Societal Change in the Age of AI: Inclusion as Defense Against the Augmented Insider

Sander Zeijlemaker & Ola Alhaddad

Abstract: Despite widespread expectations that artificial intelligence will drive productivity and growth, evidence shows persistent failure in scaling AI and realizing value. Organizations face a paradox of high investment but limited returns, alongside rising social inequality, security risks, and governance gaps. This raises the question of why AI adoption fails to deliver consistent outcomes.

This study employs a qualitative system dynamics approach, combining literature review with six expert consultations across security, management, and inclusion domains. Through iterative model building, key variables and feedback loops are identified and structured into a causal system. The model emphasizes endogenous dynamics, capturing how interactions within economic, social, and security domains shape AI adoption outcomes.

The analysis reveals that AI adoption is governed by interacting reinforcing and balancing feedback loops across four domains: investment dynamics, social inequality, security escalation, and the emergence of the "augmented insider." These interactions produce systemic contradictions, where initial gains are offset by rising costs, distrust, and risks. AI adoption thus behaves as a complex, co-evolving system rather than a linear technological progression.

The findings suggest that technical or isolated policy interventions are insufficient. Sustainable AI adoption requires integrated strategies addressing inclusion, governance, workforce transformation, and security simultaneously. Inclusion emerges as a critical stabilizing mechanism to counter reinforcing risks. Future research should focus on model quantification and empirical validation to support policy design and enable simulation of long-term system behavior.

Key words: AI Adoption, Qualitative Analysis, Augmented Insider, Inclusion, Equity

From Promise to Paradox: Failures in AI Adoption

Artificial intelligence (AI) is widely framed as a general-purpose technology capable of transforming productivity, economic growth, and societal well-being. This expectation has driven rapid investment across public and private sectors. However, emerging empirical evidence challenges this narrative. Rather than delivering consistent benefits, AI adoption shows persistent gaps between expectations and realized outcomes, pointing to structural issues in how it is understood and managed.

At the firm level, the inability to translate AI investments into sustained value is well documented. Boston Consulting Group (2024) finds that 74% of companies fail to scale AI beyond pilots, while Gartner (2025) projects that over 40% of agentic AI initiatives will be

discontinued by 2027. High-profile setbacks, such as the cancellation of major infrastructure projects (Shilov, 2026), further reinforce this instability. Despite this, organizations continue to promote “responsible AI” as a driver of growth (Industry Dive, 2025), even though 95% of executives report AI-related failures and only 2% meet responsible use standards (ETtech, 2025). This indicates a disconnect between governance narratives and operational reality.

At the macro level, policy responses reflect similar inconsistencies. The cancellation of $5.1 billion in IT contracts (Ammachchi, 2025) contrasts with growing dependence on digital infrastructure. At the same time, regulatory actions, such as those affecting Anthropic, can unintentionally destabilize innovation ecosystems (Dave, 2026).

AI adoption also produces measurable societal and political effects. Surveillance technologies are increasingly used in repression, disproportionately targeting women activists (Bachlakova, 2025), while actors such as Palantir raise concerns about the influence of AI on democratic processes (McCordick, 2026).

These patterns reveal systemic contradictions: rising investment alongside limited value, strong governance narratives alongside weak implementation, and technological progress alongside increasing social risk. System dynamics offers a lens to understand these outcomes as the result of interacting feedback structures rather than isolated failures.

All combined this leads to the research question: *Why are large-scale AI investments failing to consistently deliver value, and how do interacting economic, social, and security dynamics shape these outcomes over time?*

Our work is an initial attempt to provide more clearance in this field. This paper first explains our qualitative model building approach followed by explaining the systemic structure that explains AI adoption. Although this work is initial and qualitative in nature we provide already some preliminary policy analysis. The last part of this paper is reflection on the work and opportunities for future research.

Model Creation

Model building in System Dynamics is inherently an interactive and iterative process (Forrester, 1961; Pruyt, 2013; Sterman, 2000). In line with this tradition, the present study adopts a structured yet participatory approach to model development, integrating insights from literature review, group dialogues, and expert consultation. This approach ensures that the resulting model is not only theoretically grounded but also closely aligned with real-world organizational dynamics.

To operationalize this process, our model building was supported by literature review and multiple expert consultations. Expert consultation was conducted through both group dialogues and individual interviews. Insights from the literature review informed the initial model structure and guided the identification of key variables and feedback mechanisms.

In the context of this AI adoption study, this participatory approach was particularly valuable, as it enabled the integration of domain expertise from multiple disciplines into a coherent model

structure. However, due to participant availability and preferences, consultations were used more extensively than group building session sessions. These consultations were semi-structured and aimed at eliciting causal relationships, key variables, and system boundaries.

The model development process consisted of six supportive expert consultations. Participants included information security consultants, diversity, and inclusion specialists, as well as executive and senior management across different industries. These stakeholders contributed diverse perspectives on AI adoption, including economic, organizational, ethical, social, and security dimensions. The iterative dialogue between participants revealed important interdependencies that were not initially apparent. One key outcome was the emergence of the concept of the *augmented insider*, a central construct in the model, as well as the recognition of the importance of inclusion and equality in AI adoption challenges.

The use of expert knowledge in model construction aligns with established system dynamics practice, where models are built with individuals who possess direct experience with the system under study (Ford & Sterman, 1998). This ensures that the model reflects the structure and behaviour of the real system rather than abstract assumptions. In addition, tools and scripts provided by the System Dynamics Society were used to support the consistency and rigor of the modelling process. We also used two sessions to provide feedback on the modelling work.

Given the qualitative nature of this study, the model-building and validation processes were adapted accordingly. The model captures key behavioural mechanisms rather than precise quantitative relationships.

A central requirement in system dynamics modelling is endogeneity, meaning that system behaviour should emerge from interactions within the model boundary (Forrester, 1961; Richardson & Pugh, 1981). To assess this, bull's-eye diagrams were used (Ford, 2010). These diagrams categorize variables into endogenous and exogenous groups, illustrating the model's focus, inputs, and boundaries. In accordance with system dynamics principles, the models developed in this study emphasize endogenous variables and feedback loops, ensuring that system performance arises primarily from internal dynamics rather than external inputs. In this work the different qualitative model structures (see Figure 1 to 4 show) show that most system performance arises primarily from internal dynamics.

Following the development of the conceptual structure, the model can be further enriched through quantification and formalization, enabling future integration with simulation and analysis techniques as well as further validation.

In summary, the model creation process combines rigorous system dynamics methodology with participatory design and expert-informed insights. This approach provides a good foundation for an initial analyses of a complex phenomena such as AI adoption.

Delineating the Systemic Structure of AI adoption

AI adoption generates interconnected dynamics across four domains. First, *AI Investment Growth and Constraints* captures the reinforcing push for investment, alongside limits created

by rising costs, diminishing differentiation, and delayed returns. Second, *AI-Induced Social Inequality and Trust Erosion* shows how biased systems, exclusion, and power concentration can weaken legitimacy and reduce participation. Third, *Security Escalation and Risk Automation* highlights how AI intensifies both cyber conflict and automation-driven disruption, creating new social and security vulnerabilities. Fourth, The *Rise of the Augmented Insider,* which is a specific emerging dynamic with security escalation, reflects how displacement, exclusion, and retained technical knowledge can elevate insider-like threats. Together, these domains form an interdependent system in which economic, social, and security feedback co-evolve and shape the trajectory of AI adoption. Figure 1 shows an aggregated level of these interconnected sub models. In the forthcoming sections each of these four sub model structures will be explained in more detail. Feedback mechanisms are labelled through the perspective of AI adoption.

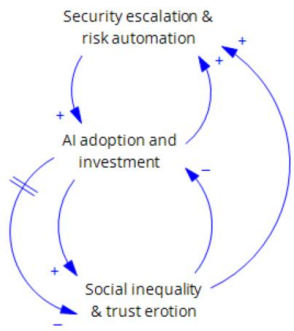

Figure 1. Systemic Structure of AI investment and adoption

AI Investment Growth and Constraints

Artificial intelligence is reshaping competitive dynamics through a reinforcing pattern of investment, capability development, and market response (Pengwei & Sun 2025). This substructure is visible in Figure 2. At the core of this process is a reinforcing feedback loop (**R1:** AI investment and adoption cycle): expectations of productivity gains stimulate increased investment in AI, which enhances capabilities and improves operational efficiency. As these gains become visible, competitive pressure intensifies, prompting further adoption across firms. This creates a self-reinforcing dynamic in which organizations continue to invest not only to advance, but to avoid falling behind.

This expansion is, however, moderated by several balancing feedback mechanisms, as economics are characterized by both reinforcing and balancing mechanisms (Joffe et al, 2021). There are limits to the growth of scaling AI (Bhardwaj et al., 2025).

First, **B1** (Advantage erosion) reflects the diminishing ability of AI to differentiate firms as adoption becomes widespread. What initially provides a strategic advantage gradually becomes a standard capability, reducing the incremental business value of continued investment. Especially as 73% of AI initiatives fail to scale beyond pilots (Dzreke et al, 2026) and 95% of the

enterprises do not show measurable profit impacts despite these investments (Wolfe et al., 2025).

Second, **B2** (Cost escalation and delayed value) captures the increasing reliance on scarce and costly resources (Avezov et al., 2025; McKinsey, 2024), including advanced chips, energy, data, and infrastructure. As demand for these inputs rises, costs increase, while the realization of business value is often delayed due to integration challenges and immature applications. This results in a widening gap between expenditure and returns. This resulted in a famous investor betting against vested tech firms in this domain by going short (Edwards, 2026).

A third constraint emerges through **B3** (Skepticism and capital correction). When anticipated returns fail to materialize within expected timeframes, firms experience financial underperformance and losses. This, in turn, drives investor skepticism, limits access to capital, and slows the pace of further investment (Avezov et al., 2025). Together, these balancing mechanisms counteract the initial reinforcing dynamic and shape the overall trajectory of AI adoption.

These interacting feedback loops are consistent with established system dynamics archetypes. The overall pattern reflects a Limits to Growth structure, in which early expansion driven by reinforcing feedback (**R1**) is progressively constrained by rising costs, reduced differentiation, and delayed value realization (**B1, B2, B3**). In addition, elements of a Fixes that Fail archetype are evident: aggressive investment in AI initially improves efficiency, but over time contributes to competitive convergence, cost inflation, and weaker returns, thereby redefining a new equilibrium of economic reality.

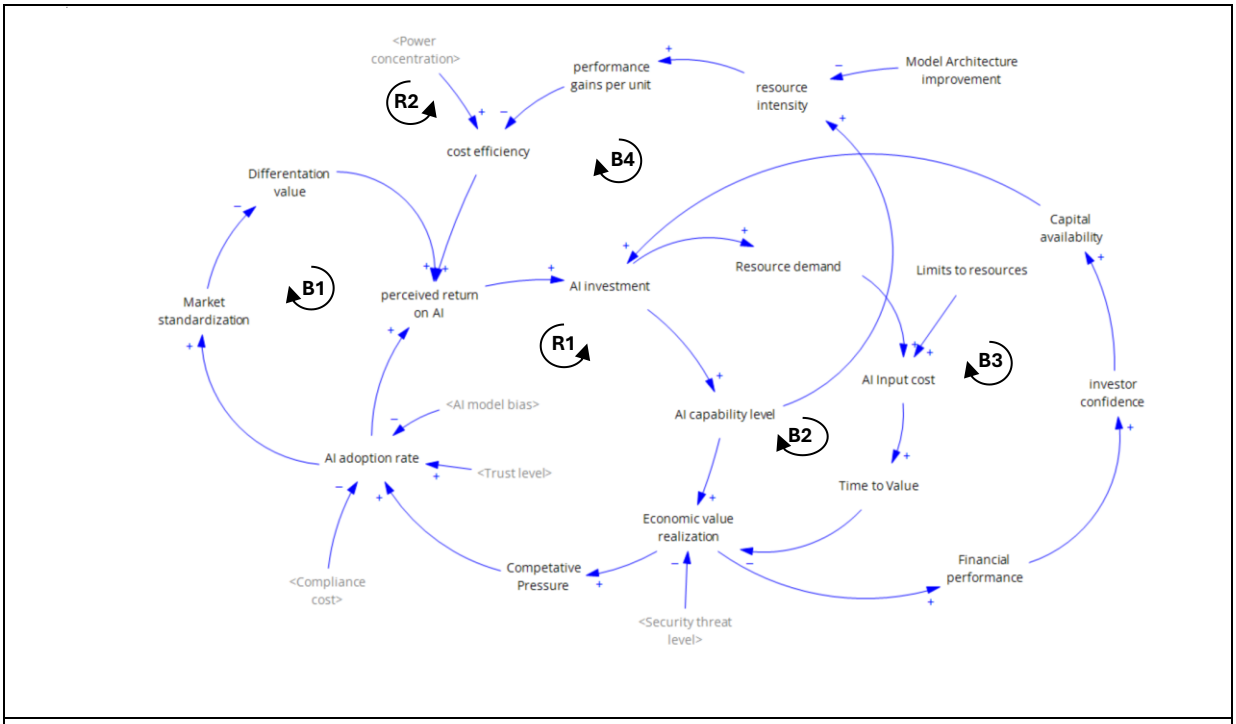

Figure 2. AI Investment Growth and Constraints

These economic dynamics are further compounded by technical constraints captured in **B4** (Performance ceiling and scaling limits). Current AI model architectures exhibit diminishing performance gains relative to the resources required to scale them. As improvements in model quality no longer keep pace with increases in cost, energy consumption, and infrastructure requirements, both technical and economic limits become increasingly binding. This reinforces the broader balancing forces within the system. Currently, new architecture evolves around the use of smaller models (Wang et al., 2025) multi orchestrated agent models (Meier et al, 2025), or new reasoning approaches for models (Kilcher, 2026).

Taken together, the interaction between **R1** and **B1–B4** produces a recurring cycle of rapid investment, rising expectations, constrained returns, and eventual correction. Breaking this pattern will require a shift away from continued scaling of existing approaches toward new model architectures and more efficient deployment strategies capable of delivering sustained and differentiated business value.

AI-Induced Social Inequality and Trust Erosion

Artificial intelligence is not only reshaping economic dynamics but also generating reinforcing social patterns that influence trust, inclusion, and long-term system quality. This substructure is Visible in Figure 3. A critical mechanism is captured in **B5** (Bias feedback loop). AI systems trained on historical data inherit existing societal biases (Pagan et al., 2023), which can lead to discriminatory outcomes (Leavy et al., 2020). These outcomes contribute to the exclusion of affected groups, increasing distrust and fear of AI. As trust declines, participation in AI-mediated systems decreases, reducing the diversity of data available for future training. Lower data diversity, in turn, further amplifies model bias, reinforcing the cycle. Ultimately increasing model bias limits AI adoption.

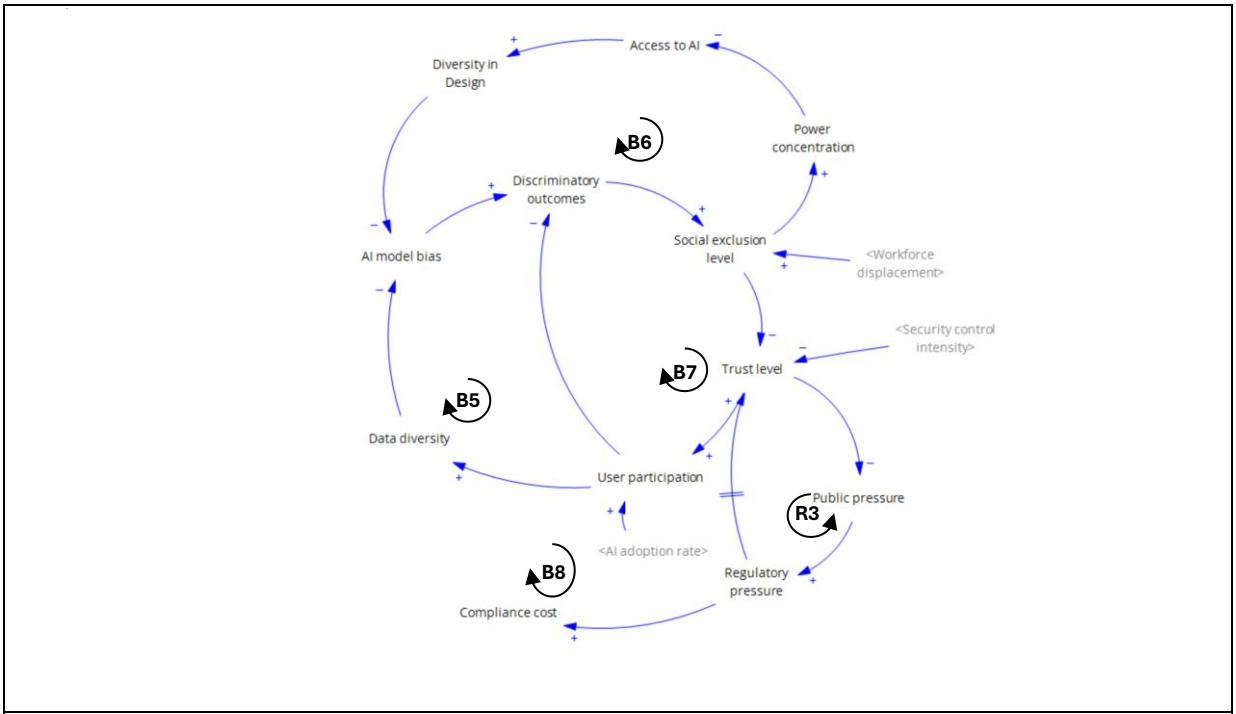

Figure 3. AI-Induced Social Inequality and Trust Erosion

A second balancing process that negatively affects AI adoption, **B6** (Power concentration loop), emerges from unequal access to AI capabilities that affect economic performance, social visibility, and long-term stability of social systems (Nabin, 2026). The increasingly higher cost of computing, data, and specialized talent limits access to advanced AI systems. This concentrates development and decision-making power among well-resourced organizations. These actors are better positioned to shape system design and priorities, often with limited representation of marginalized groups. Reduced diversity in design and governance increases the likelihood that biases persist or intensify, further reinforcing unequal outcomes. This concentration of power, however, coexists with achieving economies of scale and lower cost favoring AI adoption (**R2** Economics of scale – see Figure 2).

These dynamics are complemented by **B7** (Distrust–exclusion loop), in which discriminatory outcomes directly increase public fear and reduce engagement as it is being perceived as untrustworthy and unfair (Shams et al, 2025). Lower participation deepens exclusion, which in turn further amplifies distrust. Together, these loops create a self-sustaining pattern in which bias, exclusion, and fear mutually reinforce one another. This ultimately affects AI adoption negatively.

At the same time, reinforcing mechanisms begin to emerge. **R3** (Trust and legitimacy) reflects how rising public concern and regulatory scrutiny can limit AI deployment or reshape its development. The European AI act (AI ACT) and the development of digital trust frameworks (Dobrygowski & Valkhof 2024) can be an example of this intervention. An adept to strengthen AI adoption through trust and legitimacy. A side effect is that compliance may increase cost of AI adoption (**B8** – compliance cost increase). However, as distrust grows, organizations face pressure to improve transparency, fairness, and accountability, which can slow down adoption and increase compliance costs in the short turn but increase higher levels of trust and adoption eventually. This introduces a counteracting force to unchecked expansion.

These interacting feedback mechanisms align with established system dynamics archetypes. The system reflects a Success to the Successful pattern, where those with greater resources gain disproportionate access and influence, further strengthening their position. It also exhibits characteristics of Fixes that Fail, as reliance on historical data provides a fast and scalable solution but reproduces structural inequalities, undermining long-term system effectiveness.

Overall, the interaction between reinforcing loops (**R2–R3**) and balancing constraints (**B5 – B8**) in AI adoption produces a trajectory in which social inequality, power concentration, and public distrust can escalate unless deliberate interventions improve inclusivity, data diversity, and equitable access.

Security Escalation and Risk Automation

A third set of dynamics emerges in the domain of security, where artificial intelligence creates a continuous cycle of mutual escalation. This substructure is shown in Figure 4. This process is captured in (AI security arms race). The use of AI by malicious actors increases the scale and sophistication of attacks, raising the level of threat faced by organizations (Zeijlemaker et al, 2025). In response, firms invest in AI-enabled detection and defense systems to improve speed and accuracy (Zeijlemaker et al., 2025). Unfortunately, these defender-oriented investments

take time to implement due. However, these defensive improvements place pressure on attackers to further enhance their own capabilities, reinforcing a cycle in which each side's progress drives the other forward (Zeijlemaker & Siegel, 2023). This dynamic reflects a classic escalation pattern, where neither side can disengage without increasing its relative vulnerability. Technological evolution in this adversarial rat-race will have reinforcing and balancing mechanisms depending on adversarial (**B9**) or defense (**R4**) dominance which is trigger by time needed to improve and anticipate towards the opponent.

Alongside this, organizations respond to both security pressures and broader efficiency goals by increasing automation. This introduces **R5** (Cost reduction through automation), where AI-driven automation reduces operating costs and initially alleviates pressure to improve efficiency (Acemoglu & Restrepo, 2018; 2019) and favor AI adoption. However, this effect generates significant secondary consequences.

Most notably, **B10** (Automation–inequality loop) captures the social implications of this shift. Increased automation reduces demand for labor, leading to layoffs and workforce displacement. Even when business objectives are not realized organizations feel pressure to lower staff levels. Many global acting big tech firms show such behavior (Egan, 2025; Trax Technologies, 2025). As job losses accumulate, income inequality and social insecurity rise, placing greater strain on public welfare systems (Heinz, 2025; Bell & Korinek, 2023). These effects can reinforce themselves over time, as reduced economic stability that further limits opportunities for affected groups (Lei & Kim. 2024).

At the same time, automation reshapes rather than simply eliminates work. Through organizational learning, firms gradually generate demand for new roles (**R6**) that complement automated systems, such as in data analysis and human–machine coordination. However, this shift often occurs only after a time delay, during which displacement effects dominate. The resulting opportunities are uneven and require different skills than those displaced, creating structural mismatches. Without sufficient reskilling and institutional support, this transition can still reinforce inequality despite eventual job creation. Recently IBM started recruiting many juniors employees due to limitations to AI – adoption (Gonzales, 2026) as an example of these favorable effects.

A further balancing mechanism, **B11** (Displacement–security risk loop), links labor market outcomes back to security dynamics. Workforce displacement (Nabin, 2026; Idrisi, et al., 2024) creates a larger pool of affected Individuals. One particular group that can do material harm to organizations is the disgruntled employee (Yogi & Mundru, 2022; Rauf et a., 2023) which can be strengthened by these effects. A disgruntled employee is an (ex-) worker who is dissatisfied or unhappy with their job, often leading to negative attitudes or behavior, or even adverse behavior (ibid.). Particularly technically skilled individuals can expand the pool of such individuals that are vulnerable to participation in illicit or adversarial automated activities. This can increase the capability of malicious actors, feeding back into the broader security arms race.

These interacting loops correspond to established system dynamics archetypes. The security domain reflects an Escalation structure, while the reliance on automation as a response to cost and threat pressures exhibits features of Fixes that Fail, where short-term efficiency gains generate longer-term social and security risks. Together, the reinforcing loops (**R4–R6**) and the

balancing mechanism (**B9 – B11**) create a tightly coupled system in which technological progress, labor disruption, and security risks evolve simultaneously.

The Raise of the Augmented Insider

Although literature recognizes the insider threat caused by discretion employees (Moore et al, 2018; Martinez-Moyano et al., 2008; Colwill, 2009; Akeju et al., 2018) the recognition of the augmented insider has not been established. Here our expert consultation makes a significant contribution. A further dynamic emerges at the intersection of workforce displacement and security risk, where knowledge and exclusion interact to create a new category of threat. The model structures are also visible in Figure 4.

This is captured in **B12** (Exclusion–grievance loop). Organizational restructuring and AI-driven automation increase the number of displaced IT professionals. Displacement can generate feelings of abandonment, exclusion, and loss of professional identity. These conditions may increase the motivation to misuse retained knowledge, not necessarily driven by ideology, but by perceived injustice or marginalization. As a result, the likelihood of insider-like threats increases. Increasing adversarial effects lowers the value of AI and therefore increases the costs of AI adoption.

This risk is amplified by **B13** (Knowledge-enabled threat loop). These displaced professionals retain deep understanding of systems, architectures, and vulnerabilities. This knowledge lowers the barriers to executing high-impact attacks, increasing both the effectiveness and potential damage of malicious actions. The combination of capability and motivation gives rise to what can be described as an “augmented insider”: an actor who operates with insider-level knowledge but outside formal organizational boundaries.

A third balancing mechanism, **B14** (Defensive exclusion loop), arises from organizational responses to such threats (Zeijlemaker et al, 2025; Zeijlemaker & Siegel, 2023). Security incidents typically lead to stronger monitoring, tighter controls, and more restrictive access policies. While these measures aim to reduce risk, they can also create a culture of distrust and exclusion if implemented without consideration of their broader impact. This can weaken trust and belonging among both current and former employees, reinforcing the very grievances that contribute to insider risk.

At the same time, a reinforcing mechanism is present in **R7** (Threat suppression loop). Enhanced controls and monitoring can reduce the immediate likelihood and success of attacks by making exploitation more difficult (Zeijlemaker & Siegel, 2023). However, this effect may be temporary if underlying social drivers remain unaddressed.

These dynamics align with established system dynamics archetypes. The pattern reflects Fixes that Fail, where increased controls provide short-term risk reduction but may unintentionally deepen exclusion and future vulnerability. Elements of Shifting the Burden are also evident, as organizations rely on technical controls while neglecting the underlying social and organizational causes of risk.

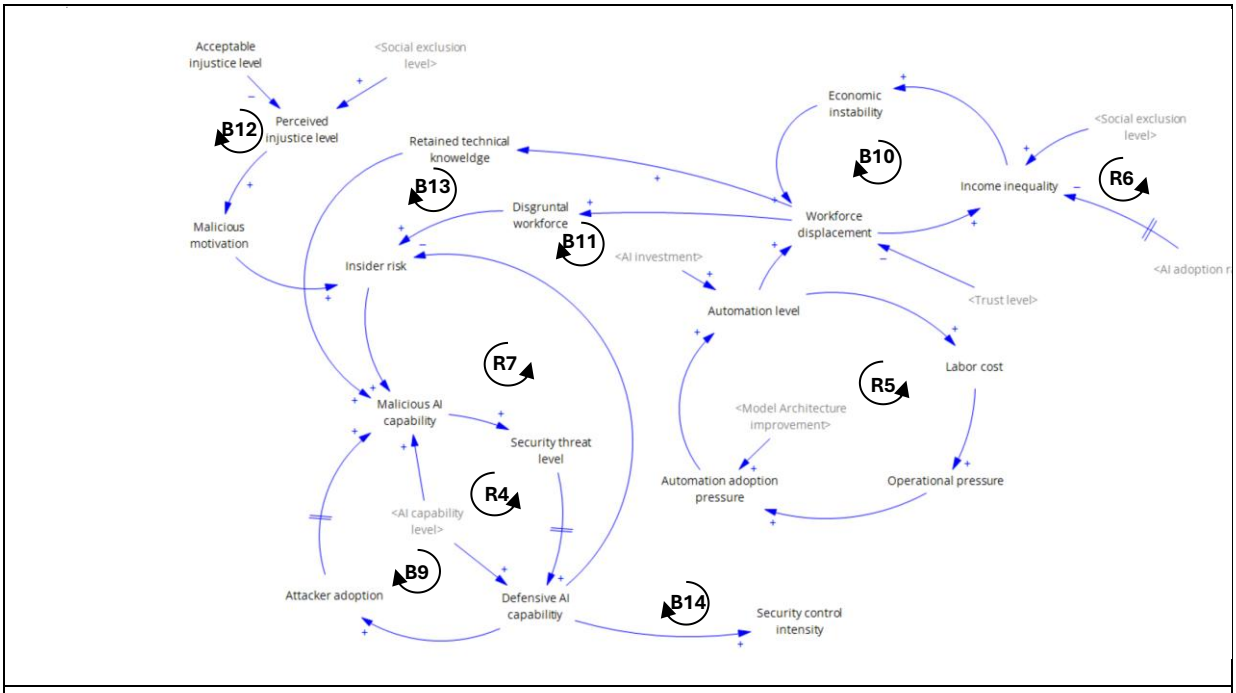

Figure 4. Security Escalation and Risk Automation (incl the rise of augmented insider)

Taken together, the interaction between reinforcing loop (**R7**) and the balancing mechanism (**B12– B14**) suggests that managing the augmented insider threat requires not only stronger security measures, but also careful attention to workforce transitions, trust, and *inclusion*.

Systemic Structure Summary

The systemic structure of AI adoption reveals a tightly interwoven set of economic, social, and security dynamics that evolve through interacting reinforcing and balancing feedback loops. What begins as a productivity-driven investment cycle quickly expands into a broader system in which gains in capability are counterbalanced by rising costs, diminishing differentiation, and delayed returns. This economic "limits to growth" dynamic does not operate in isolation; it is deeply entangled with social and institutional consequences.

As AI diffuses, reinforcing mechanisms of bias, exclusion, and power concentration shapes who benefit from these technologies and who is left behind. Declining trust and participation reduce the diversity and quality of data, further degrading system performance, and legitimacy. At the same time, security dynamics introduce an additional layer of complexity, where AI accelerates both defensive and offensive capabilities, creating an ongoing escalation cycle. Efforts to improve efficiency through automation simultaneously generate labor displacement and inequality, feeding back into both social instability and security risk.

The emergence of the "augmented insider" illustrates how these domains converge. Economic displacement, social exclusion, and retained technical knowledge combine to produce new forms of risk that cannot be addressed through technical controls alone. Instead, short-term

fixes often reinforce long-term vulnerabilities, highlighting classic system dynamics patterns such as escalation, fixes that fail, and shifting the burden.

Overall, AI adoption should be understood not as a linear technological transition, but as a co-evolving system. Sustainable outcomes will depend on breaking reinforcing cycles through coordinated interventions that will balance innovation with inclusion, efficiency with resilience, and security with trust.

Way forward to foster global AI adoption

The systemic analysis of AI adoption highlights that isolated or purely technical interventions are insufficient to address the reinforcing dynamics of inequality, security risk, and diminishing economic returns. Effective policy responses must therefore adopt an integrated and inclusive lens that simultaneously addresses ethical, social, economic, and security dimensions. This is consistent with the model's core insight that system behaviour emerges from interconnected feedback loops rather than linear cause–effect relationships.

First, the establishment of robust ethical standards is foundational. Policies must move beyond compliance toward embedded accountability mechanisms that address bias, transparency, and fairness throughout the AI lifecycle. This directly weakens reinforcing loops of bias and distrust by improving data diversity, decision explainability, and stakeholder inclusion. However, ethical frameworks alone are insufficient without enforceable governance structures. Governments and organizations should implement adaptive AI governance models that combine regulation, auditability, and continuous monitoring, ensuring that innovation remains aligned with societal values while maintaining strategic flexibility.

Second, workforce transformation must be treated as a strategic priority rather than a secondary effect. The analysis shows that automation-driven displacement reinforces inequality and security risks. Targeted reskilling and upskilling programs, particularly for vulnerable and transitioning workers, are critical to breaking these cycles. Importantly, these initiatives should not only focus on technical skills but also on interdisciplinary capabilities that enable human–AI collaboration. A diverse workforce further enhances system resilience by improving model design, reducing bias, and strengthening organizational adaptability.

Third, equitable access to AI capabilities is essential to counteract power concentration. Policies should promote shared infrastructure, open innovation ecosystems, and public–private partnerships that lower barriers to entry. This ensures that the benefits of AI are more broadly distributed, preventing the entrenchment of dominant actors and supporting more competitive and inclusive markets.

Fourth, security must be reframed as a socio-technical challenge. While investments in AI-driven defense systems are necessary, they must be complemented by organizational policies that address trust, inclusion, and workforce stability. Mitigating the "augmented insider" risk requires combining technical safeguards with human-centered approaches, including transparent communication, fair transition processes, and cultural integration.

Ultimately, strategic effectiveness lies in aligning innovation with long-term societal value. Policies that integrate ethical governance, workforce inclusion, equitable access, and systemic security can dampen reinforcing risks while strengthening sustainable growth. Such an approach not only stabilizes the AI adoption trajectory but also enhances overall quality of life by ensuring that technological progress translates into shared economic and social benefits.

Reflections and Future Research Directions

This study provides a comprehensive systemic perspective on AI adoption by integrating economic, social, and security dynamics into a coherent feedback structure. A key strength lies in its participatory model-building approach, which incorporates multidisciplinary expert insights and ensures that the model reflects real-world complexity rather than abstract assumptions. The identification of interconnected reinforcing loops and particularly the integration of the "augmented insider" concept, represents a novel contribution that bridges workforce dynamics and cybersecurity risk.

However, expert feedback highlights that the current model, while robust in capturing reinforcing dynamics, underrepresents stabilizing mechanisms and institutional safeguards. This creates a potential structural bias toward risk amplification without fully explaining how systems can regain equilibrium. Future iterations should therefore explicitly incorporate balancing loops such as reskilling systems, inclusive governance structures, legal accountability, and government interventions to market concentrations. Also, the integration of a unifying "inclusion–resilience" loop, as highlighted in the expert feedback, provides a powerful bridging mechanism across different system sub-models, positioning inclusion not as a peripheral consideration but as a fundamental driver of long-term system stability and resilience. These additions would improve the explanatory power of the model by capturing not only how risks emerge, but also how they can be mitigated and managed over time.

In addition to structural refinement, a critical next step is model quantification and formalization. The current qualitative model captures causal relationships but does not yet allow for simulation or advanced scenario analysis. Future research should translate key variables into measurable constructs, define functional relationships, and estimate parameters using empirical data. This would enable the development of stock-and-flow models capable of testing policy interventions, exploring time delays, and identifying tipping points within the system. Hybrid approaches combining system dynamics with agent-based modeling or AI-driven simulation could further enhance analytical depth.

Finally, longitudinal validation and empirical grounding will be essential. Future work should incorporate case studies, industry data, and policy experiments to evaluate model assumptions and refine its predictive capacity. By advancing both structural completeness and quantitative rigor, the model can evolve into a decision-support tool that enables organizations and policymakers to navigate AI adoption in a way that is not only economically effective, but also socially inclusive and systemically resilient.

Literature

Acemoglu, D., & Restrepo, P. (2019). *Automation and new tasks: How technology displaces and reinstates labor*. **Journal of Economic Perspectives, 33**(2), 3–30. https://doi.org/10.1257/jep.33.2.3

Ammachchi, N. (2025). *U.S. cancels $5.1 billion in IT contracts, citing cost-cutting measures*. Nearshore Americas. https://nearshoreamericas.com/u-s-cancels-5-1-billion-in-it-contracts-citing-cost-cutting-measures/

Annic Kilcher. (2026). *$1 billion funding to find new kind of AI – not LLM scaling* [Video]. YouTube. https://www.youtube.com/watch?v=fL-lohKjELI

Artificial Intelligence Act. (n.d.). *EU Artificial Intelligence Act*. https://artificialintelligenceact.eu/

Bachlakova, P. (2025). *The market for spyware is growing – and it's being used differently against women*. The Fuller Project. https://www.fullerproject.org/editions/revolutions/spyware-women-activists-palantir-pegasus-transnational-repression/

Bell, S.A., & Korinek, A. (2023). AI's Economic Peril. *Journal of Democracy 34*(4), 151-161. https://dx.doi.org/10.1353/jod.2023.a907696.

Bhardwaj, E., Alexander, R., & Becker, C. (2025). *Limits to AI growth: The ecological and social consequences of scaling*. arXiv. https://doi.org/10.48550/arXiv.2501.17980

Bisola Akeju, Joseph Edivri, Jolly I. Ogbole, Precious Osobhalenewie Okoruwa, Oladapo Fadayomi, Toyosi O Abolaji "Conceptual Model for Insider Threat Classification and Risk Modeling in Complex Digital Systems" Iconic Research And Engineering Journals Volume 1 Issue 9 2018 Page 476-492 https://doi.org/10.64388/IREV1I9-1713778

Boston Consulting Group. (2024, October 24). *AI adoption in 2024: 74% of companies struggle to achieve and scale value*. https://www.bcg.com/press/24october2024-ai-adoption-in-2024-74-of-companies-struggle-to-achieve-and-scale-value

Carl Colwill, Human factors in information security: The insider threat – Who can you trust these days?, Information Security Technical Report, Volume 14, Issue 4, 2009, Pages 186-196, ISSN 1363-4127, https://doi.org/10.1016/j.istr.2010.04.004.

Daron Acemoglu, D., & Pascual Restrepo, P. (2018). *Artificial intelligence, automation, and work*. **SSRN Electronic Journal**. https://doi.org/10.2139/ssrn.3098384

Dave, P. (2026, March 9). *Anthropic claims Pentagon feud could cost it billions*. Wired. https://www.wired.com/story/anthropic-claims-business-is-in-peril-due-to-supply-chain-risk-designation/

Dobrygowski, D., & Valkhof, B. (2024, November 14). *Explainer: What is digital trust in the intelligent age?* World Economic Forum.

Dzreke, S. S., & Dzreke, S. E. (2026). *The compound advantage: AI as the engine for unassailable business growth*. *Engineering Science & Technology Journal, 7*(1), 19–35. https://doi.org/10.51594/estj.v7i1.2190

Edwards, J. (2026, January 10). *"Big Short" investor Michael Burry makes put-options bet against Oracle stock (ORCL).* Fortune. https://fortune.com/2026/01/10/big-short-investor-michael-burry-put-options-bet-oracle-stock-orcl/

Egan, M. (2025, October 19). *Firms are blaming AI for job cuts. Critics say it's a "good excuse."* **CNBC**. https://www.cnbc.com/2025/10/19/firms-are-blaming-ai-for-job-cuts-critics-say-its-a-good-excuse.html

ETtech. (2025, August 14). *AI mishaps hit 95% executives, only 2% firms meet responsible use standards: Infosys study*. The Economic Times. https://economictimes.indiatimes.com/tech/artificial-intelligence/ai-mishaps-hit-95-executives-only-2-firms-meet-responsible-use-standards-infosys-study/articleshow/123305693.cms

Farrukh Avezov. (2025). The Economic Impact of AI Adoption: Measuring Productivity and Competitive Advantage in International Enterprises. The American Journal of Management and Economics Innovations, 7(06), 22–29. https://doi.org/10.37547/tajmei/Volume07Issue06-04.

Ford, A. (2010). Modeling the Environment. Island Press.

Ford, D. N., & Sterman, J. D. (1998). Expert knowledge elicitation to improve formal and mental models. System Dynamics Review, 14(4), 309–340.

Forrester, J. W. (1961). Industrial Dynamics. Pegasus Communications.

Gartner, Inc. (2025, June 25). *Gartner predicts over 40% of agentic AI projects will be canceled by end of 2027*. https://www.gartner.com/en/newsroom/press-releases/2025-06-25-gartner-predicts-over-40-percent-of-agentic-ai-projects-will-be-canceled-by-end-of-2027

González, L. (2026, February 13). *Gen Z jobs aren't dead yet: $240 billion tech giant IBM says it's rewriting entry-level jobs—and tripling down on hiring young workers*. **Fortune**. https://fortune.com/2026/02/13/tech-giant-ibm-tripling-gen-z-entry-level-hiring-according-to-chro-rewriting-jobs-ai-era/

Idrisi, M.J., Geteye, D. & Shanmugasundaram, P. Modeling the Complex Interplay: Dynamics of Job Displacement and Evolution of Artificial Intelligence in a Socio-Economic Landscape. *Int J Netw Distrib Comput* **12**, 185–194 (2024). https://doi.org/10.1007/s44227-024-00025-0

Ignacio J. Martinez-Moyano, Eliot Rich, Stephen Conrad, David F. Andersen, and Thomas R. Stewart. 2008. A behavioral theory of insider-threat risks: A system dynamics approach. ACM Trans. Model. Comput. Simul. 18, 2, Article 7 (April 2008), 27 pages. https://doi.org/10.1145/1346325.1346328

Industry Dive. (2025). *Execs use responsible AI to drive growth, prevent risks*. Yahoo Finance. https://finance.yahoo.com/news/execs-responsible-ai-drive-growth-070000174.html

Joffe, M. (2021). Equilibrium, Instability, Growth and Feedback in Economics. In: Cavana, R.Y., Dangerfield, B.C., Pavlov, O.V., Radzicki, M.J., Wheat, I.D. (eds) Feedback Economics. Contemporary Systems Thinking. Springer, Cham. https://doi.org/10.1007/978-3-030-67190-7_3

Leavy, S., O'Sullivan, B., & Siapera, E. (2020). Data, Power, and Bias in Artificial Intelligence. *ArXiv, abs/2008.07341*.

Lei, Y.-W., & Kim, R. (2024). *Automation and augmentation: Artificial intelligence, robots, and work*. **Annual Review of Sociology, 50**(1), 251–272. https://doi.org/10.1146/annurev-soc-090523-050708

Karl Jochen Heinz. The AI consumption spiral: an underestimated systemic risk for economy and society. International Journal of Research and Review. 2025; 12(10): 77-87. DOI: https://doi.org/10.52403/ijrr.20251009

McCordick, J. (2026, March 13). *Palantir CEO makes shocking confession on disrupting Democratic power*. The New Republic. https://newrepublic.com/post/207693/palantir-ceo-karp-disrupting-democratic-power

McKinsey & Company. (2024, September 17). *How data centers and the energy sector can sate AI's hunger for power*. https://www.mckinsey.com/industries/private-capital/our-insights/how-data-centers-and-the-energy-sector-can-sate-ais-hunger-for-power

P. Moore, T. M. Cassidy, M. C. Theis, D. Bauer, D. M. Rousseau, and S. B. Moore, "Balancing Organizational Incentives to Counter Insider Threat," *2018 IEEE Security and Privacy Workshops (SPW)*, San Francisco, CA, USA, 2018, pp. 237-246, doi: 10.1109/SPW.2018.00039.

Nabin, M. H. (2026). A Diagnostic Framework for Socially Sustainable AI Diffusion. *Sustainability*, *18*(3), 1153. https://doi.org/10.3390/su18031153

Nicolò Pagan, Joachim Baumann, Ezzat Elokda, Giulia De Pasquale, Saverio Bolognani, and Anikó Hannák. 2023. A Classification of Feedback Loops and Their Relation to Biases in Automated Decision-Making Systems. In Proceedings of the 3rd ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization (EAAMO '23). Association for Computing Machinery, New York, NY, USA, Article 7, 1–14. https://doi.org/10.1145/3617694.3623227

Pengwei, Z., & Sun, Y. (2025). *Assessing impact of strategic investments on development of sustainable artificial intelligence dynamics*. *Sustainable Futures, 10*, 100884. https://doi.org/10.1016/j.sftr.2025.100884

Pruyt, E. (2013). Small system dynamics Models for Big Issues: Triple Jump towards Real World complexity. Delft: TU Delft Library. 324.

Rauf, U., Mohsen, F., & Wei, Z. (2023). *A taxonomic classification of insider threats: Existing techniques, future directions & recommendations*. **Journal of Cyber Security and Mobility, 12**, 221–252. https://doi.org/10.13052/jcsm2245-1439.1225

Richardson, G., & Pugh, A. (1981). Introduction to system dynamics modeling with DYNAMO. Cambridge.

Shams, R.A., Zowghi, D. & Bano, M. AI and the quest for diversity and inclusion: a systematic literature review. *AI Ethics* **5**, 411–438 (2025). https://doi.org/10.1007/s43681-023-00362-w

Shilov, A. (2026, March 8). *OpenAI's massive Stargate data center canceled as firm can't reach terms with Oracle, operator struggles with reliability issues*. Tom's Hardware.

https://www.tomshardware.com/tech-industry/artificial-intelligence/openais-massive-stargate-data-center-canceled-as-firm-cant-reach-terms-with-oracle-operator-struggles-with-reliability-issues-meta-said-to-be-interested-in-snatching-excess-capacity

Sterman, J. (2000). Business Dynamics: System thinking and modelling for a complex world. Irwin MC Graw-Hill.

Sven Meier, Pratik Narendra Raut, Felix Mahr, Nils Thielen, Jörg Franke, Florian Risch, Structured knowledge-based causal discovery: Agentic streams of thought, Information Processing & Management, Volume 62, Issue 5, 2025, 104202, ISSN 0306-4573, https://doi.org/10.1016/j.ipm.2025.104202.

Trax Technologies. (2025, November 7). *Tens of thousands of layoffs blamed on AI—But where are the actual returns?* https://www.traxtech.com/ai-in-supply-chain/tens-of-thousands-of-layoffs-blamed-on-ai-but-where-are-the-actual-returns

Wang, C., Wan, Z., Kang, H., Chen, E., Xie, Z., Krishna, T., Reddi, V. J., & Du, Y. (2025). *Slm-mux: Orchestrating small language models for reasoning*. arXiv. https://doi.org/10.48550/arXiv.2510.05077

Wolfe, D. A., Choe, A., & Kidd, F. (2025). *The architecture of AI transformation: Four strategic patterns and an emerging frontier*. arXiv. https://doi.org/10.48550/arXiv.2509.02853

Yogi, M. K., & Mundru, Y. (2022). *Detecting insider threat through psychometric scores and work environment*. **Journal of Soft Computing Paradigm, 4**(3), 200–212. https://doi.org/10.36548/jscp.2022.3.008

Zeijlemaker, Sander and Siegel, Michael, "Capturing the Dynamic Nature of Cyber Risk: Evidence from an Explorative Case Study" (2023). *Hawaii International Conference on System Sciences 2023 (HICSS-56)*. 5. https://aisel.aisnet.org/hicss-56/os/cybersecurity/5

Zeijlemaker S, Lemiesa YK, Schröer SL, Abhishta A, Siegel M. How Does AI Transform Cyber Risk Management? *Systems*. 2025; 13(10):835. https://doi.org/10.3390/systems13100835