AI without these moderators (Non-TPS−). Quantitative analyses demonstrated significant differences across conditions. Teams in the transparent and participative TPS condition (TPS+) reported substantially higher psychological safety ($M = 5.11$) than those in the Non-TPS− condition ($M = 2.70$), $F(3, 996) = 972.56$, $p < .001$. Similar effects emerged for team performance ($M = 5.54$ vs. 2.56; $F(3, 996) = 1453.27$, $p < .001$) and AI adoption ($M = 4.61$ vs. 3.00; $F(3, 996) = 347.28$, $p < .001$). Chi-square analyses further confirmed significant categorical differences in pay outcomes ($\chi^2(6) = 836.07$, $p < .001$) and job redundancy ($\chi^2(15) = 686.91$, $p < .001$) across scenarios. These findings suggest that transparency and participative decision-making positively moderate AI's impact on team dynamics, producing the consistent performance hierarchy TPS+ > Non-TPS+ > TPS− > Non-TPS−. While the data are synthetic, the results offer preliminary support for the theoretical integration of psychological safety and AI augmentation frameworks within human-centered production systems. Future research should validate these trends through human-subject surveys and ethnographic case studies in real manufacturing contexts.

Keywords: Toyota Production System (TPS); Psychological Safety; Artificial Intelligence (AI); Transparency; Participative Decision-Making; Synthetic Data Simulation

Corresponding author: Cayden Bunjamin, E-mail: Cayden.Bunjamin@gmail.com.

Accepted December 4, 2025
https://doi.org/10.70251/HYJR2348.36781790

INTRODUCTION

The Toyota Production System (TPS) has long been considered the gold standard in production management, renowned for its culture of continuous improvement, also known as "kaizen," lean practices, and employee participation at all organizational levels (1). Central to this framework is the concept of psychological safety, a condition that enables individuals to take interpersonal risks without fear of negative consequences. Harvard researcher Edmondson (2) defines psychological safety as a shared belief that a team is safe to speak up, contribute ideas, and learn from mistakes. This concept is closely aligned with TPS's emphasis on collective problem-solving and learning-by-doing, suggesting that psychological safety forms the social foundation upon which lean processes and continuous improvement thrive.

At the same time, artificial intelligence (AI) usage is on the rise and rapidly transforming the manufacturing environment. AI is being used to automate decision-making processes and time-consuming operational tasks that teams traditionally handled (3). While AI allows for efficiency and data-driven improvement, it creates a new challenge for team dynamics, communication, and perceived agency. Workers may feel less responsible for their outcomes, feel more replaceable, and be hesitant to ask questions about decisions made by AI, all of which can weaken the psychological safety necessary for sustained learning. Wilson & Daugherty emphasize that successful AI-driven redesigns require a balanced partnership between human judgment and algorithmic assistance, ensuring that automation complements rather than replaces human sensemaking and collaboration (3).

This study investigates how varying levels of AI integration, transparency, and participative decision-making influence team psychological safety in scenarios inspired by the Toyota Production System (TPS). Because no direct Toyota data were available, the study simulates these organizational dynamics, utilizing synthetic survey data generated by large language models (LLMs) to explore their theoretical effects. Maintaining psychological safety during AI adoption is critical to preserving TPS's learning-oriented culture. Using synthetic data to simulate four organizational conditions (TPS+, TPS−, Non-TPS+, and Non-TPS−), this research examines how human-centered factors moderate the relationship between AI augmentation and psychological outcomes. The results aim to advance the theoretical understanding of how socio-technical systems can integrate AI responsibly while sustaining trust, engagement, and continuous improvement.

LITERATURE REVIEW

The Toyota Production System (TPS) is globally recognized as the benchmark for production and operations management, renowned for its ability to drive operational excellence. This level of excellence is achieved by following key concepts of lean manufacturing, including continuous improvement (kaizen) and frontline participation. At the core of TPS is a structure that enables employees at all levels to be able to identify inefficiencies, raise concerns, and make suggestions without relying on the hierarchical approval (1). This decentralized system creates both operational flexibility and psychological empowerment, allowing improvement initiatives to emerge organically from the front line. As Spear and Bowen (1) emphasize, effective decentralization requires both structural discipline and a supportive social environment where management trusts workers to act and workers feel safe to contribute ideas.

That willingness for upper management to step back and for team members to speak up requires what Edmonson defines as psychological safety, a shared belief that the team is safe to take interpersonal risks (2). Psychological safety allows team members to voice their concerns, ask questions, and admit mistakes without the fear of shame or punishment, forming the social foundation of continuous learning. Later work by Edmonson and Lei reinforces its importance, showing that psychological safety predicts adaptive performance and collaboration in fast-changing, complex environments. Within it, the culture of continuous improvement would revert to compliance-based management (4).

As AI becomes increasingly integrated into organizational processes, new pressures emerge for sustaining psychological safety within systems such as TPS. AI applications increasingly automate decision-making and operational tasks that human teams have traditionally managed. While these technologies can improve efficiency and consistency, they also shift how responsibility and authority are distributed, which may reduce employees' sense of control and willingness to question automated outcomes. Such dynamics can weaken the learning culture that TPS depends on. Wilson and Daugherty (3) emphasize that effective AI adoption requires collaborative intelligence, humans and machines working together so that algorithms augment, rather than replace, human expertise and judgment. Maintaining this balance is essential to preserving trust and engagement in human-centered production systems.

There is substantial literature on the distinction of AI usage as a tool for augmentation versus automation. Brynjolfsson *et al*. argue that AI's organizational impact depends on whether it enhances or replaces human labor (5). When used for automation, AI can support decision-making and its development, reinforcing psychological safety by empowering workers. Conversely, automation tends to displace workers and increase uncertainty and anxiety, reducing psychological safety. In empirical manufacturing contexts, Brynjolfsson *et al*. found that the adoption of robots in manufacturing led to reduced employment in low-skilled positions, but that automation improved high-skilled wages (6). This suggests that the relationship between AI integration and employee well-being is moderated by organizational design and leadership transparency.

Within TPS, where improvement depends on collaborative problem-solving and continuous feedback, AI tools are more likely to augment human judgment rather than replace it. This contrasts with traditional manufacturing systems that vary in transparency and participative decision-making. Accordingly, this study models how AI's effects on team outcomes differ across organizational structures that vary in transparency and participative decision-making.

H1: Transparency and participative decision-making will positively moderate the relationship between AI implementation type (augmentation vs automation) and employee outcomes, specifically, psychological safety and perceived compensation stability. Teams in high-transparency, participative TPS conditions (TPS+) will report higher psychological safety and compensation than those in low-transparency conditions (TPS−, Non-TPS−).

Recent literature has explored the concept of Creative Displacement Anxiety (CDA), the fear that one's human capabilities, such as creativity, judgment, or intuition, are becoming obsolete by AI. Caporusso notes that Generative AI differs from prior technological shifts because it mimics human cognitive outputs, challenging identity and legitimacy among knowledge workers (7). This anxiety is not limited to only the creative industries; it can emerge across professions. However, perceived control and participative culture moderate this effect: employees in distributed systems like TPS, which empower decision-making at all levels, experience the opposite. Burton *et al*. found that algorithm aversion decreases when individuals retain decision authority, directly linking perceived autonomy to trust in AI systems (8).

H2: Organizational structure will influence AI-related displacement anxiety such that employees in participative TPS environments will report lower CDA and higher psychological safety than those in hierarchical manufacturing systems.

One of the most challenging parts of implementing AI is that employees need to develop the skills and mindset to work alongside these tools. Generative AI can be deployed in two ways: automation, where it replaces human labor by performing tasks independently, and augmentation, where it enhances human work by acting as a tool that employees actively use to improve their decision-making, creativity, and problem-solving. The latter of the two emphasizes the collaboration between humans and AI rather than substitution and is more consistent with the Toyota Production System's emphasis on learning and continuous improvement.

Pavone found that individuals with higher psychological safety and self-confidence were more motivated to engage with AI tools creatively, whereas those experiencing performance anxiety tended to avoid or resist AI use (9). These findings suggest a reciprocal relationship between psychological safety and AI adoption: psychologically safe teams are more likely to use AI constructively, and successful AI augmentation can reinforce that safety by enabling positive learning experiences. Edmondson similarly observed how teams with higher team psychological safety are more willing to take risks, share more feedback, and overall are more open to learning, and these are in line with the core components of TPS (2). Contrastingly, traditional manufacturing systems are lacking these learning structures, which often results in a slower and more superficial AI adoption, driven primarily by automation goals rather than collaborative augmentation. This theoretical distinction frames the next hypothesis.

H3: The level of AI adoption will be significantly higher in TPS contexts than in traditional manufacturing, due to greater learning orientation, feedback culture, and psychological safety.

Given the higher level of AI adoption in TPS, the impact on team psychological safety will be moderated by transparency and participative decision-making. Higher transparency will lead to higher team psychological safety at the same level of AI adoption. Same for participative decision making. Worker satisfaction and trust with AI systems depend on whether the workers feel empowered or undermined by the new addition of AI tools. Burton *et al*. discovered that AI was more accepted when it augmented human judgment calls rather than overriding them (8). When workers were able to have some control over the system's recommendations, they were more likely to trust and use the AI tools. This aligns with self-determination theory, which posits that perceived autonomy is a central determinant of motivation and engagement.

TPS is designed around human agency and emphasizes the decentralization of decision-making and structured autonomy at the team level. This structure allows workers to view AI as a supportive tool that enhances analytical capacity and decision speed, thereby reinforcing their sense of competence and contribution. Conversely, in traditional manufacturing structures, AI is likely to be introduced to the top-down, giving employees less control and clarity over the tool, which can weaken morale.

H4: AI tools that enhance perceived autonomy in problem-solving will produce higher job satisfaction, team performance, and psychological safety in TPS environments compared to traditional manufacturing.

This literature review integrates psychological, technological, and organizational perspectives to construct a socio-technical model linking AI adoption, transparency, and participative decision-making to psychological safety. The framework provides the theoretical foundation for the study's synthetic simulation of four conditions, TPS+, TPS−, Non-TPS+, and Non-TPS−, to examine how AI's organizational impact depends on both human-centered design and system transparency.

METHODS AND MATERIALS

A mixed-methods approach was employed, combining a literature review with a conceptual model supported by synthetic data, to investigate the impact of AI adoption on team psychological safety within the Toyota Production System (TPS). The literature review focuses on existing research related to TPS practices, psychological safety in team settings, and the integration of AI in the redesign process. This helps create a theoretical foundation for understanding the relationship between the factors and identifying gaps in how the implementation of AI may reshape communication, trust, and learning within a TPS framework. Due to the unavailability of direct data from Toyota, the study simulates TPS-inspired environments through synthetic data generation, allowing for controlled modeling of theoretical relationships between AI integration and team psychological safety.

Scenarios Definitions

The data collected examined how varying conditions of AI integration, transparency, and participative decision-making influenced team psychological safety, team performance, AI adoption, and perceived workforce outcomes across four simulated organizational contexts:

TPS+: You are a factory worker in a Toyota Production System environment where AI tools have been adopted as augmentation tools, not automation. Decisions are transparent, and employees are included in problem-solving.

TPS−: You are a factory worker in a Toyota Production System environment where AI tools have been adopted as augmentation tools, not automation. Decisions are not made transparently, and employees' opinions and thoughts have been left out of the problem-solving processes.

Non-TPS+: You are a factory worker in a traditional car manufacturing environment where AI tools have been adopted as automation tools, not augmentation. Decisions are transparent, and employees are included in problem-solving.

Non-TPS−: You are a factory worker in a traditional car manufacturing environment where AI tools have been adopted as automation tools, not augmentation. Decisions are not made transparently, and employees' opinions and thoughts have been left out of the problem-solving processes.

Synthetic Data Generation

Synthetic data is created through specific prompting of Large Language Models (LLMs) to generate structured and theoretically coherent patterns that mimic human-like response variability. Prior research on synthetic data (10, 11) suggests that such data can reveal conceptual relationships, but they do not replicate human cognition or social dynamics. Thus, in this study, synthetic data serves as a probe for theoretical relationships, not to measure or predict real behavioral outcomes.

Conceptual Model

The conceptual model (Figure 1) maps the interactions between AI integration type (augmentation vs. automation), psychological safety, and TPS's emphasis on continuous improvement and frontline participation. It predicts that transparency and worker participation will moderate the impact of AI adoption on team-level outcomes, sustaining trust and open communication that is central to TPS. The model is tested using synthetic datasets to examine how different scenarios of AI integration may impact psychological safety and team-based learning scenarios. The dataset comprises LLM-generated survey-style responses to questions on team psychological safety, team performance, AI integration, and perceived compensation, using questions both adapted conceptually and taken directly from Edmondson's (2) work.

Data Processing and Software

Data were generated using the Google Gemini 2.0 Flash model, selected for its accessibility and reliable output variance. A temperature of 0.9 was applied to balance creativity and coherence, ensuring diversity across the simulated responses while still maintaining thematic consistency. In total, 1,000 responses were generated (250 per scenario), representing four AI adoption conditions: high-transparency participative TPS (TPS+), low-transparency TPS (TPS−), high-transparency traditional manufacturing (Non-TPS+), and low-transparency hierarchical manufacturing (Non-TPS−). This structure ensured a balanced dataset for comparative analysis.

All data collection and analysis were conducted using RStudio (Version 2025.05.1+513) for reproducibility and statistical transparency. The analysis pipeline included descriptive statistics, one-way ANOVA (F-tests), Tukey post-hoc comparisons, and chi-square tests to evaluate both continuous and categorical outcomes. Specifically, ANOVA was used to test for mean differences in psychological safety, team performance, and AI adoption across scenarios, while chi-square analyses examined categorical variables such as perceived pay change and job redundancy distributions. Standard deviation and variance were also calculated to assess the internal consistency and distributional realism of the synthetic data.

Variance checks and outlier removals were done manually, and the means were visualized using ggplot2 for clarity. The code was framed by that used by Bisbee *et al.* (12) and adapted to fit the structure of this project, with guidance from generative AI tools, including Google Gemini and ChatGPT, for syntax refinement and efficiency.

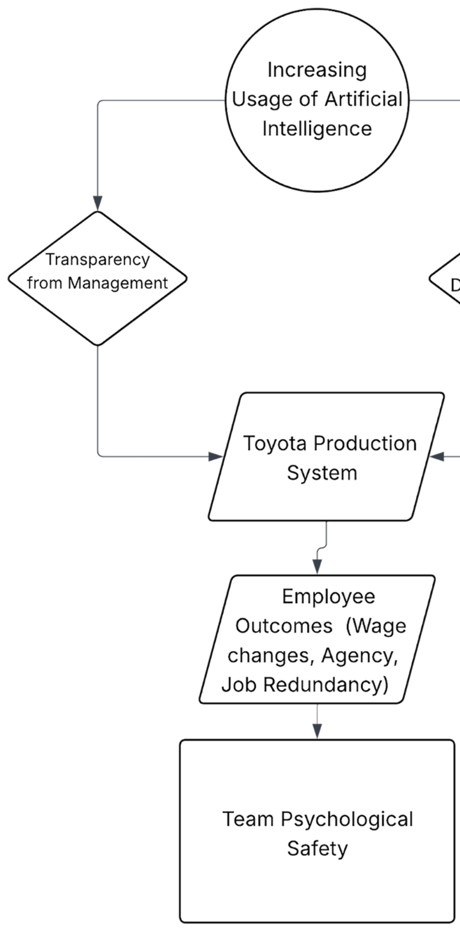

Figure 1. The model illustrates how increasing AI usage (augmentation vs. automation) interacts with organizational moderators (transparency and participative decision-making) to influence Toyota Production System (TPS) principles including agency, kaizen, and feedback loops. TPS then affects employee outcomes (perceived wage changes and job redundancy), which collectively shape team psychological safety, the primary outcome.

LLM Prompt Template

The following prompt was provided to LLM for each run. ("ROLE: You are simulating responses to a survey as if you are a human. Answer realistically based off the scenario given and vary responses to match persona across questions. Use the Likert scale (1=Strongly disagree, 2=Disagree, 3=Somewhat disagree, 4=Neutral, 5=Somewhat agree, 6=Agree 7=Strongly agree). For the question on pay change use one of ['a','b','c'] (a=increase, b=decrease, c=no change). For the question on redundancy percentage use one of ['a','b','c','d','e','f']-a=0%, b=0.1–2%, c=2–5%, d=5–10%, e=10–20%, f=>20%. Do not always give the same number; vary them to some extent, as a real person taking the survey would. Do not create patterns in which a numbers are repeated in a order, do not repeat the same number response for the entire survey, act as a human replacement. CONTEXT (SCENARIO): SCENARIOS[[scenario_id]]; TASK: sprintf(- Produce EXACTLY %d respondents for SCENARIO=%s., n, scenario_id), Return ONLY JSON that strictly adheres to the provided schema, and Output format must be an ARRAY of OBJECTS.",sep="\n")

Limitations of Synthetic Data

Finally, the study acknowledges that while synthetic LLM data enables structured testing of theoretical relationships, they do not replicate real human variance. As Anthis *et al*. (10) discuss how promising LLM social simulations are, while cautioning issues such as bias and alienness, and Xie *et al*. demonstrate that LLM agents can match human behavior at a high level (11). At the same time, limitations still exist. Responses from LLMs tend to show less variance, are highly sensitive to prompt design, and may shift as models evolve (12). Argyle *et al*. go on to write about how these limitations of LLMs are what make them far from being a universal solution (13).

RESULTS AND DISCUSSION

Overall Mean Comparisons

Mean comparisons indicated clear performance differences across the four conditions (Table 1). Psychological safety and team performance were highest under the TPS+ condition (M = 5.11, 5.54, respectively), followed by Non-TPS+ (M = 4.97, 5.26). Both systems under hierarchical (−) management exhibited substantially lower means (TPS−: M = 3.33, 3.24; Non-TPS−: M = 2.70, 2.56). AI adoption followed that pattern, with higher levels in participative systems (TPS+: M = 4.61; Non-TPS+: M = 4.13) and lower levels in hierarchical systems (TPS−: M = 3.36; Non-TPS−: M = 3.00) (Figure 2). Standard deviations were also consistent

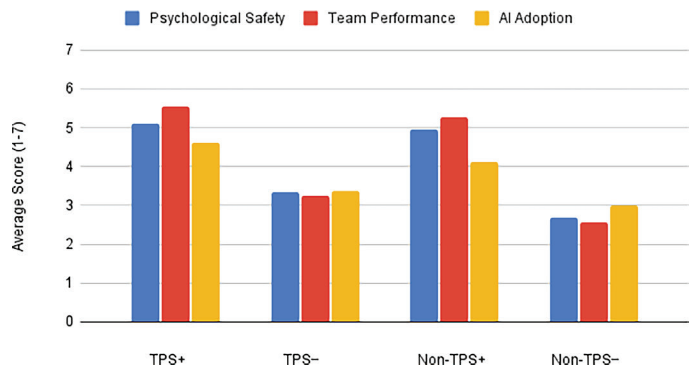

Figure 2. Mean scores (M) for each scenario (TPS+, TPS−, Non-TPS+, and Non-TPS−) are presented across three outcome measures: psychological safety, team performance, and AI adoption. Values are derived from synthetic datasets simulating team responses, illustrating conceptual differences between participative and hierarchical organizational conditions. TPS+ consistently demonstrates the highest mean scores across all categories, whereas Non-TPS− shows the lowest scores, reflecting the hypothesized moderating effects of transparency and participative decision-making.

Table 1. Summary of psychological safety, team performance, AI adoption, perceived pay change, and perceived job redundancy across four simulated organizational conditions: TPS with transparency and participative decision-making (TPS+), TPS without transparency and participation (TPS−), traditional manufacturing with transparency (Non-TPS+), and traditional manufacturing without transparency (Non-TPS−). Pay Change: A (Increase in pay); B (Decrease in pay); C (No change in pay); Job Redundancy: A (0%); B (0.1% - 2%); C (2%-5%); D (5%-10%); E (10-20%); F (>20%)

Scenario 1 (TPS+):	**Scenario 2 (TPS−):**	**Scenario 3 (Non-TPS+):**	**Scenario 4 (Non-TPS−):**
Psychological Safety Avg Score: 5.1072	Psychological Safety Avg Score: 3.3308	Psychological Safety Avg Score: 4.9688	Psychological Safety Avg Score: 2.7012
Team Performance Avg Score: 5.542666667	Team Performance Avg Score: 3.244666667	Team Performance Avg Score: 5.255333333	Team Performance Avg Score: 2.562
AI Adoption Avg Score: 4.6135	AI Adoption Avg Score: 3.363	AI Adoption Avg Score: 4.1255	AI Adoption Avg Score: 3.0045
Pay Change:	Pay Change:	Pay Change:	Pay Change:
A: 92.00%	A: 7.60%	A: 17.20%	A: 2.80%
B: 0.00%	B: 31.60%	B: 3.60%	B: 59.60%
C: 8.00%	C: 60.80%	C: 79.20%	C: 37.60%
Job Redundancy:	Job Redundancy:	Job Redundancy:	Job Redundancy:
A: 97.60%	A: 12.00%	A: 67.60%	A: 6.40%
B: 2.40%	B: 23.60%	B: 26.00%	B: 17.20%
C: 0.00%	C: 22.40%	C: 2.40%	C: 30.80%
D: 0.00%	D: 15.20%	D: 1.60%	D: 20.40%
E: 0.00%	E: 14.00%	E: 1.20%	E: 11.60%
F: 0.00%	F: 12.80%	F: 1.20%	F: 13.60%

across means (SD ≈ 0.6), indicating stable response distributions and internal reliability of the synthetic data. All analyses were conducted using a two-tailed test with an alpha level of .05.

Hypothesis Testing

Data for H1

One-way analyses of variance (ANOVA) were conducted to assess whether observed mean differences across scenarios were statistically significant. Results revealed significant effects for psychological safety ($F(3, 996) = 972.6$, $p < .001$, $\eta^2 = 0.75$), and team performance ($F(3, 996) = 1453$, $p < .001$, $\eta^2 = 0.81$).

These effect sizes indicate large between-group differences. Post hoc Tukey HSD comparisons confirmed that the TPS+ condition scored significantly higher than all other scenarios across all three means ($p < .001$), while Non-TPS− consistently scored lowest. Chi-square analyses were also conducted. The distributions differed significantly across the four scenarios with the result for pay change being $\chi^2(6, N = 1000) = 836.07$, $p < .001$.

The data supported H1. Teams in the TPS+ scenario showed the fewest job cuts and the highest rate of wage increases, while TPS− showed more rationalization and lower pay outcomes. Similarly, Non-TPS+ outperformed Non-TPS−, showing that transparency and participative decision-making (the + condition) reduced job insecurity regardless of system type. Results were the same in psychological safety, as TPS+ and Non-TPS+ outperformed their counterparts. These results suggest that transparency and participation buffer the negative impacts of AI implementation, maintaining stability and trust within teams.

Data for H2

One-way analyses of variance (ANOVA) results are used once again, but this time looking at psychological safety and AI adoption, psychological safety ($F(3, 996) = 972.6$, $p < .001$, $\eta^2 = 0.75$) and AI adoption ($F(3, 996) = 347.3$, $p < .001$, $\eta2 = 0.51$); these effect sizes once again shows large between-group differences. The chi-square test results for job redundancy are used for this hypothesis, and the results are $\chi^2(15, N = 1000) = 686.91$, $p < .001$, with TPS+ exhibiting the fewest redundancies. These results indicate that participative structures mitigate perceived AI-driven displacement. The data supports H2. TPS+ teams scored the highest in psychological safety and AI adoption (including displacement anxiety), while Non-TPS− scored the lowest. Employees in a transparent and participative environment reported feeling more secure, more included in AI-related changes, and more willing to adopt AI tools. These results align with the theory that participative systems foster collective trust and adaptive behavior during technological transitions.

Data for H3

AI adoption scores varied significantly by condition ($F(3, 996) = 347.3$, $p < .001$, $\eta^2 = 0.51$). Post hoc comparisons confirmed that TPS+ teams adopted AI effectively more frequently than all other groups ($p < .001$). H3 was supported by the consistent hierarchy of results (TPS+ > Non-TPS+ > TPS– > Non-TPS–). Notably, Non-TPS+ beat out TPS− in both AI adoption and team psychological safety scores by a significant amount, suggesting that the positive effects of transparency and participation can outweigh the structural benefits of TPS when those moderators are absent. This finding emphasizes the importance of human-centered governance in AI implementation: systems succeed not only because of process design, but because employees feel involved and respected throughout the adoption process.

Data for H4

Participative and transparent conditions yielded significantly higher psychological safety and team performance ($F(3, 996) = 972.6$, $p < .001$, $\eta2 = 0.75$) and team performance ($F(3, 996) = 1453$, $p < .001$, $\eta2 = 0.81$). H4 was supported. TPS+ teams demonstrated the highest job satisfaction and performance indicators, followed by Non-TPS+. Using psychological safety scores as a proxy for job satisfaction, the results suggest that participative, transparent environments maintain morale and motivation even amid automation pressures.

Overall, the results provide strong conceptual evidence that transparency and participation meaningfully enhance both team-level and individual-level outcomes in the context of AI integration. TPS+ consistently outperformed traditional, hierarchical environments on all key indicators. In contrast, the Non-TPS− scenario exhibited the weakest outcomes, characterized by lower psychological safety, reduced performance, limited AI adoption, and more negative workforce perceptions. While these results are robust statistically, they are derived from synthetic data intended for conceptual modeling rather than for empirical generalization. The magnitude of F- and χ^2-values should be interpreted as reflecting the theoretical strength of the modeled relationship rather than observed real-world effect sizes.

Collectively, the findings support the hypothesis that participatory and transparent systems promote higher collaboration, trust, and openness toward AI integration, forming a strong conceptual basis for future empirical validation in manufacturing contexts.

The results of this study demonstrated that organizational transparency and participative decision-making (the TPS moderators) strongly influence the relationship between AI integration and team outcomes. Across all scenarios, the pattern followed the theoretical expectation: TPS+ > TPS− > Non-TPS+ > Non-TPS−. This supports the idea that participative structures enhance psychological safety, job satisfaction, and openness to AI adoption by aligning human and technological systems. When comparing across systems, TPS− underperformed Non-TPS+ in all five categories measured, reinforcing that participative culture is more critical to successful AI adoption than structure alone. Still, TPS− outperformed Non-TPS−, confirming that even without strong moderators, TPS principles offer some advantage over traditional manufacturing systems. These findings underscore that transparency and participation act as powerful moderators that are able to shape how organizations adapt to the introduction of AI. Leaders who integrate these values can reduce resistance, sustain trust, and improve team outcomes.

Limitations

Limitations of this approach include the reliance on synthetic data rather than real-world team interactions and the inherent simplifications of the conceptual model. However, this method enables the exploration of the complex dynamics that occur when aligning AI adoption with the human-centered learning culture of TPS, providing actionable insights for manufacturing leaders seeking responsible AI integration.

Given these limitations, the use of real-world surveys in the future will be necessary to back the tentative findings from our studies. In addition, refining the survey instruments and adjusting the scenarios to have LLM models be able to distinguish the effects of the moderators in each scenario will further strengthen the validity of future analyses.

CONCLUSION

This study explored how the integration of AI affects team psychological safety within the Toyota Production System. Using a synthetic dataset across four scenarios, the results show that AI integration in TPS environments can reinforce psychological safety and team performance, but only when transparency and participative decision-making are preserved. TPS structures with high levels of participation (TPS+) constantly outperform traditional manufacturing environments (Non-TPS+ and Non-TPS−) and TPS without these moderators (TPS−). Conversely, the absence of these moderators often eroded the benefits of TPS so significantly that TPS underperformed against traditional manufacturing.

These findings highlight the importance of workplace culture in shaping AI's impact. For organizations that are seeking responsible AI adoption, the lesson is clear: technology alone cannot guarantee better outcomes. Psychological safety, supported by open communication and shared decision-making, remains the foundation for continuous improvement.

Future research should validate these conceptual findings with real-world data. Mixed-method designs, such as surveys, longitudinal field studies, or ethnographic observations, could capture how employees actually experience AI integration over time. Examining how different levels of AI maturity influence psychological safety would also help identify when and how participative practices are most critical. By combining empirical validation with human-centered design principles, future work can guide leaders toward AI strategies that enhance both organizational performance and employee well-being.

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to my mentor, Julian Corj, for their support, guidance, and encouragement throughout the entire process of writing. I am sincerely thankful for their time, patience, and insight, which were invaluable in helping me understand how to structure and write a paper from start to finish. I would also like to thank Lumiere Education for providing me with the resources I needed to improve my writing skills.

FUNDING SOURCES

The author did not receive any funding for the conduct of this research or the preparation of this article.

CONFLICT OF INTEREST

The author declares that there are no conflicts of interest regarding the publication of this article.

REFERENCES

1. Spear S, Bowen HK. Decoding the DNA of the Toyota Production System. *Harv Bus Rev.* 1999 Sept; 77 (5): 96.
2. Edmondson A. Psychological Safety and Learning Behavior in Work Teams. *Adm Sci Q.* 1999 June 1; 44 (2): 350-83. https://doi.org/10.2307/2666999
3. Wilson HJ, Daugherty PR. The Secret to Successful AI-Driven Process Redesign: Strong leaders put business transformation in the hands of all employees. *Harv Bus Rev.* 2025; 104 (1-2): 45-51.
4. Edmondson A, Lei Z. Psychological Safety: The History, Renaissance, and Future of an Interpersonal Construct. *Annu Rev Organ Psychol Organ Behav.* 2014 Mar 21; 1 (1): 23-43. https://doi.org/10.1146/annurev-orgpsych-031413-091305
5. Brynjolfsson E, Li D, Raymond LR. Generative AI at work. *Oxf Univ Press.* 2025 May; 140 (2): 889-942. https://doi.org/10.1093/qje/qjae044
6. Brynjolfsson E, Buffington C, Goldschlag N, Li JF, Miranda J, Seamans R. Robot Hubs and the Use of Robotics in US Manufacturing Establishments. *AEA Pap Proc*. 2025 May 1; 115: 24-8. https://doi.org/10.1257/pandp.20251000
7. Caporusso N. Generative Artificial Intelligence and the Emergence of Creative Displacement Anxiety: Review. Res Directs Psychol Behav [Internet]. 2023 Oct 16 [cited 2025 Aug 8]; 3 (1). Available from: https://www.researchdirects.com/index.php/psychology/article/view/95. https://doi.org/10.53520/rdpb2023.10795
8. Burton JW, Stein M, Jensen TB. A systematic review of algorithm aversion in augmented decision making. *J Behav Decis Mak*. 2019 Oct; 33 (2): 220-39. https://doi.org/10.1002/bdm.2155
9. Pavone G. Generative AI in the Learning Process: Threat or Tool? Understanding the Role of Self-Esteem and Academic Anxiety in Shaping Student Motivations. 2025 June; https://doi.org/10.1177/02734753251346857
10. Anthis JR, Liu R, Richardson SM, Kozlowski AC, *et al*. LLM Social Simulations Are a Promising Research Method [Internet]. arXiv; 2025 [cited 2025 Sept 5]. Available from: http://arxiv.org/abs/2504.02234
11. Xie C, Chen C, Jia F, Ye Z, *et al*. Can Large Language Model Agents Simulate Human Trust Behavior? *Adv Neural Inf Process Syst.* 2024 Dec; 37: 15674-729. https://doi.org/10.52202/079017-0501
12. Bisbee J, Clinton JD, Dorff C, Kenkel B, Larson JM. Synthetic Replacements for Human Survey Data? The Perils of Large Language Models. *Polit Anal.* 2024 Oct; 32 (4): 401-16. https://doi.org/10.1017/pan.2024.5
13. Argyle LP, Busby EC, Fulda N, Gubler JR, Rytting C, Wingate D. Out of One, Many: Using Language Models to Simulate Human Samples. *Polit Anal.* 2023 July; 31 (3): 337-51. https://doi.org/10.1017/pan.2023.2

APPENDIX

LLM Prompting Framework: Simulated Survey Items

The following items were included in prompts to the Google Gemini 2.0 Flash model to generate structured synthetic responses simulating human survey data. No actual human participants completed these items.

Item Development and Attribution:

Questions directly adapted from or taken from Edmondson (2): PS2, PS4, PS7, TP1, TP2

Questions developed for this study: All AI adoption items, pay change, and job redundancy questions

Simulated Survey Items:

Psychological Safety

1. I feel my contributions are valued by the team
2. I trust the members of this team to support me when I take a risk*
3. People on this team feel safe to admit when they've made an error
4. I often notice that people on my team ask questions when they are unsure about something*
5. Everyone on this team has a chance to contribute during discussions
6. People on this team are sometimes hesitant to share new ideas as they are worried about how others might react (reverse scale for analysis)
7. If you make a mistake on this team, it is often held against you*
8. If I notice a problem that is slowing team progress, I feel I can raise it
9. When people make a mistake on this team, they are usually used as opportunities to learn rather than as a place to blame
10. It is acceptable to disagree with others on this team

Team Performance

1. My team continuously improves our processes*
2. Our team consistently meets the quality standards expected in our work*
3. Our team learns from past projects to improve future work
4. Constructive feedback is shared within the team to help us improve
5. We regularly discuss how we can improve our team's processes
6. We complete tasks on time as a team

AI

1. We use AI in our workflows (adoption)
2. We use AI in our workflows to improve efficiency (adoption + improvement)
3. I am worried AI will replace my role (threat) (reverse scale for analysis)
4. AI tools will help me make decisions more confidently (augmentation)
5. I feel that AI tools could negatively impact how our team collaborates (threat) (reverse scale for analysis)
6. I am aware of the ways AI is impacting our work processes (awareness)
7. I have received adequate training to effectively use AI tools in my work (learning)
8. I feel the AI tools we use improve the quality of my work (adoption + improvement)

Pay

1. Report expected change in pay (increase, decrease, no change)
2. What is the percentage of jobs at your company that have been made redundant as a consequence of the introduction of generative AI? (a) 0% b) 0.1% - 2% c) 2%-5% d) 5%-10% e) 10-20% f) >20%)