IACIS
Proceedings of the 66th International Association for Computer Information Systems Conference - October 7 - 10, 2026 | Pensacola, FL

ARTIFICIAL INTELLIGENCE, NEWS-FINDS-ME PERCEPTION, AND PERCEIVED INFORMEDNESS: AN EXPLORATORY PLS-SEM STUDY ACROSS PUERTO RICO AND THE DOMINICAN REPUBLIC

Carlos Javier Sánchez, AGM University - Online, cjsanchez@agmu.edu
Rafael Padilla Vega, Universidad AGM - Gurabo, padillar1@agmu.edu
Angel Ojeda Castro, Universidad AGM - Gurabo, ojedaa1@agmu.edu

ABSTRACT

This study examines how artificial intelligence (AI) relates to Perceived Informedness and evaluates the explanatory role of the News-Finds-Me Perception (NFM) in algorithmically mediated information environments. Data were collected from a pooled sample of 434 respondents residing in Puerto Rico (n = 222, 51.2%) and the Dominican Republic (n = 212, 48.8%) and analyzed using Partial Least Squares Structural Equation Modeling (PLS-SEM). The results indicate that AI Perception significantly predicts Perceived Informedness ($\beta = 0.41$, $p = .006$), whereas the path from NFM to Perceived Informedness was not supported ($\beta = 0.14$, $p = .095$). The relationship between AI Perception and NFM likewise did not emerge as significant ($\beta = -0.004$, $p = .754$), suggesting that algorithmic exposure and passive news-consumption beliefs operate through distinct pathways within this sample. Social Media Exposure showed a significant, though modest, effect on NFM ($\beta = 0.24$, $p = .037$). Within this exploratory, pooled Caribbean sample, the findings suggest that NFM offers limited explanatory value in AI-mediated environments and point toward the need to incorporate algorithmic constructs into information behavior models.

Keywords: Artificial Intelligence; News-Finds-Me Perception; Algorithmic Curation; Perceived Informedness; PLS-SEM; Social Media

INTRODUCTION

Digital technologies have transformed how individuals access and process information, particularly in platform-based environments where content is curated rather than actively sought (Castells, 2010; Van Dijck, 2013). Artificial intelligence (AI) occupies a central role in this transformation, enabling algorithmic filtering, personalization, and automated content delivery (Ågerfalk, 2020; Davenport & Ronanki, 2018).

The News-Finds-Me Perception (NFM) offers a theoretical account of passive information consumption, describing the belief that individuals can remain informed without actively searching for news (Gil de Zúñiga et al., 2017; Strauß et al., 2021). This perception has been associated with lower levels of knowledge acquisition and greater reliance on social cues (Lazer et al., 2018; Tucker et al., 2018).

This study proposes a model that integrates social and algorithmic pathways of information exposure. Social Media Exposure is expected to influence NFM, consistent with the mechanisms through which passive information beliefs are formed (Gil de Zúñiga et al., 2017; Van Dijck, 2013). AI Perception, in turn, is conceptualized as a direct driver of Perceived Informedness,

reflecting the mediating role of algorithmic curation (Ågerfalk, 2020; Davenport & Ronanki, 2018), while a further path tests whether AI Perception also shapes NFM directly.

NFM was originally conceptualized within a social media context in which information diffusion occurs through interpersonal networks. Contemporary digital environments, however, are increasingly shaped by algorithmic systems that actively curate information rather than merely facilitating its diffusion (Pariser, 2011; Bodó et al., 2019). This shift raises the question of whether Perceived Informedness continues to be driven primarily by passive, network-based exposure or is increasingly shaped by algorithmic systems.

This study addresses this question by examining the relationships among AI Perception, Social Media Exposure, NFM, and Perceived Informedness. Data were collected from respondents in Puerto Rico and the Dominican Republic and pooled into a single dataset to develop an exploratory regional model; the study does not conduct a formal multi-group comparison between the two countries, and the findings should be interpreted as reflecting shared patterns across the pooled sample rather than country-specific effects.

RQ1: Does AI Perception influence Perceived Informedness?
RQ2: Does AI Perception influence NFM?
RQ3: Does NFM influence Perceived Informedness?
RQ4: Does Social Media Exposure influence NFM?

This study contributes to the information systems literature by empirically evaluating whether algorithmic systems complement or operate independently of traditional passive information mechanisms, and by offering an early PLS-SEM application of the NFM framework in an underrepresented Caribbean context.

LITERATURE REVIEW

The News-Finds-Me Perception (NFM) has been widely used to explain passive information consumption in digital environments (Gil de Zúñiga et al., 2017; Strauß et al., 2021). Individuals with high levels of NFM tend to rely on incidental exposure rather than active information seeking, which has been associated with lower levels of political knowledge (Lazer et al., 2018; Tucker et al., 2018).

Artificial intelligence has transformed information systems by acting as a digital agent that curates and delivers content (Ågerfalk, 2020; Seaver, 2017). Algorithmic systems prioritize content based on engagement and behavioral patterns, shaping what users see and how they interpret information (Pariser, 2011; Bodó et al., 2019), a process that actively constructs the information environment rather than simply facilitating access to it (Van Dijck, 2013; Castells, 2010); as a result, users may develop a sense of being informed based on curated content rather than deliberate engagement.
Perceived Informedness reflects a subjective sense of knowledge that may not correspond to actual understanding (Lazer et al., 2018). Algorithmic systems may reinforce this perception by repeatedly exposing users to relevant content, strengthening familiarity and perceived competence (Shin, 2021; Glikson & Woolley, 2020). Two competing mechanisms are suggested in the literature: AI may reinforce passive exposure by making information more accessible, or it may

shape Perceived Informedness through a pathway independent of passive exposure (Ågerfalk, 2020; Seaver, 2017). This study evaluates these competing explanations empirically.

Trust in algorithmic systems has also become central to understanding digital information behavior (Glikson & Woolley, 2020; Shin, 2021). As individuals rely increasingly on AI-driven systems, trust in these technologies may substitute traditional mechanisms of information verification (Venkatesh et al., 2012). Algorithmic systems further introduce issues of transparency and explainability that shape how users interpret curated content (Shin, 2020); a limited understanding of how algorithms operate may lead users to overestimate the reliability of curated content, reinforcing perceptions of being informed (Koohang et al., 2023; Davenport & Ronanki, 2018).

Presentation of the Main Contribution / Research Model

The proposed model reflects the coexistence of two mechanisms of information exposure: social diffusion and algorithmic mediation. While NFM captures passive exposure through social networks, AI Perception captures the growing role of automated systems in shaping informational environments (Ågerfalk, 2020; Seaver, 2017). Although social media platforms are themselves algorithmically curated, AI Perception and Social Media Exposure are conceptually distinct: the former captures respondents' awareness of algorithmic mediation itself, while the latter captures reliance on social/messaging channels for political news and a related source-verification behavior, regardless of whether that mediation is recognized. This distinction is examined empirically through discriminant validity testing (see Discriminant Validity).

The model specifies the following relationships: Social Media Exposure → NFM; AI Perception → NFM; NFM → Perceived Informedness; and AI Perception → Perceived Informedness. Based on this model, the following hypotheses are proposed:

H1: AI Perception positively influences Perceived Informedness.
H2: AI Perception positively influences NFM.
H3: NFM positively influences Perceived Informedness.
H4: Social Media Exposure positively influences NFM.

Figure 1 presents the conceptual research model.

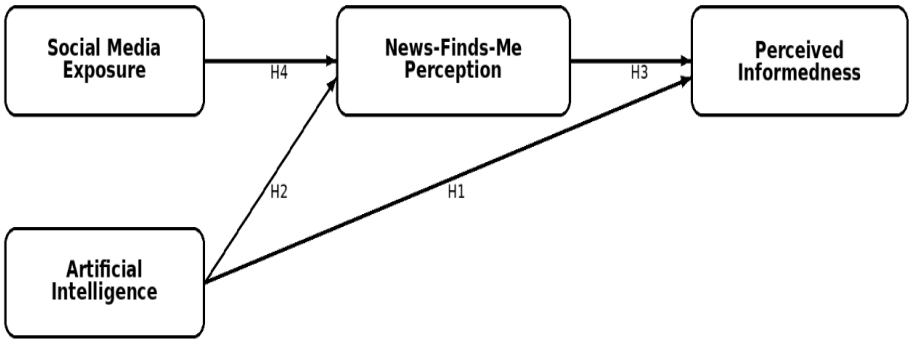


Figure 1. Conceptual research model.

Conceptualizing the Artificial Intelligence Perception Construct

AI Perception, as specified in this model, is a reflective latent construct capturing individuals' subjective awareness of algorithmic mediation in their information environment. The construct is anchored in five interrelated facets: (a) awareness of algorithmic curation, understood as the recognition that content is filtered and prioritized by automated systems rather than presented neutrally; (b) the perception that AI actively shapes the political information individuals encounter; (c) the perception of AI-generated political content, including synthetic or algorithmically assembled material; (d) perceived algorithmic intervention in the sequencing and visibility of information; and (e) awareness of AI-assisted recommendation systems that surface content based on inferred user preferences.

These facets are treated as reflective indicators of a single underlying perception because they are expected to co-vary as manifestations of one latent belief about algorithmic mediation, consistent with the measurement logic applied to NFM by Gil de Zúñiga et al. (2017). This specification distinguishes AI Perception from adjacent constructs such as algorithmic trust or explainability, which capture evaluative rather than perceptual dimensions of algorithmic mediation (Shin, 2020; Glikson & Woolley, 2020), and from Social Media Exposure, which captures behavioral frequency of platform use rather than awareness of algorithmic processes.

METHODOLOGY

Research Design

This study employs a quantitative, cross-sectional, explanatory research design consistent with theory-testing objectives in information systems research (Hair et al., 2019). The explanatory orientation reflects the study's aim of testing directional hypotheses derived from an a priori conceptual model rather than generating theory inductively. Partial Least Squares Structural Equation Modeling (PLS-SEM) was selected as the analytical technique because of its suitability for prediction-oriented research involving reflective constructs, exploratory theoretical extensions, and moderate sample sizes (Hair et al., 2019; Sarstedt et al., 2020).

Sampling

Data were collected from a pooled sample of 434 respondents, comprising 222 participants from Puerto Rico (51.2%) and 212 participants from the Dominican Republic (48.8%). Respondents were reached using a Random Digit Dialing (RDD) telephone methodology, generating numbers to broadly cover residential and mobile telephone users in each country. Trained interviewers administered a standardized structured questionnaire in Spanish and entered responses directly into the survey system during the call. Inclusion criteria required that respondents be adults (18 years of age or older) and residents of Puerto Rico or the Dominican Republic; participation was voluntary and anonymous, with verbal informed consent obtained before each interview began.

A pilot test involving 15 telephone interviews was conducted prior to full-scale data collection to assess item clarity, respondent comprehension, and interview length. Pilot feedback indicated ambiguity between AI Perception items referring to algorithmic curation versus general technology use; wording was revised accordingly before the interview protocol was finalized.

Consistent with the exploratory, pooled design of this study, the Puerto Rico and Dominican Republic subsamples were combined into a single dataset for model estimation rather than analyzed as independent groups. This approach was adopted to develop a regional model of AI-related information behavior rather than to test for cross-country differences. Formal multi-group comparison (e.g., measurement invariance testing and PLS multi-group analysis) was outside the scope of the present analysis; accordingly, the comparative framing implied by earlier drafts of this manuscript has been removed from the title and abstract, and the study is presented as an exploratory, pooled-sample investigation.

Data Collection Procedure

Data collection followed a computer-assisted telephone interviewing (CATI) protocol built around Random Digit Dialing. Interviewers received standardized training on the interview script, neutral probing, and handling refusals and callbacks, ensuring consistent administration across both countries. Quality control included supervisor monitoring of a call subset and periodic verification of completed interviews for protocol adherence. Respondents were informed of the study's academic purpose, advised that participation was voluntary and could be discontinued at any time, and assured that responses would remain anonymous and reported only in aggregate form. Verbal informed consent was recorded before each interview, no personally identifying information was retained, and completed interviews were transferred directly into the analysis dataset.

Instrument

The questionnaire measured four constructs using Likert-type items (ranging from strongly disagree to strongly agree): AI Perception, News-Finds-Me Perception (NFM), Social Media Exposure, and Perceived Informedness (Venkatesh et al., 2012). AI Perception was operationalized according to the five facets described above. NFM was measured using items adapted from the original News-Finds-Me scale, capturing the belief that one can remain informed through incidental, network-based exposure without actively seeking news (Gil de Zúñiga et al., 2017). Social Media Exposure was operationalized with two retained indicators: reliance on social media or messaging platforms as a primary news channel, and a self-reported source-verification behavior (Gil de Zúñiga et al., 2017; Van Dijck, 2013); both exhibited strong outer loadings (above 0.89), though their conceptual heterogeneity is addressed as a measurement limitation below. Perceived Informedness captured respondents' subjective sense of being knowledgeable about current events and public affairs, independent of objective knowledge measures (Lazer et al., 2018). Each item was read aloud by the interviewer exactly as written, and answers were recorded in real time during the call.

Table 1. Construct Definitions and Sources

Construct	Definition	Source
AI Perception	Subjective awareness of algorithmic curation, AI-mediated political content, and AI-assisted recommendation systems in one's information environment.	Adapted for this study; informed by conceptual work on algorithmic mediation

Construct	Definition	Source
		(Ågerfalk, 2020; Seaver, 2017)
News-Finds-Me Perception (NFM)	Belief that one can remain adequately informed through incidental, network-based exposure without actively seeking news.	Adapted from Gil de Zúñiga et al. (2017)
Social Media Exposure	Reliance on social media/messaging platforms as a primary channel for political news, together with a self-reported source-verification behavior retained in the same two-indicator reflective construct (see Limitations).	Adapted from Gil de Zúñiga et al. (2017); Van Dijck (2013)
Perceived Informedness	Subjective sense of being knowledgeable about current events and public affairs, independent of objective knowledge.	Adapted from subjective informedness measures in political communication research (Lazer et al., 2018)

Perceived Informedness was modeled as a reflective latent construct comprising three indicators capturing respondents' subjective perception of being informed about current public and political affairs. Standardized outer loadings exceeded 0.90, supporting the construct's reliability within the proposed model. Table 1 summarizes the definition and source of each construct.

PLS-SEM Data Analysis Procedure

Data analysis followed the two-stage PLS-SEM procedure recommended by Hair et al. (2019). The measurement model (outer model) specifies the relationships between each latent construct and its observed indicators and was evaluated prior to structural model testing. Indicator reliability was assessed through outer loadings, with values above 0.70 considered ideal and values above 0.60 acceptable for exploratory research (Hair et al., 2019; Sarstedt et al., 2020). Internal consistency reliability was assessed using Cronbach's alpha and Composite Reliability (CR), both of which estimate the extent to which the indicators of a construct consistently measure the same underlying concept, with 0.70 generally considered an acceptable threshold (Hair et al., 2019). Convergent validity was assessed through the Average Variance Extracted (AVE), representing the average proportion of indicator variance explained by its associated construct; values above 0.50 indicate that a construct explains more than half of the variance in its indicators (Hair et al., 2019; Henseler et al., 2009).

Discriminant validity, which assesses whether constructs are empirically distinct from one another, was evaluated using the Heterotrait-Monotrait ratio of correlations (HTMT); ratios below 0.85 (or the more conservative 0.90 threshold) indicate adequate discriminant validity (Hair et al., 2019). This test is particularly important given the conceptual proximity of AI Perception and Social Media Exposure noted above.

The structural model (inner model) specifies the hypothesized directional relationships among latent constructs and was evaluated using standardized path coefficients (β), indicating the strength and direction of each relationship. Statistical significance was assessed via bootstrapping with 5,000 subsamples, generating empirical t-statistics and p-values without relying on distributional assumptions (Hair et al., 2019). Multicollinearity among predictors was assessed using the Variance Inflation Factor (VIF), with values below 5 (ideally below 3) indicating that collinearity does not bias structural estimates (Hair et al., 2019). Explanatory power was assessed using the coefficient of determination (R^2), while effect size (f^2) quantifies each predictor's contribution to a given R^2. Predictive relevance was assessed using the Stone-Geisser Q^2 statistic obtained through blindfolding, with values above zero indicating predictive relevance (Hair et al., 2019; Sarstedt et al., 2020).

Because AI Perception was hypothesized to relate to Perceived Informedness both directly and indirectly through NFM, the indirect path AI Perception → NFM → Perceived Informedness was formally tested using the same bootstrapping procedure. As reported in the Results section, this indirect effect was small and non-significant, and mediation is not supported by the present data. This more conservative interpretation supersedes claims in earlier drafts of this manuscript that AI Perception "bypasses" or "replaces" NFM as a mediating mechanism.

Figure 2 presents the sequential methodological workflow employed in this study, from initial survey design through hypothesis testing.

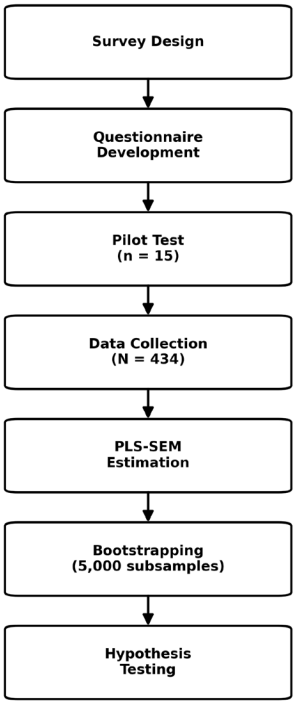


Figure 2. Methodological workflow.

RESULTS / FINDINGS

Results are reported in the sequence recommended for PLS-SEM studies: sample characteristics, measurement model assessment, discriminant validity, structural model results, and explanatory/predictive power (Hair et al., 2019).

Sample Characteristics

The final sample consisted of 434 respondents: 222 from Puerto Rico (51.2%) and 212 from the Dominican Republic (48.8%). Table 2 summarizes the demographic composition of the pooled sample. Gender and age totals are slightly lower than 434 due to item non-response (8 respondents did not report age).

Table 2. Sample Characteristics (N = 434)

Characteristic	Category	n	%
Country	Puerto Rico	222	51.2
Country	Dominican Republic	212	48.8
Country	Total	434	100.0
Gender	Male	175	40.3
Gender	Female	259	59.7
Gender	Total	434	100.0
Age	21–30 years	93	21.8
Age	31–45 years	132	31.0
Age	46–60 years	101	23.7
Age	61–65 years	100	23.5
Age	Total	426	100.0

Measurement Model Assessment

The measurement model was evaluated to assess the reliability and validity of the constructs, following established PLS-SEM guidelines (Hair et al., 2019; Henseler et al., 2009). The Perceived Informedness construct exhibited outer loadings above 0.90, indicating strong indicator reliability and internal consistency (Hair et al., 2019; Sarstedt et al., 2020). The NFM construct showed loadings ranging from approximately 0.63 to 0.79, within the range considered acceptable in social science research (Gil de Zúñiga et al., 2017; Hair et al., 2019).

The Social Media Exposure construct (SOCNEWS and VERIFY indicators) exhibited standardized outer loadings above 0.89 for both items, supporting strong indicator reliability despite the conceptual heterogeneity discussed in the Limitations section. The AI Perception construct initially included a larger set of indicators; items exhibiting low loadings were removed to improve construct reliability and validity, and the retained indicators exhibited loadings between approximately 0.54 and 0.79 (Ågerfalk, 2020; Hair et al., 2019). The lowest retained loading (0.54)

falls slightly below the 0.60 threshold conventionally cited for exploratory research; it was retained to preserve the conceptual breadth of the construct across its five facets, and this decision is acknowledged as a measurement limitation in the Limitations section. This item-purification step reflects the exploratory status of the AI Perception construct, which does not yet correspond to a widely validated scale in the information systems literature.

Indicator-level reliability was supported by the outer loading ranges reported above. Table 3 reports the outer loading range for each construct; additional reliability and validity indices are addressed in the Limitations section.

Table 3. Measurement Model: Indicator Reliability (Outer Loadings)

Construct	Outer Loadings (range)
AI Perception	0.54–0.79
News-Finds-Me Perception	0.63–0.79
Social Media Exposure	0.89–0.96
Perceived Informedness	> 0.90

Note. See Limitations for additional reliability and validity indices not reported here.

Discriminant Validity

Discriminant validity, which assesses whether constructs are empirically distinct from one another, is conventionally evaluated using the Heterotrait-Monotrait ratio of correlations (HTMT), a test of particular relevance given the conceptual proximity between AI Perception and Social Media Exposure noted in the Methodology section. HTMT ratios are addressed in the Limitations section.

Structural Model Results and Hypothesis Testing

The structural model was estimated using standardized path coefficients (β) and evaluated for statistical significance through a bootstrapping procedure with 5,000 subsamples (Hair et al., 2019). Table 5 reports the path coefficients, t-statistics, p-values, and hypothesis decisions for each of the four hypothesized relationships.

The path from AI Perception to Perceived Informedness was positive and statistically significant (β = 0.41, t = 2.74, p = .006); H1 was supported. The path from AI Perception to NFM was not statistically significant (β = -0.004, t = 0.31, p = .754); H2 was not supported. Similarly, the path from NFM to Perceived Informedness did not reach statistical significance (β = 0.14, t = 1.67, p = .095); H3 was not supported. The path from Social Media Exposure to NFM was positive and statistically significant, though modest in magnitude (β = 0.24, t = 2.09, p = .037); H4 was supported.

Table 5. Structural Model Results

Path	Hypothesis	β	t	p	Decision
AI Perception → Perceived Informedness	H1	0.41	2.736	.006	Supported

Path	Hypothesis	β	t	p	Decision
AI Perception → NFM	H2	-0.004	0.313	.754	Not supported
NFM → Perceived Informedness	H3	0.14	1.669	.095	Not supported
Social Media Exposure → NFM	H4	0.24	2.085	.037	Supported

Note. Path coefficients estimated using PLS-SEM with bootstrapping (5,000 subsamples). The bootstrapped indirect effect of AI Perception on Perceived Informedness through NFM was small and non-significant (β = -0.009, t = 0.246, p = .805), confirming that mediation is not supported by the present data.

Explanatory and Predictive Power

The structural model explained 16.8% of the variance in Perceived Informedness ($R^2 = 0.168$) and 5.6% of the variance in NFM ($R^2 = 0.056$), indicating modest explanatory power consistent with an exploratory model that includes a limited set of predictors (Hair et al., 2019). The low R^2 for NFM in particular means that weak measurement precision remains a plausible alternative explanation for the non-significant paths involving NFM, alongside a genuine absence of effect within this sample; the two explanations cannot be fully disentangled with the present data. Effect size (f^2) and predictive relevance (Q^2) are addressed in the Limitations section.

Table 6. Explanatory Power (R^2)

Endogenous Construct	R^2
Perceived Informedness	0.168
NFM	0.056

Key Findings

Within this sample, AI Perception emerged as a significant and direct predictor of Perceived Informedness, while NFM did not emerge as a significant predictor of Perceived Informedness. This pattern indicates that, in this exploratory model, algorithmic mediation and passive social exposure operate as distinct pathways rather than as a single, mediated mechanism.

The non-significant path between AI Perception and NFM further suggests that algorithmic exposure, as measured in this study, was not associated with the passive information beliefs captured by NFM within this sample. The evidence indicates that these two constructs may capture largely independent dimensions of contemporary information behavior. Social Media Exposure remained a significant, though modest, predictor of NFM, consistent with the construct's original conceptualization in social-network contexts (Gil de Zúñiga et al., 2017).

Discussion of Findings

The significant direct effect of AI Perception on Perceived Informedness (β = 0.41, p = .006) is the most consequential finding of this study. It suggests that algorithmic systems shape individuals' subjective sense of being informed through a pathway independent of the passive

exposure mechanisms traditionally emphasized in communication research (Ågerfalk, 2020; Glikson & Woolley, 2020), consistent with the conceptualization of AI as a form of digital agency that actively mediates users' interaction with information (Seidel et al., 2013; Koohang et al., 2023).

The non-significant relationship between AI Perception and NFM is similarly noteworthy. Contrary to expectations derived from prior research, AI Perception did not predict NFM within this sample (Strauß et al., 2021; Seaver, 2017), suggesting that algorithmic exposure and the passive, network-based mechanisms underlying NFM may constitute largely separate pathways of information delivery (Bodó et al., 2019; Van Dijck, 2013).

The non-significant path from NFM to Perceived Informedness must be read alongside the limited variance explained in NFM ($R^2 = 0.056$); a model that accounts for so little of NFM's variance cannot, on its own, establish that NFM is theoretically irrelevant, only that it did not emerge as a significant predictor within this particular model and sample. While prior work suggests that individuals who rely on passive exposure may nonetheless feel informed (Gil de Zúñiga et al., 2017; Lazer et al., 2018), this relationship did not hold here, perhaps because algorithmically curated content, delivered in a tailored and timely manner, increasingly substitutes for socially mediated exposure as a source of Perceived Informedness, consistent with research on trust in AI systems (Glikson & Woolley, 2020; Shin, 2021; Venkatesh et al., 2012).

The significant, though modest, relationship between Social Media Exposure and NFM ($\beta = 0.24$, $p = .037$) is consistent with the continued relevance of social platforms in facilitating passive information consumption (Gil de Zúñiga et al., 2017; Tucker et al., 2018). Because this construct also includes a source-verification indicator, the path may partly reflect verification-conscious users rather than passive exposure alone, and the limited explanatory power of NFM overall suggests that social exposure no longer fully accounts for how individuals in this sample come to feel informed (Strauß et al., 2021; Lazer et al., 2018).

Taken together, these findings point to a broader shift in the structure of information environments, in which algorithmic systems shape user perceptions alongside, rather than solely through, network-based diffusion (Van Dijck, 2013; Castells, 2010). This shift carries implications for political attitudes, including populist orientations (Akkerman et al., 2014), and for civic participation more broadly (Lazer et al., 2018): if individuals derive their sense of being informed substantially from algorithmically curated content, questions arise about the diversity and accuracy of the information underlying public opinion (Tucker et al., 2018; Pariser, 2011).

IMPLICATIONS FOR RESEARCH AND PRACTICE

Theoretical Implications

These findings extend the information systems literature by operationalizing an AI-related construct within a PLS-SEM framework and by empirically testing the explanatory limits of NFM in AI-mediated environments (Gil de Zúñiga et al., 2017; Ågerfalk, 2020). The non-significant NFM paths suggest that models of information behavior may benefit from incorporating algorithmic mediation as a distinct theoretical construct rather than treating it as an extension of existing passive-exposure frameworks (Seidel et al., 2013; Koohang et al., 2023). Future research

should further examine constructs such as algorithmic trust, perceived transparency, and explainability alongside AI Perception and NFM (Shin, 2020; Shin, 2021).

Managerial Implications

For platform operators, policymakers, and media organizations, the findings underscore the growing role of algorithmic systems in shaping how individuals perceive their own level of information (Davenport & Ronanki, 2018; Ågerfalk, 2020). Because AI Perception was not associated with the passive, socially mediated beliefs captured by NFM, organizations should not assume that algorithmic curation simply reinforces existing social information channels; algorithmic transparency and explainability merit direct attention in their own right (Shin, 2021; Glikson & Woolley, 2020). Digital and algorithmic literacy initiatives, which help users understand how AI systems curate content, may help mitigate risks of overconfidence and selective exposure associated with algorithmically curated information (Tucker et al., 2018; Bodó et al., 2019).

Limitations

This study has several limitations that should be considered when interpreting the findings (Hair et al., 2019; Sarstedt et al., 2020). First, the cross-sectional design limits the ability to establish causal relationships among constructs, as data were collected at a single point in time; longitudinal or experimental designs would strengthen causal inference regarding the effects of algorithmic exposure on Perceived Informedness (Hair et al., 2019; Henseler et al., 2009).

Second, the study relies on self-reported data, which may be subject to social desirability, recall bias, and common method variance (Podsakoff et al., 2003; Hair et al., 2019); the present study did not implement a statistical remedy such as Harman's single-factor test or a marker-variable approach, and future replications should incorporate such procedures. Because Perceived Informedness itself reflects a subjective assessment rather than objective knowledge, future research should also incorporate objective knowledge measures to examine the correspondence between perceived and actual informedness (Lazer et al., 2018; Akkerman et al., 2014).

Third, the measurement model reported here is provisional in several respects. The item-purification procedure used to refine the AI Perception construct improved statistical reliability but reduced conceptual breadth, and the lowest retained loading (0.54) falls slightly below the conventional 0.60 exploratory threshold (Hair et al., 2019). The Social Media Exposure construct further combines two indicators with strong loadings but non-equivalent conceptual content, reliance on social media as a news channel and a source-verification behavior, and should be treated as an empirically convenient rather than conceptually pure operationalization. Beyond indicator-level outer loadings and the explained variance (R^2) reported for the structural model, additional statistics conventionally reported in PLS-SEM studies (Cronbach's alpha, Composite Reliability, Average Variance Extracted, HTMT ratios, f^2, and Q^2) were not extracted from the SmartPLS output prior to manuscript finalization and are not reported here; these statistics should be computed and reviewed before this measurement model is used in confirmatory research.

Fourth, although RDD provided broad telephone coverage, individuals without telephone access or who declined participation were not represented, introducing coverage and nonresponse bias.

The sample is also limited to Puerto Rico and the Dominican Republic, two culturally distinct but geographically proximate Caribbean contexts, constraining generalizability to other regions (Tucker et al., 2018; Castells, 2010). Because the two national samples were pooled for this exploratory analysis rather than compared directly, country-specific patterns could not be examined; future studies should consider multi-group analyses (e.g., measurement invariance testing and PLS-MGA) to test for path differences across countries (Van Dijck, 2013; Koohang et al., 2023).

CONCLUSIONS

This study contributes to the information systems literature by empirically testing the relationships among AI Perception, NFM, and Perceived Informedness within a pooled Caribbean sample using PLS-SEM, offering an early empirical account of how algorithmic mediation relates to an established passive-exposure framework.

Theoretically, the study contributes a reflective operationalization of AI Perception, specified across five interrelated facets of algorithmic-mediation awareness, and provides an initial test of how this construct relates to NFM. The results suggest that information behavior models built primarily around social and network-based mechanisms may need to incorporate algorithmic mediation as a distinct explanatory pathway (Ågerfalk, 2020; Gil de Zúñiga et al., 2017; Seidel et al., 2013).

Practically, the findings underscore the importance of transparency, explainability, and algorithmic literacy initiatives as organizations and policymakers navigate AI's growing role in information environments (Shin, 2021; Davenport & Ronanki, 2018). Because users' sense of being informed was linked to AI Perception rather than to passive social exposure, efforts to support informed digital citizenship may need to address algorithmic awareness specifically, rather than relying solely on traditional media or social-media literacy programs (Tucker et al., 2018).

Future research should extend this exploratory model by validating the AI Perception construct across additional cultural contexts, testing the model using longitudinal or experimental designs to support causal claims, incorporating objective knowledge measures alongside Perceived Informedness, conducting formal multi-group comparisons between Puerto Rico and the Dominican Republic or other national contexts, and incorporating additional constructs such as algorithmic trust, media literacy, or political engagement alongside AI Perception and Perceived Informedness (Hair et al., 2019; Van Dijck, 2013; Shin, 2021; Glikson & Woolley, 2020; Ågerfalk, 2020; Seaver, 2017).

Acknowledgements

The author(s) acknowledge the use of AI-assisted tools during the preparation and language editing of this manuscript, consistent with the IIS AI Use Policy (https://iacis.org/iis/AI_Policy.php). All research design, data collection, analysis, and interpretation were conducted by, and are the sole responsibility of, the author(s).

REFERENCES

Akkerman, A., Mudde, C., & Zaslove, A. (2014). How populist are the people? Measuring populist attitudes in voters. Comparative Political Studies, 47(9), 1324–1353.

Ågerfalk, P. J. (2020). Artificial intelligence as digital agency. European Journal of Information Systems, 29(1), 1–8.

Bodó, B., Helberger, N., Eskens, S., & Möller, J. (2019). Interested in diversity: The role of user attitudes, algorithmic feedback loops, and policy in news personalization. Digital Journalism, 7(2), 206–229.

Castells, M. (2010). The rise of the network society (2nd ed.). Wiley-Blackwell.

Davenport, T. H., & Ronanki, R. (2018). Artificial intelligence for the real world. Harvard Business Review, 96(1), 108–116.

Gil de Zúñiga, H., Weeks, B., & Ardèvol-Abreu, A. (2017). Effects of the news-finds-me perception in communication. Journal of Computer-Mediated Communication, 22(3), 105–123.

Glikson, E., & Woolley, A. W. (2020). Human trust in artificial intelligence. Academy of Management Review, 45(4), 627–660.

Hair, J. F., Hult, G. T. M., Ringle, C. M., & Sarstedt, M. (2019). A primer on partial least squares structural equation modeling (PLS-SEM) (2nd ed.). Sage.

Henseler, J., Ringle, C. M., & Sinkovics, R. R. (2009). The use of partial least squares path modeling in international marketing. Advances in International Marketing, 20, 277–319.

Koohang, A., Sargent, C. S., Zhang, J. Z., & Marotta, A. (2023). Big data analytics: From leadership to firm performance. Industrial Management & Data Systems, 123(12), 2976–2996.

Lazer, D. M. J., Baum, M. A., Benkler, Y., Berinsky, A. J., Greenhill, K. M., Menczer, F., Metzger, M. J., Nyhan, B., Pennycook, G., Rothschild, D., Schudson, M., Sloman, S. A., Sunstein, C. R., Thorson, E. A., Watts, D. J., & Zittrain, J. L. (2018). The science of fake news. Science, 359(6380), 1094–1096.

Pariser, E. (2011). The filter bubble: How the new personalized web is changing what we read and how we think. Penguin.

Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review. Journal of Applied Psychology, 88(5), 879–903.

Sarstedt, M., Ringle, C. M., & Hair, J. F. (2020). Partial least squares structural equation modeling. In Handbook of market research (pp. 1–40). Springer.

Seaver, N. (2017). Algorithms as culture: Some tactics for the ethnography of algorithmic systems. Big Data & Society, 4(2), 1–12.

Seidel, S., Recker, J., & von Brocke, J. (2013). Sensemaking and sustainable practicing: Functional affordances of information systems in green transformations. MIS Quarterly, 37(4), 1275–1300.

Shin, D. (2020). User perceptions of algorithmic decisions in the personalized AI system: Perceptual evaluation of fairness, accountability, transparency, and explainability. Journal of Broadcasting & Electronic Media, 64(4), 541–565.

Shin, D. (2021). The effects of explainability and causability on perception, trust, and acceptance: Implications for explainable AI. International Journal of Human-Computer Studies, 146, 102551.

Strauß, N., Huber, B., & Gil de Zúñiga, H. (2021). Structural influences on the News Finds Me perception: Why people believe they don't have to actively seek news anymore. Social Media + Society, 7(2), 20563051211024966.

Tucker, J. A., Guess, A., Barberá, P., Vaccari, C., Siegel, A., Sanovich, S., Stukal, D., & Nyhan, B. (2018). Social media, political polarization, and political disinformation: A review of the scientific literature. Political Polarization and Political Disinformation.

Van Dijck, J. (2013). The culture of connectivity: A critical history of social media. Oxford University Press.

Venkatesh, V., Thong, J. Y. L., & Xu, X. (2012). Consumer acceptance and use of information technology: Extending the unified theory of acceptance and use of technology. MIS Quarterly, 36(1), 157–178.

ARTIFICIAL INTELLIGENCE AS A CATALYST FOR BUSINESS INNOVATION

Hesham Allam, Ball State University, hesham.allam@bsu.edu
Edward J. Lazaros, Ball State University, ejlazaros@bsu.edu
Allen D. Truell, Ball State University, atruell@bsu.edu
Fred Kitchens, Ball State University, fkitchens@bsu.edu

ABSTRACT

Artificial intelligence (AI) as a driver of innovation in modern business environments is well recognized. This is because AI can boost productivity and innovation in sectors by improving firms' data analytics and decision-making capabilities. However, AI technology deployment also comes with its own set of challenges, including algorithmic bias, lack of transparency, disruption to the workforce, and data governance issues. This paper critically assesses AI as an engine of innovation and as a strategic risk. The paper, based on an integrative literature review of recent academic work on the use of AI in business settings, shows that the adoption of AI is a necessary driver of sustainable innovation. Nevertheless, such an innovation strategy needs to be coupled with socially responsible AI practices that consider the ethical implications of applying AI technologies. The findings show that organizations need to adopt a socio-technical perspective on AI, which recognizes the technology's dual nature as both an enabler of innovation and a risk to their strategy.

Keywords: artificial intelligence, business innovation, responsible AI, digital transformation, workforce change

INTRODUCTION

Artificial intelligence has been growing in different companies through large scale data generation and improvements in computing (Brynjolfsson & McAfee, 2017; Makridakis, 2017). AI enables organizations to go beyond automation of processing to sound decision-making, prediction and foresight (Davenport & Ronanki, 2018). Currently, organizations leverage AI to identify market trends, allocate resources efficiently, and create innovations (Ardito et al., 2025). Therefore, the introduction of AI becomes an important element determining the development of continuous innovation in technology-led companies.

From a strategic management perspective, the adoption of artificial intelligence is not merely a technological choice. It is a transformational event associated with the development of innovative capacity and organizational flexibility. According to the resource-based view, organizations can gain a competitive advantage when they possess valuable, rare, inimitable, and well-organized resources or capabilities (Barney, 1991; Ferreira & Ferreira, 2025). Recent AI research suggests that organizations gain a competitive advantage not just from AI technologies themselves, but from the combination of proprietary data, organizational knowledge, and AI-enabled capabilities embedded in business processes (Wade et al., 2024).In the digital arena, this could mean proprietary databases, sophisticated analytical skills, knowledge of AI technology, and state-of-the-art technological platform infrastructure (Salerno & Maçada, 2025).

On the other hand, the use of artificial intelligence poses a challenge to organizational operations. There are numerous risks associated with the use of artificial intelligence, including bias, lack of transparency in the decision-making process, job loss, and information security risks, that need to be considered (Dwivedi et al., 2021; Floridi et al., 2018). If these risks are not considered, there will be negative impacts on innovation and potential legal ramifications.

This study aims to analyze the influence of AI on innovation and provide recommendations for the effective application of this technology to facilitate innovation while minimizing associated risks. The theoretical background of this study will include resource-based and dynamic capabilities theories. Based on these approaches, a conceptual framework will be proposed. The paper will include a literature review, methodological considerations, results, and conclusions.

LITERATURE REVIEW

The current literature review explores the role of artificial intelligence (AI) as an innovation driver, as well as the associated ethics and organizational implications. The primary focus of the review is on three related topics: AI as an innovation driver, organizational and ethical issues, and the governance of AI.

AI as a Driver of Business Innovation

Artificial intelligence technologies enable companies to improve the efficiency of business processes and drive innovation through predictive analytics and automation. Previous empirical studies have shown that companies that use AI technologies can achieve significantly higher efficiency, thanks to optimized resource allocation and improved processes (Makridakis, 2017). For example, in a manufacturing environment, artificial intelligence-powered predictive maintenance can improve equipment availability while reducing the duration of unavailability (Davenport & Ronanki, 2018). On the other hand, recommendation systems can generate higher revenues through better customer experience in retail companies (Brynjolfsson & McAfee, 2017).

Apart from improving operations, AI significantly enhances the quality of decision-making by analyzing relevant data and simulating different scenarios. Present-day studies demonstrate that the employment of AI generators results in better performance and higher quality of outcomes, with the impact being more noticeable among employees lacking experience, and hence, can be considered as a method of increasing efficiency and expanding skills (Brynjolfsson et al., 2023). The managers utilize the results generated by AI in forecasting, evaluating risks and setting prices (Davenport & Ronanki, 2018).

Decision making, however, relies on investments in human capital and learning efforts. Raisch and Krakowski (2021) use the term “automation-augmentation paradox” to describe the above scenario, which implies that the automation of tasks and the improvement of decision-making skills by human managers through AI occur at once. Companies that can handle the dilemma enjoy the advantage of integrating information generated by AI and professional decision-making.

Moreover, artificial intelligence is one of the main enablers of data-driven innovations that help in creating products and services. Data-driven innovations allow organizations to spot shifts in customer preferences and reduce time for the creation of new products (Ardito et al., 2025). It is

possible to regard AI as a general-purpose technology since it causes productivity effects on the whole economy similarly to other technological progress like electricity and the Internet (Brynjolfsson & McAfee, 2017).

Ethical and Organizational Issues

While being extremely beneficial to businesses, artificial intelligence creates numerous ethical challenges that should be carefully managed. First, there is a possibility of algorithmic bias when AI is used to make decisions, including hiring and lending people money, based on machine learning models trained on biased datasets (Floridi et al., 2018; Singhal et al., 2024; Cheong, 2024). The described issue may lead to discrimination, litigation, and negative media coverage.

Additionally, the lack of transparency hinders organizations' capacity for self-accountability. The decisions that AI makes are difficult to explain due to complex machine learning algorithms, particularly in areas such as healthcare, policing, and finance (Chaudhary, 2024; Rudin, 2019). The issue of explainability is very significant in fields such as healthcare, law enforcement, and banking, where accountability is important (Akhtar et al., 2024; Batool et al., 2025). This problem makes it difficult for organizations to manage their AI technologies and audit their decisions under the new AI governance regulations (Cheong, 2024; Ismail & Ahmad, 2025).

Finally, another major ethical issue that arises from the large amounts of private data AI technologies rely on is data privacy. The reason is that businesses will lose out since they cannot keep their data private (Dwivedi et al., 2021). They may face various consequences, ranging from legal action to fines.

Fourth, there is a significant organizational challenge in transforming the workforce. Research shows that AI automation causes job losses in manual and cognitive jobs, but also creates a demand for more analytical, innovative, and soft skills (Bessen, 2018). According to the World Economic Forum, AI will simultaneously displace and create jobs and will increase the demand for analytical, technological, and human-centered skills (World Economic Forum, 2025). Recent research from the International Labor Organization shows that AI will affect many job tasks, but more by changing them than by fully replacing workers (International Labor Organization, 2025). Acemoglu and Restrepo (2020) find two effects of technology on the labor market: it tends to automate some tasks and complement other human work when applied correctly.

Responsible AI and Governance Frameworks

In order to avoid such problems, it is important that experts draw attention to the importance of AI system management. Effective management involves the application of ethical principles like fairness, transparency, accountability, privacy, and risk management, which have been considered in risk management of AI systems frameworks, such as the NIST AI Risk Management Framework (NIST, 2024), and the recent AI governance literature (Batool et al., 2025). Companies develop such governance approaches to address AI-related risks and ethical concerns.

Artificial intelligence in the socio-technical framework highlights the significance of looking at both the technological aspect as well as organizational management (Jarrahi, 2018). This implies

that the proper use of AI will depend on various aspects including organizational learning, skills development among employees, stakeholder participation, and performance evaluation.

Thematic Synthesis of Findings

As a result, the following seven themes were identified through thematic analysis of the literature on the adoption of artificial intelligence within organizations.

Firstly, AI is seen as a means to increase organizational efficiency. Several studies suggest the advantages of implementing AI through automation, prediction, and optimization; nevertheless, these advantages could be reduced in effectiveness if the organizations rely heavily on the automation tool without putting necessary controls in place.

Second, the use of AI leads to better decision-making. This is through the use of predictive analytics, forecasts, and scenario analysis. However, data quality issues, explainability, and lack of transparency can affect the quality of the decisions made.

Third, AI contributes a lot to innovation of products and services in the organization. The organization will be able to innovate by improving customer experience, experimenting, and digital products and services. Furthermore, it is pointed out that collecting too much customer data may affect the privacy of the organization and its reputation.

Fourthly, the use of AI increases the organization's competitive advantage by developing learning and adaptability capabilities. It should be noted that different firms reap different benefits, as companies with better digital resources, more sophisticated analytics, and advanced organizational capabilities get the most out of technology.

Finally, the issue of transformation of the workforce is a huge problem with the implementation of AI within the organization. While AI helps in simplifying tasks for employees and even gives them an opportunity to analyze and think creatively, there are many other problems that should be taken into account by companies. Successful implementation usually requires training and development of the workers.

Additionally, the topics of trust, ethics, and organizational legitimacy remain very important for organizations. In this regard, the authors indicate that the absence of transparency, weak accountability systems, and bias can undermine stakeholders' trust and pose risks to sanctions and organizational reputation. Hence, the proper use of AI is critical.

Lastly, the sustainability of innovation benefits requires integrating AI into organizational systems rather than merely applying technologies. The literature indicates that successful long-term value creation depends on leadership, cross-functional cooperation, organizational learning, and the adaptation of governance measures to technological change.

All these themes demonstrate that artificial intelligence must be understood as a socio-technical capability rather than a mere technical resource.

METHODOLOGY

In this study, an integrative literature review serves as the principal research design to assess the implications of AI adoption for business innovation and the corresponding strategic risks. The rationale for choosing such a methodology lies in its effectiveness in summarizing diverse bodies of literature and generating a holistic view of complex, multilevel phenomena (Whittemore & Knafl, 2005). Given the fragmented nature of the existing literature regarding the interplay between AI innovation and strategic risks, this type of analysis seems particularly appropriate.

Search Strategy

For this literature review, relevant literature was identified via searches of the main bibliographic databases, including Scopus, Web of Science, and Google Scholar. The search used Boolean operators to combine keywords within the following categories:

- **AI and Innovation:** ("artificial intelligence" or "AI" OR "machine learning") and ("business innovation" OR "digital transformation" OR "innovation outcomes")
- **Risks and Ethics:** ("algorithmic bias" or "transparency " or "data privacy" or "workforce displacement" OR "ethics") an ("AI" OR "artificial intelligence")
- **Governance:** ("AI governance" or "responsible AI" or "socio-technical" or/and "governance framework") and ("business" OR "organization")

Only journal articles and conference papers, as well as official documents released by reputable international organizations, published between 2015 and 2025 have been considered. A total of 487 potentially relevant sources has been identified through this procedure.

Selection and Screening

The following procedure for screening the collected sources was adopted.

First, titles and abstracts were independently evaluated by two reviewers for relevance to the topic under study. The exclusion criterion was to focus on issues other than AI in the business context. As a result of the filtering stage, 124 articles remained.

The whole text of the rest of the papers was further reviewed for the following criteria: relevance to the topic, consideration of innovation, risk, ethical or governance issues, and making an empirical or theoretical contribution, and finally, published in English. Through this process, 40 papers were found to be relevant to the research.

Data Extraction and Thematic Synthesis

An analysis of these sources was conducted using thematic synthesis, a methodology for examining themes that emerge in qualitative and/or quantitative studies (Thomas & Harden, 2008). This process entailed the following steps:

1. **Initial Coding:** Key concepts, theories, evidence, and recommendations found in each paper have been identified.
2. **Theme Development:** Emerging themes have been generated by grouping codes after comparative analysis and discussion.

3. **Synthesis and Integration**: Identified themes have been compared to determine points of agreement and disagreement. The results of this phase are shown at the end of the literature review, which combines innovation opportunities, strategic risks, business impacts, and suggested AI governance measures in seven main thematic categories.

Illustrative Industry Data

To demonstrate how the concepts discussed apply to specific industries, data on AI usage has been obtained from the Organization for Economic Co-operation and Development (OECD) ICT Access and Usage Database (OECD, 2026). This data allows highlighting differences between industries in AI adoption and will be used descriptively rather than analytically in the subsequent discussion.

Framework Development

The conceptual model, formulated based on findings from thematic synthesis and industry example analysis, is shown in Figure 2 below. The relationships between AI adoption, capabilities, innovations, strategic risks, and AI governance are shown in Figure 2.

ANALYSIS AND DISCUSSION

This part includes important insights regarding the impact of AI on business innovation and competitiveness. From theoretical perspectives and the integration of multiple studies, the following conclusions can be drawn.

Innovation Opportunities and Competitive Advantage

AI has transformed from an additional innovation tool into a core driver of innovations and competitive advantage. Apart from making business operations more efficient, AI has enabled advancements in forecasting, optimization, experimentation, and decision-making. Firms that adopt AI have an advantage over others, as they enjoy higher efficiency, quicker responses to external factors, and faster innovation (Makridakis, 2017; Davenport & Ronanki, 2018).

From the standpoint of resource-based theory, the value of resources, including data and analytical skills, increases with AI adoption. Moreover, the difficulty of replicating the value created by these resources increases if they are combined with managers' competencies and organizational learning processes (Barney, 1991). Consistent with the dynamic capability's perspective, AI enables companies to sense environmental changes, exploit opportunities for innovation, and adjust organizational activities to dynamic conditions (Teece et al., 2016).

According to industry-level data, adoption of AI accelerates across all sectors, although in some sectors it is faster than others. Specifically, industries with high knowledge intensity and high levels of digitalization experience rapid AI adoption compared to other industries. According to the International Labor Organization (2025), there are significant disparities across professions in the degree of their exposure to AI.

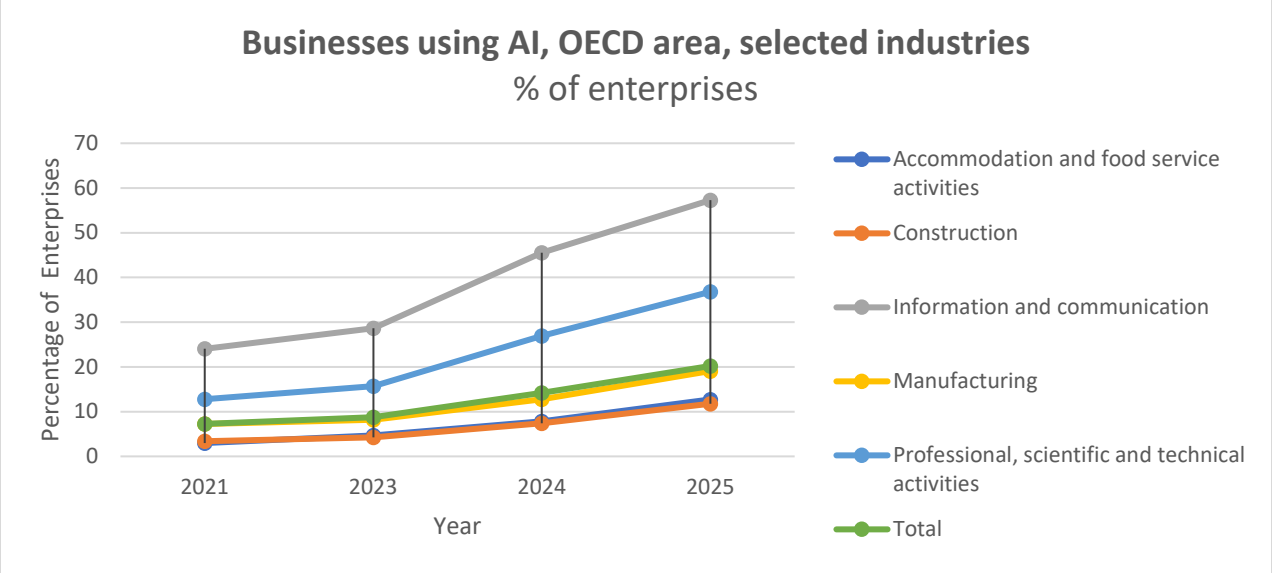


Figure 1. Adoption of Artificial Intelligence Across Selected Industries in the OECD Area 2021-2025, adapted from: OECD ICT Access and Usage by Businesses Database (2026), Businesses Using AI Technologies (% of enterprises), selected industries.

As shown in Figure 1, there is a sharp rise in the implementation of AI technologies in OECD countries, with particular emphasis on ICT, professional services, and finance. On the other hand, adoption levels remain quite low in construction, accommodation, and manufacturing. This implies that the effectiveness of innovation is not only dependent on the technology itself but also on the maturity of organizations and employees' competence within the industry. In general, the findings clearly support the notion that AI technology is a strategic resource that needs to be mobilized properly.

Strategic Challenges Associated with Ethical Risks

On the one hand, the findings reported in Figure 1 indicate a significant innovation potential. On the other hand, research suggests that adopting AI poses significant strategic challenges for organizations, with major implications for innovation performance.

First, algorithmic bias represents an important strategic issue. Biased machine learning algorithms lead to increased discrimination in employment opportunities, consumer finance, CRM programs, and many other areas where high-stakes decisions are made (Floridi et al., 2018). In addition to being unethical, biased decision-making exposes organizations to legal consequences and to the loss of trust from various stakeholder groups. Therefore, AI governance should address such problems at the strategic level.

Second, the problem of lack of transparency poses a significant strategic challenge. Since many AI models remain opaque, it becomes difficult for organizations to make proper decisions and exercise control over the process (Rudin, 2019). A transparency deficit can be particularly detrimental in regulated industries, where explanations are legally required.

Third, the risk related to personal and proprietary data raises additional strategic issues. In today's landscape of sophisticated machine learning techniques, reliance on big data creates additional vulnerabilities that affect organizations through cybersecurity threats and privacy breaches (Dwivedi et al., 2021). As a result, AI governance should address data-related risks to ensure privacy and security while pursuing innovation goals.

Finally, the problem of workforce transformation adds yet another strategic aspect to AI adoption. On the one hand, innovations in this domain contribute to workforce empowerment. On the other hand, the implementation of AI solutions leads to job loss across all categories of tasks (Bessen, 2018). Successful transformation of work involves strategic investments into workforce development and change management (Acemoglu & Restrepo, 2020). According to the ILO (2025), jobs, rather than occupations, will undergo transformation in the future if companies adopt proactive approaches.

The discussion above implies that ethical issues should be incorporated into any core AI strategy to ensure that strategic benefits materialize in innovation performance. In other words, AI governance needs to account for the potential risks associated with the technology to ensure innovation and competitiveness. This conclusion provides a rationale for further elaboration of the conceptual framework described below.

Conceptual Framework: AI Innovation and Responsible Governance

Based on the analysis and existing theory, a conceptual model is proposed to explain the linkages between AI adoption, capability development, innovation performance, and sustainable competitive advantage (see Figure 2). First, following resource-based and dynamic capabilities theories, the model assumes that innovation results depend on organizational capabilities and governance.

In particular, the model proposes four key propositions:

Proposition 1 claims that the adoption of AI technologies—such as machine learning, natural language processing, predictive analytics, and automation—helps build and enhance three organizational capabilities: (a) the capability of processing big data and making sense of information; (b) the capability to automate routine operations; and (c) the ability to adapt and respond to environmental changes.

According to Proposition 2, developing such organizational capabilities allows firms to generate three main types of innovation performance: (a) improvements in efficiency achieved due to process optimization and better use of resources; (b) improved decision-making based on data insights and predictive modeling; and (c) market differentiation resulting from innovations in product or service offerings, personalization, time-to-market, and customer experience enhancement.

Finally, Proposition 3 specifies that specific governance mechanisms play a vital role in translating AI capabilities into innovation outcomes. Two essential components are necessary for governing AI in the way that ensures its benefits and prevents strategic failures:

- Firstly, firms must ensure proper management of their AI practices through practices such as conducting audits to remove any bias, enhancing transparency, instituting accountability standards, and ensuring data security, among other methods of preventing any negative impact from the practice of AI.
- Secondly, AI governance must ensure that organizations adapt due to changes in their workforces.

Lastly, the **Fourth Proposition** suggests that responsible AI governance can help an organization achieve excellence in its innovations while maintaining legitimacy, stakeholder trust, and compliance. It will result in a competitive advantage.

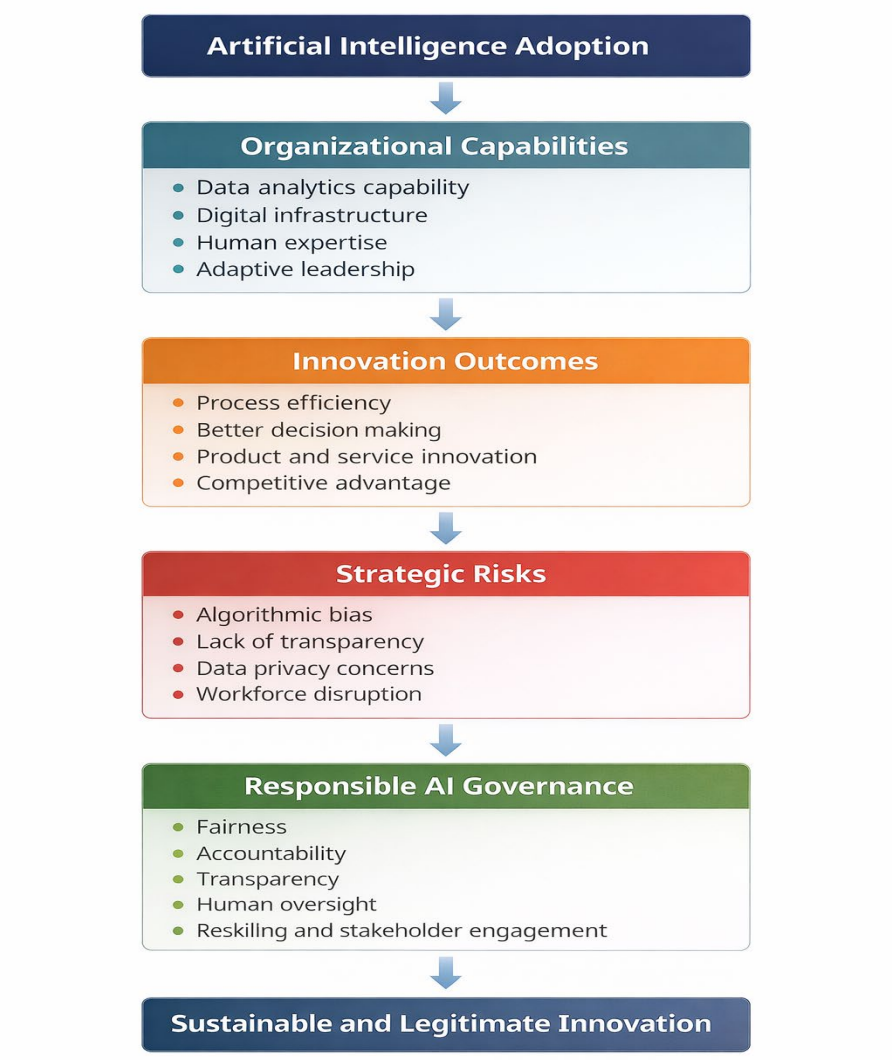


Figure 2. Conceptual Framework of AI-Driven Business Innovation and Responsible Governance

The theoretical value of the framework in relation to existing literature is associated with the following aspects: (a) the combination of the resource-based and dynamic capability theories with socio-technical governance theory; (b) the formulation of empirically testable hypotheses; (c) the reconstruction of the concept of responsible AI through the lens of strategic enabler; and (d) the focus on the contingency of the effects of AI innovations. This conceptual integration provides a more nuanced understanding of AI's dual role in contemporary organizations.

DISCUSSION SUMMARY

As discussed above, the analysis reveals AI as a dual facilitator of strategy – acting both as a catalyst and a potential strategic risk factor. On one hand, AI enables efficiency gains, better decisions, and differentiation through innovations in goods and services. On the other hand, ethical risks that are not controlled (algorithmic bias, transparency problems, privacy concerns, and disruption due to restructuring of the workforce) may hinder innovation success.

The proposed conceptual framework addresses this duality by examining specific ways in which governance shapes the relationship between AI adoption and innovation. Specifically, the framework argues that organizational success, in terms of competitive advantage and sustainable innovation, requires not only the aggressive development of capabilities but also active efforts toward responsible governance and adaptation.

Thus, the proposed framework helps examine the question of responsible AI innovation in a different way – shifting from viewing it as a barrier to innovation to seeing it as an indispensable element of sustainable competitive advantage.

Findings

Based on the analysis, key lessons and insights are the following:

Firstly, the research highlights the importance of adopting a socio-technical perspective on AI innovation – one that addresses the interrelations among technological systems, human capabilities, and organizational context. Organizations that successfully integrate these elements achieve sustainable innovation outcomes and uphold their ethical standards and legal responsibilities.

Secondly, the lack of attention to socio-technical integration in the development of new technologies is likely to lead to strategic failures and undermine the potential benefits of innovation.

CONCLUSION

As the title implies, the main objective of this study was to analyze and evaluate the dual nature of artificial intelligence as a facilitator and generator of strategic risk in relation to innovation. An integrative literature review based on resource-based theory and dynamic capabilities showed that AI is an important disruptive capability that enables improvements in business efficiency and market differentiation through innovation. However, achieving such results depends on the governance of associated risks.

The main contribution of this study is the discussion of a framework that addresses how to combine AI capabilities with governance practices to enable successful innovation and competitive advantage. Specifically, the proposed framework incorporates resource-based theory, dynamic capabilities, and socio-technical governance perspectives to present several propositions about the process of innovation through the responsible adoption of AI capabilities.

Governance is considered extremely important because it determines whether AI will bring competitive advantages or lead to failure within the organization. This point of view departs from overly simplistic approaches to technological innovation and offers a more comprehensive view of the issue.

The end of the literature review section illustrates the synthesized information about innovation opportunities, strategic risks, business consequences, and governance measures in seven thematic areas. Figure 1 shows the use of AI in different sectors. When analyzed together, this data enables managers to understand what steps should be undertaken to effectively utilize innovation potential of AI and prevent any strategic risks.

Managerial Implications

For today's managers, there is a need to transition from testing AI capabilities to systematic innovation grounded in their effective use. This requires the adoption of a two-fold strategy consisting of: (a) vigorous pursuit of AI development and adoption; and (b) establishing responsible practices of its governance. Some specific recommendations would include:

- establishment of multi-stakeholder governance committees;
- regular fairness audits;
- implementation of bias mitigation protocol;
- workforce reskilling and job redesign;

In other words, it is crucial to foster an organizational culture in which responsible AI practices are seen as a means of gaining a competitive advantage.

Policy Implications

Based on the above, some policy implications are as follows:

- Development of responsible AI standards;
- Promotion of workforce training programs;
- Development of sector-specific strategies for AI adoption;
- International collaboration on issues related to AI innovation.

Limitations and Future Research

There are some limitations associated with this study, which could serve as suggestions for future research. First, the integrative review used in the paper did not include any primary data. Thus, there is a need to conduct empirical research in order to test propositions presented here. Secondly, there was no consideration of AI innovation at the sectoral, national, and international levels. Also,

the industry data presented are merely descriptive; thus, there is a need to conduct statistical analyses of the causes of differences in adoption levels and outcomes.

Future research priorities:

- longitudinal studies evaluating the long-term impact of AI innovation;
- comparative analysis of governance practices in responsible AI innovation;
- investigation of organizational leadership in fostering AI innovation.

Acknowledgment

Grammarly AI was used only for language clarity and proofreading. All academic content, analysis, and conclusions were developed independently by the author(s).

REFERENCES

Akhtar, M. A. K., Kumar, M., & Nayyar, A. (2024**).** *Transparency and accountability in explainable AI: Best practices.* In *Towards Ethical and Socially Responsible Explainable AI*. Springer.

Acemoglu, D., & Restrepo, P. (2020). Robots and jobs: Evidence from US labor markets. *Journal of Political Economy, 128*(6), 2188–2244. https://doi.org/10.1086/708716

Ardito, L., Filieri, R., Raguseo, E., & Vitari, C. (2025). Artificial intelligence adoption and revenue growth in SMEs. *Internet Research*. Advance online publication. https://doi.org/10.1108/INTR-01-2024-0082

Barney, J. (1991). Firm resources and sustained competitive advantage. *Journal of Management, 17*(1), 99–120. https://doi.org/10.1177/014920639101700108

Batool, A., Zowghi, D., & Bano, M. (2025)**.** *AI governance: a systematic literature review.* AI and Ethics. https://link.springer.com/article/10.1007/s43681-024-00653-w

Bessen, J. E. (2018). Artificial intelligence and jobs: The role of demand. In A. Agrawal, J. Gans, & A. Goldfarb (Eds.), *The economics of artificial intelligence: An agenda* (Chap. 10). University of Chicago Press. https://www.nber.org/books-and-chapters/economics-artificial-intelligence-agenda/artificial-intelligence-and-jobs-role-demand

Brynjolfsson, E., Li, D., & Raymond, L. R. (2023). *Generative AI at work* (NBER Working Paper No. 31161). National Bureau of Economic Research. https://doi.org/10.3386/w31161

Brynjolfsson, E., & McAfee, A. (2017). *Machine, platform, crowd: Harnessing our digital future*. W. W. Norton.

Chaudhary, G. (2024). *Unveiling the black box: Bringing algorithmic transparency to AI.* Masaryk University Journal of Law and Technology. https://journals.muni.cz/mujlt/article/view/36881

Cheong, B. C. (2024). *Transparency and accountability in AI systems: Safeguarding wellbeing in the age of algorithmic decision-making.* Frontiers in Human Dynamics. https://www.frontiersin.org/articles/10.3389/fhumd.2024.1421273

Davenport, T. H., & Ronanki, R. (2018). Artificial intelligence for the real world. *Harvard Business Review, 96*(1), 108–116.

Dwivedi, Y. K., Hughes, L., Ismagilova, E., Aarts, G., Coombs, C., Crick, T., Duan, Y., Dwivedi, R., Edwards, J., Eirug, A., Galanos, V., Ilavarasan, P. V., Janssen, M., Jones, P., Kar, A. K., Kizgin, H., Kronemann, B., Lal, B., Lucini, B., ... Williams, M. D. (2021). Artificial Intelligence (AI): Multidisciplinary perspectives on emerging challenges, opportunities,

and agenda for research, practice, and policy. *International Journal of Information Management, 57*, 101994. https://doi.org/10.1016/j.ijinfomgt.2019.08.002

Ferreira, N. C., & Ferreira, J. J. (2025). The field of resource-based view research: mapping past, present, and future trends. *Management Decision*, *63*(4), 1124-1153.

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, P., & Vayena, E. (2018). AI4People—An ethical framework for a good AI society. *Minds and Machines, 28*(4), 689–707. https://doi.org/10.1007/s11023-018-9482-5

International Labour Organization. (2025). *Generative AI and jobs: A refined global index of occupational exposure*. https://www.ilo.org/publications/generative-ai-and-jobs-refined-global-index-occupational-exposure

Ismail, O., & Ahmad, N. (2025). *Ethical and Governance Frameworks for Artificial Intelligence: A Systematic Literature Review.* International Journal of Interactive Mobile Technologies.

Jarrahi, M. H. (2018). Artificial intelligence and the future of work: Human-AI symbiosis in organizational decision making. *Business Horizons, 61*(4), 577–586. https://doi.org/10.1016/j.bushor.2018.03.007

Makridakis, S. (2017). The forthcoming artificial intelligence (AI) revolution: Its impact on society and firms. *Futures, 90*, 46–60. https://doi.org/10.1016/j.futures.2017.03.006

National Institute of Standards and Technology. (2024). *Artificial intelligence risk management framework (AI RMF 1.0)*. U.S. Department of Commerce. https://www.nist.gov/itl/ai-risk-management-framework

Organization for Economic Co-operation and Development (OECD). (2026). *ICT access and usage by businesses database*. https://www.oecd.org/en/data/datasets/ict-access-and-usage-by-businesses-database.html

Raisch, S., & Krakowski, S. (2021). Artificial intelligence and management: The automation–augmentation paradox. *Academy of Management Review, 46*(1), 192–210. https://doi.org/10.5465/amr.2018.0072

Rudin, C. (2019). Stop explaining black-box machine learning models for high-stakes decisions; use interpretable models instead. *Nature Machine Intelligence, 1*(5), 206–215. https://doi.org/10.1038/s42256-019-0048-x

Singhal, A., Neveditsin, N., Tanveer, H., et al. (2024). *Toward fairness, accountability, transparency, and ethics in AI for social media and health care: Scoping review.* JMIR Medical Informatics. https://medinform.jmir.org/2024/1/e50048/

Teece, D. J., Peteraf, M., & Leih, S. (2016). Dynamic capabilities and organizational agility: Risk, uncertainty, and strategy in the innovation economy. *California Management Review, 58*(4), 13–35. https://doi.org/10.1525/cmr.2016.58.4.13

Thomas, J., & Harden, A. (2008). Methods for the thematic synthesis of qualitative research in systematic reviews. *BMC Medical Research Methodology, 8*(1), 45. https://doi.org/10.1186/1471-2288-8-45

Whittemore, R., & Knafl, K. (2005). The integrative review: Updated methodology. *Journal of Advanced Nursing, 52*(5), 546–553. https://doi.org/10.1111/j.1365-2648.2005.03621.x

World Economic Forum. (2025). *The future of jobs report 2025*. World Economic Forum. https://www.weforum.org/reports/the-future-of-jobs-report-2025

CYBERSECURITY COMPLIANCE FOR OHIO COUNTIES: UNDERSTANDING OHIO'S 2025 CYBERSECURITY STANDARDS AND GOVERNANCE CHALLENGES

Nathan McVicar, Marymount University, n0m01886@marymount.edu

ABSTRACT

County governments increasingly rely on digital infrastructure to deliver public services, manage records, and support public safety operations. In Ohio, the implementation of House Bill 96 and related state guidance requires political subdivisions to adopt formal cybersecurity programs designed to protect information systems and respond to cybersecurity incidents. However, county governments often operate with limited cybersecurity staffing, fragmented information technology governance, and varying levels of centralized authority. This paper uses a qualitative conceptual policy analysis supported by structured document review to examine how Ohio's 2025 cybersecurity requirements apply to county governments. The analysis synthesizes Ohio statutory requirements, Auditor of State guidance, CyberOhio resources, Criminal Justice Information Services (CJIS) and Law Enforcement Automated Data System (LEADS) obligations, recognized cybersecurity frameworks, and public sector information technology governance literature. The paper contributes a source-mapped compliance-readiness framework that distinguishes mandatory legal requirements, state guidance, public safety compliance obligations, recognized cybersecurity best practices, and inferred implementation recommendations. The original contribution of this paper is its integration of Ohio's cybersecurity requirements, county-level governance fragmentation, public safety compliance obligations, and recognized cybersecurity frameworks into a practical implementation model for resource-constrained county governments. The framework is intended to help Ohio county officials evaluate cybersecurity readiness, clarify governance responsibilities, and prioritize implementation actions within resource-constrained local government environments.

Keywords: Ohio cybersecurity, local government cybersecurity, public sector cybersecurity, county IT governance, cybersecurity compliance

INTRODUCTION

Local governments have steadily expanded their reliance on digital infrastructure over the past two decades as governments adopted electronic records management systems, online public portals, and interconnected public safety networks. Programs, like the federal E-Government Act of 2002, accelerated this transition by encouraging state and local agencies to digitize services and modernize administrative workflows (E-Government Act of 2002, 2002). While these advancements improved efficiency and accessibility, they simultaneously expanded the cybersecurity attack surface for government institutions (Organisation for Economic Co-operation and Development, 2024). Cyber threats targeting public sector entities have increased substantially during the same period, including ransomware attacks affecting state and local governments (Organisation for Economic Co-operation and Development, 2024).

Local governments have become particularly attractive targets for cybercriminals because they store large volumes of sensitive data, often while operating with constrained budgets and limited cybersecurity staffing. Research indicates that many county governments operate with fewer than five dedicated cybersecurity personnel, which could limit their ability to implement advanced security controls or maintain continuous network monitoring (National Association of Counties, 2023). Workforce shortages, legacy systems, and decentralized governance structures could further complicate cybersecurity readiness at the county level (Villani, 2022).

Ohio counties operate within a unique statutory environment that shapes how cybersecurity responsibilities are distributed across government offices. Ohio Revised Code §307.842 authorizes the establishment of a county automatic data process board, which is typically chaired by the county auditor, to coordinate data processing and technology decisions across county departments. At the same time, the statute allows elected officials to maintain independent information technology infrastructure and staffing. This structure often results in multiple semi-autonomous IT environments operating within a single government (Ohio Rev. Code § 307.842, 2011).

The coexistence of centralized coordination and departmental autonomy produces what could be described as a hybrid governance structure. Some counties operate under highly coordinated IT departments, while others maintain fragmented technology environments with varying cybersecurity policies and operational practices. This variation creates significant differences in cybersecurity maturity across counties.

Recognizing the growing cyber risk to local governments, the State of Ohio implemented new cybersecurity requirements in 2025, through House Bill 96, and the associated CyberOhio standards. These requirements represent an important shift toward standardized cybersecurity practices across Ohio's political subdivisions.

This paper provides a concise overview of the new cybersecurity expectations affecting Ohio counties and outlines practical strategies that county governments could adopt to strengthen cybersecurity readiness, while maintaining the operational flexibility required in public administration. This paper addresses a gap in the public sector cybersecurity governance literature by examining how statewide cybersecurity mandates are translated into county-level implementation responsibilities within a decentralized local government environment. While existing cybersecurity frameworks provide broad technical guidance, less attention has been given to how counties with fragmented authority, limited staffing, and varying levels of IT centralization can operationalize state cybersecurity requirements. This paper contributes a source-mapped compliance-readiness framework for Ohio counties by distinguishing statutory requirements, state guidance, public safety information security obligations, recognized cybersecurity best practices, and implementation recommendations inferred from public sector information technology (IT) governance literature.

BACKGROUND AND LITERATURE REVIEW

Prior research on public-sector information technology governance shows that technology implementation is shaped by organizational authority, resource availability, procurement constraints, leadership coordination, and institutional capacity. These issues are especially

important in local government cybersecurity because counties must protect sensitive data and critical services while operating within fragmented governance structures and limited staffing environments. This section synthesizes literature related to county cybersecurity risk, decentralized information technology governance, vendor and legacy system risk, and resource constraints to establish the foundation for analyzing Ohio's 2025 cybersecurity requirements.

Cybersecurity Risk in County Governments

County governments face several structural challenges that influence their cybersecurity posture. Unlike federal or state agencies, counties often operate with limited financial and personnel resources dedicated to information security. Many counties rely on small IT departments responsible for maintaining networks, managing software systems, and responding to cybersecurity incidents simultaneously.

Research on local government IT implementation has long demonstrated that public institutions often adopt technological innovations more slowly than private sector organizations because of funding constraints, procurement procedures, and organizational complexity (Beaumaster, 2002). These constraints are especially evident in cybersecurity, where constant monitoring, rapid incident response, and continuous system updates are required to maintain security.

Fragmented governance structures could further amplify cybersecurity vulnerabilities. When multiple departments operate separate IT environments, cybersecurity practices may vary significantly between offices. This inconsistency could lead to uneven patch management, weak access controls, and uncoordinated incident response procedures. Studies of decentralized IT environments have demonstrated that such fragmentation frequently produces duplication of technology investments and inconsistent security standards (Chen et al., 2010).

County governments must also manage risks associated with legacy systems and vendor relationships. Older systems may not receive regular security updates, while outsourced technology services could introduce additional complexity if cybersecurity responsibilities are not clearly defined in contracts. Research suggests that contracting decisions in cybersecurity environments could be particularly challenging due to the specialized nature of security services (Nussbaum & Park, 2018).

These challenges illustrate why cybersecurity must be understood not solely as a technical issue, but also as a governance issue. Recent research on local governments similarly indicates that technological vulnerability and limited organizational knowledge of cyber incidents can affect public sector innovation and cybersecurity management, reinforcing the importance of leadership awareness and structured governance practices (Fusi et al., 2025). The organizational structure of IT operations can significantly influence how effectively cybersecurity policies are implemented across a government entity.

Governance Challenges in Ohio Counties

Ohio counties operate within a statutory governance framework that combines centralized coordination with departmental autonomy. The county data board, authorized under Ohio Revised Code §307.842, is intended to coordinate technology decisions across county offices. However,

the statute does not include enforcement authority, which would require elected officials to adopt the board's recommendations. As a result, technology governance often depends on voluntary cooperation among county departments.

Research on public sector IT governance has demonstrated that governance structures strongly influence the success of digital transformation initiatives. Studies have found that digital initiatives frequently fail when governance structures do not align with implementation demands, or when leadership does not share the same strategic direction (Ioannou et al., 2023).

Centralized IT governance could improve coordination by standardizing policies, consolidating resources, and enabling consistent cybersecurity practices across departments. In contrast, decentralized governance structures could create duplication of resources and inconsistent security controls (Chen et al., 2010).

At the same time, decentralization offers certain advantages in public administration. Department autonomy could allow elected officials to respond more quickly to operational needs and maintain control over specialized systems (Denford, 2016). Consequently, counties must balance these competing priorities while ensuring compliance with state cybersecurity requirements. Cybersecurity readiness in Ohio counties is influenced not only by technical security practices, but also by governance structures that shape how resources, authority, and decision-making responsibilities are distributed across local government institutions.

The governance challenges described above can also be understood through the broader lens of public sector information technology governance. Research suggests that organizational structures strongly influence how effectively technology initiatives and security policies are implemented within government institutions (Chen et al., 2010). In decentralized environments, decision-making authority, budget control, and operational responsibility for technology systems may be distributed across multiple departments or offices, which can complicate the implementation of standardized cybersecurity controls. Studies of cybersecurity governance in county governments similarly emphasize that fragmented authority and limited institutional coordination often constrain the ability of local governments to adopt consistent security practices (Villani, 2022). Consequently, the effectiveness of Ohio's new cybersecurity requirements will depend not only on the adoption of technical safeguards, but also on the governance structures that shape how counties coordinate cybersecurity responsibilities, allocate resources, and enforce security policies across departments.

Table 1 compares centralized, decentralized, and hybrid county IT governance models and summarizes how each structure could affect cybersecurity implementation. This comparison is important because Ohio counties do not operate under a single technology governance model. Some counties rely on centralized IT departments, while others allow elected offices or departments to maintain separate systems, staff, and operational practices. Hybrid governance is likely the most realistic model for many counties because it allows shared cybersecurity coordination, while preserving some office autonomy.

Table 1. Comparison of County IT Governance Models

Governance Model	Description	Strengths	Cybersecurity Risks	Compliance Implication
Centralized county IT	One primary county IT department manages most systems and security functions	Consistent policies, centralized monitoring, clear accountability	Could be slower to respond to department-specific needs and security incidents	Easier to implement countywide standards and document compliance
Decentralized IT	Multiple elected offices or departments manage their own systems and IT staff	Operational autonomy and department-specific flexibility	Inconsistent controls, duplicated tools, uneven patching, and fragmented incident response	Higher risk of compliance gaps unless formal coordination exists
Hybrid IT governance	County IT provides shared services while elected offices retain some system control	Balances coordination and autonomy	Requires strong communication and defined responsibilities	Most realistic model for many Ohio counties; requires formal governance agreements

This comparison in Table 1 demonstrates that cybersecurity compliance is not only a technical issue, but also a governance issue. Centralized IT governance could make it easier to implement uniform cybersecurity policies, maintain asset inventories, and document compliance activities. Decentralized governance could preserve operational flexibility, but it increases the risk of inconsistent controls, duplicated technology investments, and fragmented incident response. Hybrid governance could offer a practical balance, but only when counties establish clear responsibilities, shared policies, and formal coordination mechanisms across elected offices and departments.

Resource Constraints and Cybersecurity Capacity in Smaller Counties

While Ohio's cybersecurity standards establish important baseline protections for political subdivisions, implementation capacity may vary significantly across counties. Larger counties often maintain dedicated cybersecurity staff, centralized IT departments, and larger technology budgets. In contrast, many rural or smaller counties operate with limited information technology personnel and constrained financial resources. Some counties may rely on a single IT administrator who is responsible for network administration, system maintenance, and cybersecurity responsibilities simultaneously.

These disparities can create uneven cybersecurity capacity across jurisdictions. Smaller counties could face difficulty implementing advanced security technologies, conducting regular

vulnerability assessments, or maintaining continuous network monitoring. Limited staffing could also reduce the ability to perform incident response coordination or security policy development.

Addressing these capacity gaps could require increased collaboration across jurisdictions and greater reliance on state and federal support programs. Shared cybersecurity services, regional collaboration among counties, and participation in statewide cybersecurity initiatives may provide practical mechanisms for smaller counties to strengthen cybersecurity protections, while operating within resource constraints. Additional funding should be sought through grants and even county insurance providers. In recent years, cyber insurance providers have increasingly required formal cybersecurity controls, risk assessments, and incident response planning as conditions for coverage, which may further encourage counties to adopt structured cybersecurity programs (Nurse et al., 2020; Scowcroft, 2025).

Ohio guidance identifies the NIST Cybersecurity Framework and CIS cybersecurity best practices as examples of generally accepted cybersecurity standards that political subdivisions should use when developing cybersecurity programs. In addition, MS-ISAC provides cybersecurity resources, information sharing support, threat intelligence, risk assessments, and cybersecurity best practices for state, local, tribal, and territorial governments (Ohio Auditor of State & County Commissioners Association of Ohio, 2025; National Association of Counties, n.d.).

Taken together, this literature suggests that county cybersecurity readiness depends on more than technical safeguards alone. Cybersecurity implementation is shaped by governance structure, leadership coordination, staffing capacity, funding, vendor relationships, and the ability to translate broad cybersecurity frameworks into local government practice. These issues provide the foundation for examining how Ohio counties could operationalize the state's 2025 cybersecurity requirements.

RESEARCH QUESTIONS

This paper is guided by three research questions. First, what cybersecurity obligations apply to Ohio counties under Ohio's 2025 cybersecurity requirements and related guidance? Second, how do centralized, decentralized, and hybrid county IT governance structures influence implementation capacity? Third, how can county officials distinguish mandatory cybersecurity obligations from recommended or inferred implementation practices when developing a county cybersecurity program?

METHODOLOGY

This study uses a qualitative conceptual policy analysis supported by structured document review. The analysis examined five categories of sources: Ohio statutory requirements and implementation guidance; state cybersecurity resources and Auditor of State materials; public safety information security requirements, including Criminal Justice Information Services (CJIS) and Law Enforcement Automated Data System (LEADS) requirements; recognized cybersecurity frameworks and best-practice sources, including National Institute of Standards and Technology (NIST), Center for Internet Security (CIS), and Multi-State Information Sharing and Analysis Center (MS-ISAC) guidance; and scholarly literature on public-sector IT governance, local government cybersecurity, and decentralized technology management.

Sources were selected based on their relevance to county cybersecurity governance, their applicability to Ohio political subdivisions, and their ability to clarify either legal obligations or implementation practices. Documents were analyzed using a source-classification approach. Each requirement or recommendation was categorized as a statutory requirement, state guidance, public safety compliance obligation, recognized cybersecurity best practice, cyber-insurance expectation, or author-inferred implementation recommendation. This classification was used to develop the compliance-readiness crosswalk and to distinguish mandatory requirements from recommended or resource-dependent practices. The documents were reviewed for references to cybersecurity governance responsibilities, required or recommended security practices, implementation responsibilities, reporting obligations, and evidence that county officials could use to demonstrate compliance readiness.

This approach is appropriate because the purpose of the manuscript is not to empirically test cybersecurity controls, but to synthesize legal, policy, governance, and cybersecurity guidance into a county-level implementation framework. Future empirical research should test the framework through surveys, interviews, or case studies involving Ohio county officials and public sector information technology leaders.

OHIO'S 2025 CYBERSECURITY LEGISLATIVE FRAMEWORK

Ohio's recent cybersecurity legislation represents a significant development in state-level efforts to improve cybersecurity resilience across local governments. House Bill 96 established new cybersecurity requirements for political subdivisions, which includes counties, municipalities, and special districts.
Ohio Auditor of State Bulletin 2025-007 provides additional implementation guidance for these requirements and clarifies that political subdivisions must adopt a formal cybersecurity program that is designed to safeguard information technology systems and data resources, while ensuring the confidentiality, integrity, and availability of those systems (Faber, 2025).

Under these standards, local governments must establish formal cybersecurity programs that align with recognized industry frameworks, such as the NIST Cybersecurity Framework (H.B. 96, 2025). The NIST Cybersecurity Framework provides a recognized structure for identifying, protecting, detecting, responding to, and recovering from cybersecurity risks (National Institute of Standards and Technology, 2018). The legislation emphasizes the need for structured cybersecurity risk management practices and requires political subdivisions to document cybersecurity policies and incident response procedures.

The CyberOhio program further clarifies the operational expectations for local governments. These requirements support compliance with Ohio cybersecurity standards and state law, while ensuring the confidentiality, integrity, and availability of county information systems. Ohio law requires political subdivisions to adopt a cybersecurity program that is designed to safeguard information systems and protect data resources from cybersecurity threats (CyberOhio, 2025; Faber, 2025).

In addition to legislative mandates, the Ohio Auditor of State and the County Commissioners Association of Ohio issued Bulletin 2025-13 to provide implementation guidance for counties. The bulletin requires local governments to define cybersecurity leadership roles, conduct cybersecurity risk assessments, and maintain documented ransomware response procedures (Ohio Auditor of State & County Commissioners Association of Ohio, 2025).

These requirements represent a shift toward standardized cybersecurity governance across Ohio's political subdivisions. However, implementation remains challenging in counties where IT authority is distributed across multiple elected offices.

DIVISION OF STATE AND COUNTY RESPONSIBILITIES

Ohio's cybersecurity requirements create responsibilities at both the state and local levels. At the state level, Ohio establishes statutory cybersecurity expectations, provides implementation guidance, receives required cybersecurity incident reports, and offers training or support resources through state cybersecurity initiatives. At the county level, political subdivisions remain responsible for adopting and implementing cybersecurity programs, documenting policies and procedures, conducting risk assessments, training employees, and responding to cybersecurity incidents.

This distinction is important because counties vary significantly in staffing, budget capacity, and IT governance structure. Smaller counties may not have dedicated cybersecurity staff or centralized IT departments, which could make implementation more difficult. As a result, state guidance, shared services, regional partnerships, grant funding, and template-based implementation resources could be necessary to reduce the compliance burden on counties with limited technical capacity.

RELATIONSHIP TO CJIS AND LEADS SECURITY REQUIREMENTS

County cybersecurity governance must account for both statewide cybersecurity requirements and criminal justice information security regulations when managing public safety information systems. While Ohio's cybersecurity law establishes baseline cybersecurity requirements for political subdivisions, these requirements do not replace or supersede existing security regulations governing criminal justice information systems. County governments that operate law enforcement agencies, such as county sheriff's offices, must continue to comply with the CJIS Security Policy and LEADS requirements when maintaining systems that access criminal justice data. Because LEADS serves as Ohio's gateway to national criminal justice information systems maintained through the FBI's Criminal Justice Information Services Division, agencies that access LEADS must comply with the CJIS Security Policy as a condition of system access (Federal Bureau of Investigation, 2024).

The CJIS Security Policy establishes minimum security requirements for any entity that accesses or manages criminal justice information (CJI). These requirements include provisions related to access control, authentication, encryption, personnel security, incident response, and system auditing (Federal Bureau of Investigation, 2024). The policy applies to state and local law enforcement agencies, as well as any governmental or contracted entities that process or store criminal justice information.

Similarly, the Ohio Law Enforcement Automated Data System LEADS provides statewide access to criminal justice databases and requires participating agencies to follow strict security and operational policies. LEADS policies establish requirements for user authentication, system monitoring, training, and data protection for agencies accessing the system (LEADS, 2023). Because the county sheriff's office relies on LEADS to access state and national criminal justice information systems, the county must maintain compliance with LEADS policies that are in addition to any other applicable cybersecurity requirements.

Ohio's cybersecurity law should therefore be viewed as an additional governance and security framework, rather than a replacement for existing public safety information security standards. Counties that operate public safety agencies must ensure that cybersecurity programs adopted under Ohio law align with CJIS and LEADS security requirements. In practice, this means that county cybersecurity policies should incorporate CJIS security controls, incident response procedures, and personnel security requirements where criminal justice systems are involved.

Failure to maintain compliance with CJIS or LEADS security requirements may result in administrative sanctions, including suspension of system access for the agency or the jurisdiction responsible for the violation (LEADS, 2023). Because access to these systems is critical for law enforcement operations, maintaining compliance with CJIS and LEADS security policies remains a fundamental responsibility for county governments operating public safety agencies.

SOURCE-MAPPED CYBERSECURITY IMPLEMENTATION RECOMMENDATIONS

The recommendations in this paper are organized according to their source and compliance function. Some actions, such as adopting a cybersecurity program, documenting incident response procedures, and following statutory reporting requirements, are tied directly to Ohio law and state guidance. Other actions, such as implementing multifactor authentication, maintaining backups, conducting risk assessments, and providing security awareness training, are recognized cybersecurity best practices and CJIS-related safeguards that support compliance, but may vary in implementation depending on county size, staffing, and budget.

Counties should avoid treating cybersecurity compliance as a generic checklist. Instead, county leaders should classify each action according to whether it is mandatory, recommended, or inferred from recognized best practices. This distinction is especially important for smaller counties because it helps leadership prioritize legally required actions, while still building toward broader cybersecurity maturity over time. This source-mapped approach also strengthens accountability because county officials can connect each implementation activity to a legal requirement, state guidance, public safety compliance, or recognized cybersecurity framework.

OPPORTUNITIES THROUGH FEDERAL CYBERSECURITY FUNDING

Federal programs offer opportunities for counties to strengthen cybersecurity capabilities. The State and Local Cybersecurity Grant Program, administered through the Federal Emergency Management Agency, provides funding for cybersecurity planning, vulnerability assessments, and infrastructure improvements (Federal Emergency Management Agency, 2025).

Participation in these programs often requires evidence of coordinated governance and risk-based cybersecurity planning. Counties that establish formal cybersecurity programs and governance structures could therefore be better positioned to compete for federal cybersecurity funding.

These funding opportunities may be particularly important for smaller counties that lack dedicated cybersecurity budgets or specialized security personnel. Grant funding can help counties perform risk assessments, acquire cybersecurity monitoring tools, develop incident response plans, and provide workforce training. By leveraging these resources, smaller jurisdictions may be able to achieve compliance with state cybersecurity requirements without placing excessive strain on local budgets.

IMPLICATIONS OF NON-COMPLIANCE WITH OHIO CYBERSECURITY REQUIREMENTS

Although Ohio's cybersecurity law does not establish direct monetary penalties for failure to adopt a cybersecurity program, political subdivisions that fail to comply with these requirements may still face significant administrative, financial, and operational consequences. Ohio Revised Code §9.64 requires the legislative authority of each political subdivision to adopt a cybersecurity program designed to safeguard the confidentiality, integrity, and availability of government information systems (Ohio Rev. Code § 9.64, 2025). Responsibility for adopting a cybersecurity program ultimately rests with the legislative authority of the political subdivision, which may include county commissioners, or other governing bodies responsible for approving cybersecurity policies and allocating resources (Ohio Rev. Code § 9.64, 2025).

Beyond regulatory considerations, cybersecurity incidents affecting county governments could disrupt critical public services, such as emergency responses, court operations, and public records access, as well as cause reputational damage for elected officials. This highlights the operational importance of maintaining effective cybersecurity programs.

Audit Findings and State Oversight

One of the primary enforcement mechanisms associated with the law occurs through the state audit process. The Ohio Auditor of State has indicated that compliance with the cybersecurity program requirement will be evaluated during regular financial audits of political subdivisions. If auditors determine that a local government has not adopted an appropriate cybersecurity program, or has failed to follow incident reporting requirements, the deficiency could result in a formal audit finding (Ohio Auditor of State, n.d.).

Audit findings can require corrective action and may remain on record until the deficiency is resolved. Repeated findings may also create reputational concerns for local governments responsible for protecting sensitive public data and maintaining secure information systems.

Incident Reporting and Compliance Obligations

The law also establishes mandatory reporting requirements for cybersecurity incidents. Political subdivisions that discover a cybersecurity incident must notify the Ohio Department of Public Safety's Division of Homeland Security within seven days of discovery, and must notify the Ohio Auditor of State within thirty days (Ohio Rev. Code § 9.64, 2025).

Failure to comply with these reporting requirements may be identified during state audits and could result in additional audit findings or oversight actions.

Financial Liability and Findings for Recovery

In some circumstances, cybersecurity-related actions could also result in financial liability for public officials. For example, Ohio guidance states that a political subdivision may only pay a ransomware demand if the local legislative authority approves the payment through a formal resolution that explains why the payment is in the public interest. Without such authorization, a ransom payment may be considered an illegal expenditure of public funds (Ohio Auditor of State, n.d.).

If auditors determine that public funds were improperly used, the Auditor of State may issue a finding for recovery, requiring repayment of the funds by the responsible officials or employees.

Governance and Operational Consequences

Beyond legal accountability, failure to implement appropriate cybersecurity safeguards may expose counties to increased cyber risk. Cyber incidents, such as ransomware attacks, data breaches, or system disruptions, could significantly impact government operations, potentially interrupt public services and expose sensitive information. Cyber attacks targeting local governments have increased in recent years, often exploiting outdated systems, limited cybersecurity budgets, and insufficient security controls (Faber, 2025).

Consequently, non-compliance with cybersecurity requirements may not only result in audit findings, but may also increase the likelihood of operational disruptions, financial losses, and damage to public trust.

COMPLIANCE-READINESS CROSSWALK FOR OHIO COUNTIES

Table 2 provides a source-mapped compliance-readiness crosswalk for Ohio counties. The table distinguishes mandatory requirements from recommended or inferred implementation practices and identifies how each area supports county cybersecurity governance. This structure is intended to help county officials understand not only what should be implemented, but also why each activity is included, and what type of evidence could demonstrate implementation.

Table 2. Compliance-Readiness Crosswalk for Ohio Counties

Requirement or Practice Area	Source Type	County Responsibility	Recommended Evidence	Small-County Implementation Note
Adopt a formal cybersecurity program	Statutory requirement	Legislative authority adopts a cybersecurity program	Approved cybersecurity policy or program document	Use state or association templates where available

Define cybersecurity leadership	State guidance / governance best practice	Assign responsibility for cybersecurity coordination	Designated IT Director, CISO, administrator, or equivalent role	Assign duties to an existing senior IT leader
Conduct cybersecurity risk assessments	State guidance / recognized best practice	Identify critical systems, vulnerabilities, and risks	Risk assessment report and remediation plan	Begin with high-risk systems, such as public safety and finance
Maintain incident response procedures	Statutory and state guidance	Document reporting, containment, and recovery procedures	Incident response plan and tabletop exercise records	Use tabletop exercises to clarify roles before an incident occurs
Report cybersecurity incidents	Statutory requirement	Notify required state entities within reporting timeframes	Incident reporting procedure and contact list	Pre-identify reporting contacts and approval chain
Address ransomware payment approval	Statutory / state guidance	Require formal legislative approval before paying a ransom demand	Ransomware response procedure and legislative approval process	Include commissioners, prosecutor, auditor, and IT leadership
Implement multifactor authentication (MFA) and access controls	Recognized cybersecurity best practice / public safety compliance / state guidance	Protect remote access, administrative accounts, and sensitive systems	MFA configuration records and access review documentation	Prioritize remote access and privileged accounts
Maintain backups and recovery procedures	Recognized cybersecurity best practice / cyber-insurance expectation	Protect data and restore critical services after disruption	Backup schedules, restore test records	Test restoration of the most critical systems first
Provide cybersecurity awareness training	State guidance	Train employees to identify malicious activity and reporting procedures	Training completion records	Use no-cost or low-cost state and federal training resources
Coordinate with external cybersecurity partners	Recommended best practice	Participate in information sharing, and grant opportunities	MS-ISAC membership and grant applications	Shared services or regional partnerships could reduce burden

The crosswalk should be used as a readiness tool and not a one-size-fits-all compliance checklist. County officials may need to adapt implementation activities based upon staffing, budget capacity, public safety responsibilities, and governance structure. However, each county should be able to document how its cybersecurity program addresses statutory requirements, state guidance, incident response obligations, risk assessment practices, employee training, and coordination with external cybersecurity partners.

IMPLICATIONS FOR COUNTY LEADERSHIP

The implementation of Ohio's cybersecurity requirements carries significant implications for county leadership and governance structures. Because Ohio counties operate under a decentralized framework where many elected officials maintain independent operational authority, cybersecurity governance requires coordination across multiple offices and departments. Ohio Revised Code §9.64 assigns responsibility for adopting a cybersecurity program to the legislative authority of the political subdivision, which typically includes the county board of commissioners, or other governing authority responsible for approving policies and allocating financial resources (Ohio Rev. Code § 9.64, 2025). As a result, county leadership plays a critical role in establishing the institutional framework that is necessary to support this cybersecurity compliance.

Effective cybersecurity governance requires clear communication and collaboration among county commissioners, elected officials, department directors, and information technology personnel. While the county automatic data processing board, authorized under Ohio Revised Code §307.842, can provide coordination for technology decisions, the statute does not grant enforcement authority over independently elected offices. Consequently, successful implementation of cybersecurity programs will often depend on voluntary cooperation and shared governance agreements among county leadership (Ohio Rev. Code § 307.842, 2011). County leaders may therefore need to facilitate cross-departmental collaboration and establish countywide cybersecurity policies that balance operational autonomy with consistent security practices.

County leadership must also address the financial and workforce implications associated with cybersecurity program implementation. Many counties operate with limited information technology budgets and small IT departments, which could constrain the ability to implement advanced cybersecurity technologies or maintain dedicated cybersecurity personnel. National research on county cybersecurity capacity indicates that local governments frequently face cybersecurity workforce shortages and limited funding for security infrastructure (National Association of Counties, 2023). County commissioners and other governing officials may therefore need to prioritize cybersecurity investments, pursue grant funding opportunities, and explore regional or shared service models to strengthen cybersecurity hygiene.

In addition to resource allocation, county leadership must consider the broader risk management implications of cybersecurity governance. Cyber incidents affecting local governments can disrupt essential services, compromise sensitive data, and create financial liability for public entities. Guidance from CIS and MS-ISAC emphasizes that effective cybersecurity governance requires leadership engagement, risk management planning, and the adoption of recognized cybersecurity frameworks to reduce organizational risk (Center for Internet Security, 2023). By actively

supporting cybersecurity governance initiatives, county leaders can strengthen institutional resilience and improve the protection of critical government systems.

Ultimately, the success of Ohio's cybersecurity requirements will depend not only on the technical controls and compliance measures, but also on the leadership structures that guide technology governance across county government. County leaders who prioritize cybersecurity governance, allocate appropriate resources, and encourage interdepartmental collaboration will be better positioned to protect public infrastructure and maintain trust in local government from their citizens.

In counties where information technology responsibilities are distributed across multiple departments or elected officials, establishing formal cybersecurity governance structures or interdepartmental cybersecurity committees may improve coordination and accountability for successful cybersecurity policy implementation.

IMPLEMENTATION RISKS AND UNINTENDED CONSEQUENCES

Although Ohio's cybersecurity requirements may strengthen local government cybersecurity readiness, they could also create implementation challenges for counties with limited staffing, fragmented governance, and constrained budgets. Smaller counties may struggle to implement formal cybersecurity programs, risk assessments, incident response procedures, training, and documentation without funding, templates, technical assistance, or shared service models.

These mandates could also create unintended consequences if counties treat compliance as documentation rather than risk management. A county could adopt policies to satisfy audit expectations while still lacking the staffing, tools, or expertise needed to implement them effectively. Decentralized counties may also face inconsistent implementation when elected offices interpret requirements differently or resist countywide coordination.

Implementation costs may be justified when cybersecurity programs reduce the likelihood of ransomware, data breaches, public safety disruption, and financial loss. However, smaller counties may need to prioritize foundational controls, such as asset inventories, access management, backups, incident response planning, and employee training, before adopting more advanced cybersecurity capabilities.

Ohio's approach for local governments is significant because it places formal cybersecurity program expectations on political subdivisions while relying on state guidance, audit processes, and local implementation.

LIMITATIONS AND FUTURE RESEARCH

This paper has several limitations that should be considered when interpreting the proposed compliance-readiness framework. First, the analysis is based upon statutory requirements, implementation guidance, recognized cybersecurity frameworks, and public sector IT governance literature rather than original empirical data collected from Ohio counties. As a result, the proposed compliance-readiness framework should be understood as a conceptual and policy-based implementation tool, rather than an empirically tested model. Second, county governments vary

significantly in size, staffing, budget capacity, public safety responsibilities, and IT governance structure, which could affect how individual counties apply the recommendations. Third, the legal and policy environment surrounding cybersecurity continues to evolve, and counties should verify current statutory requirements, Auditor of State guidance, and CyberOhio resources when implementing cybersecurity programs. Future research should empirically test the framework through surveys, interviews, or case studies involving Ohio county IT leaders, commissioners, auditors, sheriffs, records officials, emergency communication leaders, and other elected or appointed stakeholders. Future research could also compare Ohio's approach with other state cybersecurity mandates because state governments have adopted different cybersecurity policy models involving training, ransomware response, funding, insurance, and local government support (Ganapati et al., 2023).

CONCLUSION

Cybersecurity has become an essential component of modern public administration. As local governments expand their reliance on digital systems, cybersecurity risks will continue to increase in both frequency and complexity. Ohio's 2025 cybersecurity legislation represents an important step toward improving cybersecurity readiness across the state's political subdivisions. However, successful implementation will require counties to coordinate cybersecurity governance, allocate resources strategically, and adopt standardized cybersecurity practices.

Successful implementation of these requirements will depend not only on technical controls, but also on governance structures, workforce capacity, and access to cybersecurity resources across counties of varying size and capability. Counties must adopt and implement cybersecurity programs that align with these requirements no later than January 1, 2026, which further underscores the urgency of developing structured cybersecurity governance and risk management practices (Faber, 2025). By strengthening governance structures, conducting regular risk assessments, and investing in cybersecurity workforce development, Ohio counties can improve their cybersecurity posture, while continuing to deliver essential public services effectively.

As Ohio counties continue to modernize public services and expand digital infrastructure, cybersecurity governance will become an increasingly central component of effective public administration. Strengthening cybersecurity governance structures, improving coordination across county departments, and investing in cybersecurity capacity will be critical to ensuring the resilience of local government information systems and the protection of sensitive public data.

REFERENCES

Beaumaster, S. (2002). Local government IT implementation issues: A challenge for public administration. In *Proceedings of the 35th Annual Hawaii International Conference on System Sciences* (pp. 1725–1734). IEEE. https://doi.org/10.1109/HICSS.2002.994084

Center for Internet Security. (2023). Implementation guide for small- and medium-sized enterprises CIS Controls IG1. https://www.cisecurity.org/insights/white-papers/implementation-guide-for-small-and-medium-sized-enterprises-cis-controls-ig1

Chen, D. Q., Mocker, M., Preston, D. S., & Teubner, A. (2010). Information systems strategy: Reconceptualization, measurement, and implications. *MIS Quarterly*, 34(2), 233–259. https://doi.org/10.2307/20721426

CyberOhio. (2025, July 1). *New local government cyber standards*. State of Ohio. https://cyber.ohio.gov/news-and-events/all-news/new-local-government-cyber-standards

Denford, J. (2016, February 4). *Idea to retire: Decentralized IT governance*. Brookings Institution. https://www.brookings.edu/articles/idea-to-retire-decentralized-it-governance/

E-Government Act of 2002, Pub. L. No. 107-347, 116 Stat. 2899 (2002). https://www.congress.gov/bill/107th-congress/house-bill/2458

Faber, K. (2025, August 27). *Auditor of State Bulletin 2025-007: Adoption of cybersecurity program*. Ohio Auditor of State. https://ohioauditor.gov/publications/bulletins/2025/2025-007.pdf

Federal Bureau of Investigation. (2024). *Criminal Justice Information Services (CJIS) security policy*, version 6.0. U.S. Department of Justice. https://www.fbi.gov/services/cjis/cjis-security-policy-resource-center

Federal Emergency Management Agency. (2025, September 3). *Fiscal year 2025 State and Local Cybersecurity Grant Program fact sheet*. U.S. Department of Homeland Security. https://www.fema.gov/fact-sheet/fiscal-year-2025-state-and-local-cybersecurity-grant-program-fact-sheet

Fusi, F., Jung, H., & Welch, E. (2025). Technological vulnerability and knowledge of cyber-incidents: Threats to innovativeness in local governments? Public Management Review, 27(3), 545–571. https://doi.org/10.1080/14719037.2023.2250362

Ganapati, S., Ortega Franco, L., & Le, A. N. (2023). American state government cybersecurity policies. In *Proceedings of the 24th Annual International Conference on Digital Government Research: Together in the Unstable World: Digital Government and Solidarity* (pp. 637–638). Association for Computing Machinery. https://doi.org/10.1145/3598469.3598540

H.B. 96, 136th Gen. Assemb., Reg. Sess. (Ohio 2025). https://www.legislature.ohio.gov/legislation/136/hb96

Ioannou, K., Kamariotou, M., & Kitsios, F. (2023). Digital transformation strategy in the public sector: A SEM-neural network model for IS users' satisfaction. In 2023 8th International Conference on Mathematics and Computers in Sciences and Industry (MCSI) (pp. 116–121). IEEE. https://doi.org/10.1109/MCSI60294.2023.00026

LEADS. (2023). Manuals. Ohio Law Enforcement Automated Data System. https://leads.ohio.gov/Manuals

National Association of Counties. (2023). Cybersecurity: Protecting our nation's counties: Study of resilience of U.S. counties. https://www.naco.org

National Association of Counties. (n.d.). Cybersecurity for counties: Protect your county with NACo and the MS-ISAC. https://www.naco.org/page/cybersecurity-counties

National Institute of Standards and Technology. (2018). *Framework for improving critical infrastructure cybersecurity* (Version 1.1). https://doi.org/10.6028/NIST.CSWP.04162018

Nurse, J. R. C., Axon, L., Erola, A., Agrafiotis, I., Goldsmith, M., & Creese, S. (2020). The data that drives cyber insurance: A study into the underwriting and claims processes. arXiv. https://arxiv.org/abs/2008.04713

Nussbaum, B., & Park, S. (2018). A tough decision made easy? Local government decision-making about contracting for cybersecurity. In Proceedings of the 19th Annual International Conference on Digital Government Research: Governance in the Data Age

(Article 28, pp. 1–9). Association for Computing Machinery. https://doi.org/10.1145/3209281.3209368

Ohio Auditor of State. (n.d.). Cybersecurity policy program. Retrieved June 21, 2026, from https://ohioauditor.gov/fraud/cybersecurity-policy.html

Ohio Auditor of State & County Commissioners Association of Ohio. (2025, September). Bulletin 2025-13: County cybersecurity programs. Ohio Auditor of State. https://ccao.org/aws/CCAO/asset_manager/get_file/933056?ver=1

Ohio Rev. Code § 9.64 (2025). https://codes.ohio.gov/ohio-revised-code/section-9.64

Ohio Rev. Code § 307.842 (2011). https://codes.ohio.gov/ohio-revised-code/section-307.842

Organisation for Economic Co-operation and Development. (2024). *2023 OECD digital government index: Results and key findings*. OECD Publishing.

Scowcroft, J. (2025, October 6). *Top cyber insurance requirements to qualify for coverage*. SymQuest Tech Talk. https://blog.symquest.com/cyber-insurance-requirements

Villani, J. J., Jr. (2022). A cybersecurity profile of Pennsylvania county governments (Order No. 30240790). Available from ProQuest Central; ProQuest Dissertations & Theses Global. (2763378678).

IACIS
Proceedings of the 66th International Association for Computer Information Systems Conference - October 7 - 10, 2026 | Pensacola, FL

SYSTEM-LEVEL INTEGRATION OF ARTIFICIAL INTELLIGENCE IN SCADA-BASED POWER DISTRIBUTION NETWORKS: ARCHITECTURE AND PERFORMANCE EVALUATION

Matelier Numbi, Southern Utah University, mateliernumbi20@gmail.com

ABSTRACT

This paper presents a system-level integration framework for embedding artificial intelligence (AI) into SCADA-based power distribution networks to enhance operational performance and decision-making. A four-layer reference architecture maps AI inference across field devices, substation edge systems, control-center, and deterministic safety layers. The framework is evaluated on the IEEE 33 bus feeder using Monte Carlo fault injection (N = 10,000). The simulation results show that edge-resident AI reduces Fault Detection Latency (FDL) by 56.7 %, Restoration Time (RT) by 60.0 %, and False Alarm Rate (FAR) by 62.4% compared to conventional SCADA. The Operational Degradation Index (ODI) decreases from 0.180 to 0.071, indicating improved reliability. A Proximal Policy Optimization (PPO) Reinforcement Learning agent for Volt-VAR Optimization reduces active power losses by 8.6% and improves voltage compliance to 94.3%. The safety validation layer verifies that AI controls actions in 99.97% of cases to ensure safe execution. Therefore, these results illustrate that the integration of AI as a system-level component can enable SCADA to be used as a real-time decision support platform for critical infrastructure. From an information systems perspective, the proposed AI-SCADA framework functions as an integrated decision support system that coordinates data acquisition, analytics, and control logic across multiple layers. The architecture and performance evaluation directly address information flow, latency management, and system-level reliability within a critical infrastructure information system.

Keywords: artificial intelligence, SCADA systems, power distribution, fault detection, decision support systems.

INTRODUCTION

Modern power grid networks have increased operational complexity because of the rapid integration of distributed energy resources (DERs), as well as their associated bidirectional power flows. In addition, the use of smart sensors is creating a massive amount of data that is overwhelming the analytical capabilities of traditional SCADA systems. These systems were originally developed for centralized monitoring rather real-time (RT) intelligent control. (Chung & Zhang, 2023; Alhamrouni et al., 2024).

To help mitigate some of the complexities in modern electrical grid systems, artificial intelligence (AI) has been increasingly applied to address these challenges, resulting in strong performance in fault detection, predictive maintenance, and voltage regulation (Hoffmann et al., 2020; Zhang et al., 2024; Li et al., 2021). In particular, deep reinforcement learning (DRL) techniques have been successful in optimizing Volt-VAR and adapting controls (Zhang et al., 2021; Yang et al., 2020).

Edge-based AI techniques have shown a significant reduction in distribution system detection latency (Gultekin et al., 2022; Kumar & Singh, 2022).

Despite these advancements, most existing implementations treat AI as a standalone analytical tool or a loosely coupled overlay with SCADA architectures rather than an integrated system component (Hosseini & Parvania, 2021). As a result, key challenges remain unresolved, including the lack of clearly defined inference placement, limited coordination across system layers, absence of deterministic safety validation mechanisms and insufficient system-level performance evaluation frameworks.

These limitations are particularly relevant in modern distribution networks, where increasing DER penetration, communication constraints, and dynamic operating conditions require coordinated, real-time decision-making. Existing research largely focuses on improving individual AI models, while overlooking the broader architectural context in which these models operate.

This paper addresses these gaps through the proposal of a system-level AI-SCADA integration architecture that embeds AI into operational layers for coordinated data processing, decision making, and controls across multiple levels of the system while accounting for both latency budgets and safety requirements.

From an information systems standpoint, SCADA-based power distribution networks are large-scale operational information systems that integrate data collection, processing, and control across heterogeneous components The proposed architecture embeds AI as a core information-processing layer, transforming SCADA from a monitoring-only platform into a coordinated decision support system. The study therefore focusses on information system design, architecture, and performance evaluation rather than purely on hardware engineering.

The contributions of this paper are threefold:

(i) A four-layered reference architecture for AI-SCADA integration with explicit latency budgets, and defined functional roles.
(ii) A simulation-based performance evaluation using the IEEE 33-bus feeder and Monte Carlo methods (N = 10,000), and a PPO-based deep reinforcement learning approach for Volt-VAR optimization.
(iii) A sensitivity analysis to examine the impact of communication latency, DER penetration, and false alarm rates on system performance.

LITERATURE REVIEW

The current state of research shows a broad range of applications of artificial intelligence in various aspects of power distribution systems including fault detection, predictive maintenance and voltage management. For instance, machine learning (ML) and deep learning (DL) techniques have demonstrated strong performance in detecting anomalies and improving system reliability (Chung & Zhang, 2023; Alhamrouni et al., 2024; Li et al., 2021). Additionally, predictive maintenance models based on ML have also shown improvements in equipment reliability and failure predictions in medium voltage switchgear systems (Hoffmann et al., 2020). Deep

reinforcement learning (DRL) has emerged as one of the leading techniques used in voltage control for optimizing Volt-VAR. Studies conducted by Zhang et al. (2021) and Yang et al. (2020) demonstrate that DRL-based control strategies provide better results compared to rule-based control strategies when it comes to minimizing power loss and improving voltage stability. However, these studies primarily focus on algorithmic performance in isolated task environments rather than integrating them into the overall operational environment. The study by Gultekin et al. (2022) and Kumar & Singh, (2022) explores edge computing, real-time AI deployment and AI at the edge of power grids systems where edge-based AI significantly reduces detection latency and enables faster response to faults. While these approaches improve responsiveness, they often lack a structured framework for coordinating decision-making across multiple SCADA layers.

In addition, AI deployment in smart grid environments have gained considerable attention regarding cybersecurity issues, particularly in relation to data integrity, model security, and system resilience. These issues highlight the need for controlled and validated integration of AI in critical infrastructures systems. Despite the advances, there are still many limitations that must be addressed to improve the reliability of these systems. One major limitation is that most of the studies focus on individual application of AI, rather than system-level integration within SCADA architectures. Another major limitation resides on the impact of communication latency and infrastructure constraints on AI performance which are often overlooked. Lastly, few researchers include the deterministic safety validation mechanisms that ensure safe deployment of AI-driven control actions. The literature review explicitly incorporates recent studies, including edge-based fault detection, explainable AI for switchgear, and AI-driven voltage control, ensuring that the proposed framework is aligned with the latest developments in AI applications for power distribution systems.

Table 1. Summary of Related Work and Limitations

Study Authors	Year	Application	Approach	Limitation
Chung & Zhang	2023	Distribution monitoring	AI-based analytics	No system-level integration
Gultekin et al.	2022	Fault detection	Edge AI	Limited multi-layer coordination
Hosseini & Parvania	2021	System resilience	AI-based methods	No safety Validation
Zhang et al.	2021	Volt-VAR optimization	DRL	Isolated control focus
Hoffman et al.	2020	Predictive maintenance	ML models	Limited real-time capability
Yang et al.	2020	Voltage control	DRL	No SCADA integration

These observations highlight clear gaps in existing research: the absence of a unified framework allowing the integration of AI across SCADA layers while addressing latency constraints, coordination and operational safety.

METHODOLOGY

This section presents the study design, which uses a simulation-based quantitative methodology to evaluate the performance of the proposed AI-SCADA integration framework within power distribution systems. This approach combines system modeling, Monte Carlo fault simulation (N = 10,000), and machine learning-based decision-making in a unified evaluation environment. The proposed AI-integrated framework is compared to the traditional SCADA baseline across multiple performance metrics. The goal of this study is to quantify improvements in detection latency, restoration efficiency, as well as the overall operational reliability under varying system conditions.

Four-Layer Architectures Framework

To achieve the proposed study, Fig. 1 describes a four-layer hierarchical layers architecture that distributes AI functionality to ensure efficient data handling, low latency response, and safe execution of control actions.

Layer A (Field/IED)

It captures raw measurements from field devices and hosts lightweight machine learning (ML) models for rapid inference, typically within 20-80 ms. Communication is performed using IEC 61850 GOOSE to support fast event-driven messaging within substation environments (IEC, 2013).

Layer B (Substation Edge Compute)

It performs multi-IED data fusion and hosts the primary fault detection models within a 150-200 ms latency budget using IEC 61850 MMS. This layer acts as a primary point for time-critical detection and classification tasks, where minimizing communication delay is essential.

Layer C (SCADA Supervisory)

It executes system-wide optimization tasks, including deep reinforcement learning (DRL)-based Volt-VAR control, load forecasting, and predictive maintenance. These functions operate within typical SCADA scan cycles of 2-10 seconds and provide coordination across feeders and substations.

Layer D (Deterministic Safety Gate)

It serves as a mandatory validation mechanism that intercepts all AI-generated control actions before execution. It enforces N-1 contingency constraints, switching interlock logic, and confidence thresholding to ensure safe and reliable system operation.

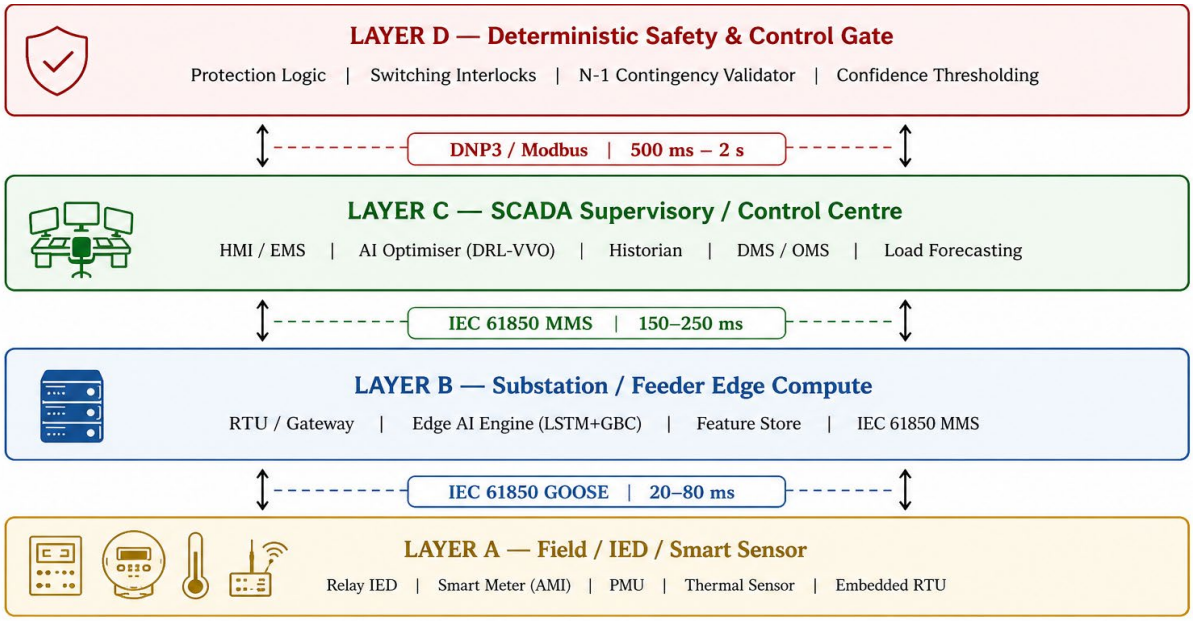


Figure 1. Four Layer AI-SCADA Reference Architecture

This layered architecture ensures that latency-sensitive decisions are handled closer to the field while computationally intensive optimization is performed at supervisory levels.

Integration Workflow

To ensure the integration of AI into the SCADA environment, a structured workflow is implemented to verify data quality, system reliability, and safe operation. This process is organized into five phases each addressing a key aspect of deployment and performance.

Phase 1: Signal Definition and Time Synchronization

It focuses on defining system signals and establishing accurate time synchronization using IEEE 1588 Precision Time Protocol (PTP). High synchronization accuracy is required, with ± 1 μs for fault detection functions; ± 1 ms for optimization tasks, to ensure consistent system behavior across distributed components.

Phase 2: Data Conditioning

It applies data conditioning through filtering, normalization, and imputation to ensure data quality and model reliability. Butterworth filtering is also applied to reduce noise in raw signals, followed by z-score normalization to standardize input values across features. To verify that missing data are handled, a multivariate imputation is applied by Chained Equations (MICE), which enables robust reconstruction of incomplete datasets. Finally, the framework assumes at least 12 months of historical data with a minimum of 95% completeness to support stable model performance.

Phase 3: Latency Budget Allocation

It is a critical factor in determining where AI functions should be deployed within the architecture. The total system latency is defined as:

$L_{total} = L_{inf} + L_{comm} + L_gate$ (ms)

Where L_inf represents inference time, L_comm communication delay, and L_gate the delay introduced by the safety validation layer. Functions requiring a total latency below 200 ms are assigned to Layer B, while less time-sensitive tasks are handled at Layer C.

Phase 4: Automation Policy Definition

It defines how AI decisions are deployed in practice. A risk-based automation strategy is adopted, consisting of three levels as follows:

- Tier 1: advisory recommendations only
- Tier 2: supervised automation
- Tier 3: autonomous operation with mandatory validation for high-risk operations.

Phase 5: Continuous Performance Monitoring

This final phase ensures long-term system reliability by continuously monitoring model performance. CUSUM and EWMA techniques are used to detect drift and identify degradation in model accuracy over time, allowing for timely updates and adjustments.

Simulation

This section presents the proposed AI-SCADA integration architecture. A simulation-based framework is used with the IEEE 33-bus distribution system and Monte Carlo (N = 10,000) fault injection to assess system performance under a wide range of operation conditions. The evaluation focuses on fault detection, restoration performance, system reliability, ad Volt-VAR optimization using a combination of machine learning and reinforcement learning models.

Test System and AI Models

This simulation is based on the IEEE 33-bus radial distribution feeder, a widely used benchmark in power system studies. The network operates at 12.66 kV with a 10 MVA base, and supports a total load of 3.175 MW across 32 branches. To reflect modern grid conditions, a 25% penetration of distributed energy resources (DERs) is introduced using photovoltaic (PV), and battery energy storage system (BESS) units.

For Volt-VAR optimization, a Proximal Policy Optimization (PPO) deep reinforcement learning (DRL) agent is implemented. The state space includes per-bus voltages, and load injections, while the action space controls the three capacitor banks (600 kVAR each) and a BESS reactive power setpoint. The agent is trained over 300 episodes, each consisting of 288 time steps.

In addition, predictive maintenance is modeled using an XGBoost-Cox survival model trained on 50 simulated switchgear assets. Failure behavior is generated using a Weibull distribution ($k = 2.3$, $\lambda = 8{,}760$ h).

Monte Carlo Fault Injection

A total of 10,000 independent Monte Carlo fault injection trials are conducted to evaluate performance. In each trial, the fault location is randomly selected from buses 2 to 32, while the fault type is sampled based on empirical probabilities (single-line-ground: 65%; line-to-line:20%; double-line-to-ground: 10%; three-phase: 5%).

Fault resistance is modeled using a log-normal distribution with a mean of 0.5 Ω, and a standard deviation of 1.2 Ω. Both traditional SCADA , and AI-integrated detection pathways are simulated independently for each trial.

To ensure statistical robustness, bootstrap resampling (B=5000) is used to compute 95% confidence intervals for all performance metrics.

Performance Metrics

To evaluate the system performance, several operational metrics are used:

Fault Detection Latency (FDL) measures the time between fault occurrence and detection

$FDL = t_{detect} - t_{fault}\ [s]$ (1)

Where t_{fault} denotes the fault occurrence time, and t_{detect} represents time at which the detection mechanism signals the fault.

Restoration Time (RT) captures the time required to restore service after detection

$RT = t_{restore} - t_{detect}\ [s]$ (2)

Where $t_{restore}$ denotes the time at which service is restored following a detected fault.

False Alarm Rate (FAR) reflects detection reliability

$$FAR = \frac{FP}{FP + TP} \quad (3)$$

Operational Degradation Index (ODI) represents the overall system impact

$$ODI = \frac{\sum di \cdot ci}{C_{total} \cdot T_{total}} \quad (4)$$

Total system latency is defined as

$L_{total} = L_{inf} + L_{comm} + L_{gate}$ (5)

For Volt-VAR optimization, the reinforcement learning reward function is defined as

$R_{(s,a)} = -0.7 \cdot Ploss_{(s,a)} - 0.3 \sum max(0, |Vi - 1| - 0.05)^2$ (6)

AI Model Specifications

To support reproducibility, the main AI models used in this study are specified in detail. The LSTM-based anomaly detection model consists of two stacked LSTM layers with 64 and 32 units, followed by a fully connected layer with 32 ReLU units and a final sigmoid output layer. The model is trained using the Adam optimizer with a learning rate of 0.001, batch size of 256, and early stopping based on validation loss after 20 epochs without improvement. Training data are split into 70% training, 15% validation, and 15% tests sets.

The predictive maintenance model uses an XGBoost-Cox survival formulation with 200 trees, maximum depth 5, learning rate 0.05, and subsample ratio 0.8. Hyper-parameters are selected through 5-fold cross-validation using concordance index as the primary metric. The Weibull-

generated failure dataset for 50 simulated switchgear assets is divided into 80% training and 20% test sets.

For Volt-VAR optimization, the PPO agent employs a neural policy network with two hidden layers of 128 and 64 units, each using ReLU activation. The agent is trained with a discount factor γ=0.99, clipping parameter ϵ=0.2, and learning rate 0.0003, using mini-batches of 1024 transitions and 10 epochs per update. Convergence is assessed based on stabilization of the cumulative reward curve and Volt-VAR performance metrics over 300 episodes.

All simulations and model training are executed on a workstation with an 8-core CPU, 32 GB RAM, and an NVIDIA GPU with 8 GB memory. Training times for the anomaly detection and PPO models are in the order of tens of minutes, and models can be retrained offline before deployment to the edge and supervisory layers.

SIMULATION RESULTS

This section presents the results of the simulation study, comparing the performance of traditional SCADA systems with the proposed AI-integrated framework. The analysis focuses on fault detection, restoration efficiency, system reliability, and optimization performance under varying operating conditions.

Primary Performance Metrics

As shown below, Table 2 summarizes the results obtained from the Monte Carlo simulation. All reported improvements are statistically significant ($p<0.0001$, Welch's t-test), confirming the effectiveness of the proposed AI-integrated architecture.

The results show a 56.7% reduction in fault detection latency, primarily due to the elimination of manual alarm acknowledgement delays through edge-based AI inference, and IEC 61850 GOOSE communication. Restoration time is reduced by 60.0%, reflecting both faster detection and the replacement of manual switching with automated Fault Location, Isolation, and Service Restoration (FLISR) operations.

In addition, system reliability improves significantly with SAIDI decreasing from 128.4 to 49.7 minutes per customer per year, and SAIFI reducing from 1.84 to 0.73 interruptions per customer per year.

It is important to note that the traditional SCADA baseline used in this study represents a conservative configuration with limited distribution automation. Many modern utilities deploy varying levels of advanced distribution automation (ADA) and FLISR schemes. As a result, the improvements reported here should be interpreted as an upper bound relative to minimally automated SCADA deployments rather than a direct comparison to the most advanced utility practices. The architectural insights remain applicable to more automated environments, but the absolute percentage gains would be smaller when compared against fully implemented ADA baselines.

Table 2. Primary Performance Metrics: Traditional vs AI-Integrated SCADA (Monte Carlo, N = 10,000)

Metric	Traditional SCADA	AI-Integrated SCADA	Improvement (% vs traditional SCADA)	95% CI (AI)
FDL (ms)	1,500 +/- 210	650 +/- 48	**56.7 %**	[627, 673]
RT (min)	12.0 +/- 2.1	4.8 +/- 0.6	**60.0 %**	[4.63, 4.97]
FAR (%)	8.5 +/- 0.9	3.2 +/- 0.4	**62.4 %**	[3.03, 3.37]
ODI	0.180 +/- 0.024	0.071 +/- 0.009	**60.6 %**	[0.066, 0.076]
SAIDI (min/cust yr)	128.4 +/- 14.2	49.7 +/- 5.8	**61.3 %**	[47.4, 52.0]
SAIFI (int./cust yr)	1.84 +/- 0.21	0.73 +/- 0.09	**60.3 %**	[0.70, 0.76]

FDL Sensitivity Latency

Fig. 2 illustrates how fault detection performance degrades as total system latency increases. Edge-deployed AI maintains more than 80% improvement in fault detection latency when total latency remains below 250 ms. However, beyond 600 ms, performance gains drop below 50% because of the backhaul delay. These results remain consistent regardless of deployment strategy, highlighting latency as a critical design constraint in AI-SCADA integration.

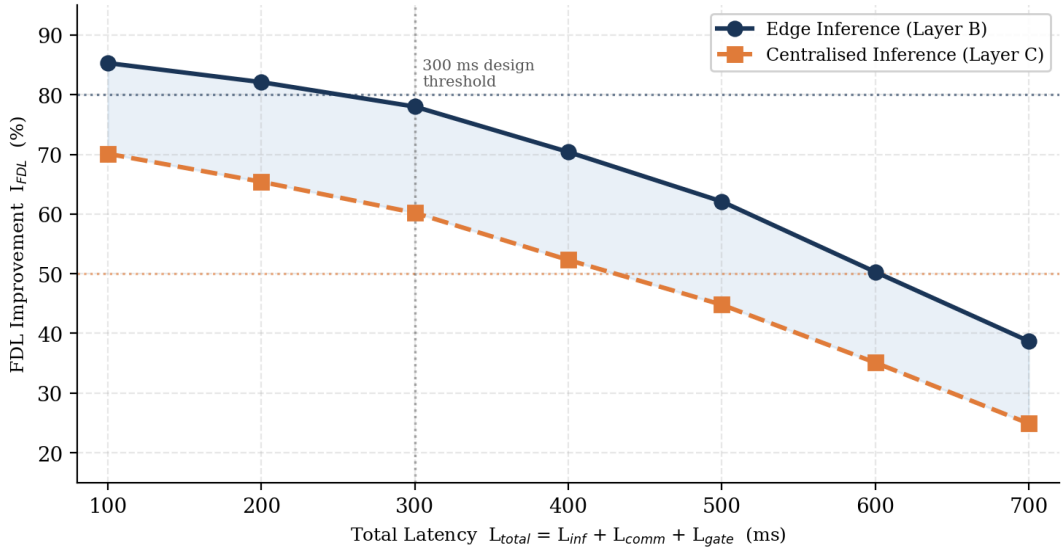


Figure 2. FDL Improvement vs. Total System Latency (L_total)

Restoration Time Comparison

Fig. 3 compares restoration time (RT) across three different operational scenarios. Full AI and Advanced Distribution Automation (ADA) integration (Tier3) achieves a mean restoration time of 4.8 min. compared to 12.0 min. for SCADA which represents a 60.0% reduction. The AI advisory-only scenario (Tier 2) yields 7.5 min., demonstrating that the 60s operator confirmation timeout is the dominant residual delay and motivating progressive expansion to Tier 3 automation as operational trust accumulates.

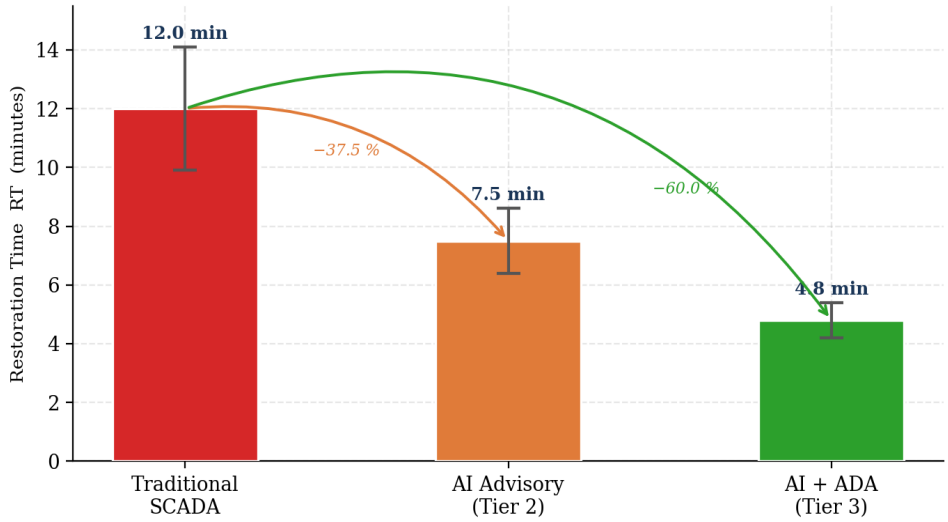


Figure 3. Restoration Time Comparison Across Three Operational Scenarios (Monte Carlo, N = 10,000)

Error bars represent ±1 standard deviation and percentages denote improvement relative to the traditional SCADA baseline.

False Alarm Rate and Hybrid Validation

Fig. 4 illustrates the relationship among of three architectures between false alarm rate (FAR), and effective restoration efficiency, defined as:

$$E_{RT} = 1 - FAR \quad (7)$$

The hybrid approach combining machine learning, rule-based and Layer D safety gate, achieves a FAR of 3.2%, which is well below the 5% operational threshold required to maintain automation trust.

In comparison, the standalone machine learning model exhibits a FAR of approximately 5.8%, while the rule-based reaches 8.5%, both exceeding acceptable operation limits under higher system variability.

The results show that standalone ML detection becomes unreliable at DER penetration levels above 30% primarily due to misclassification of load switching transients. The introduction of a

two-stage hybrid validation process effectively eliminates these false positives, significantly improving detection reliability, and enabling safe automation.

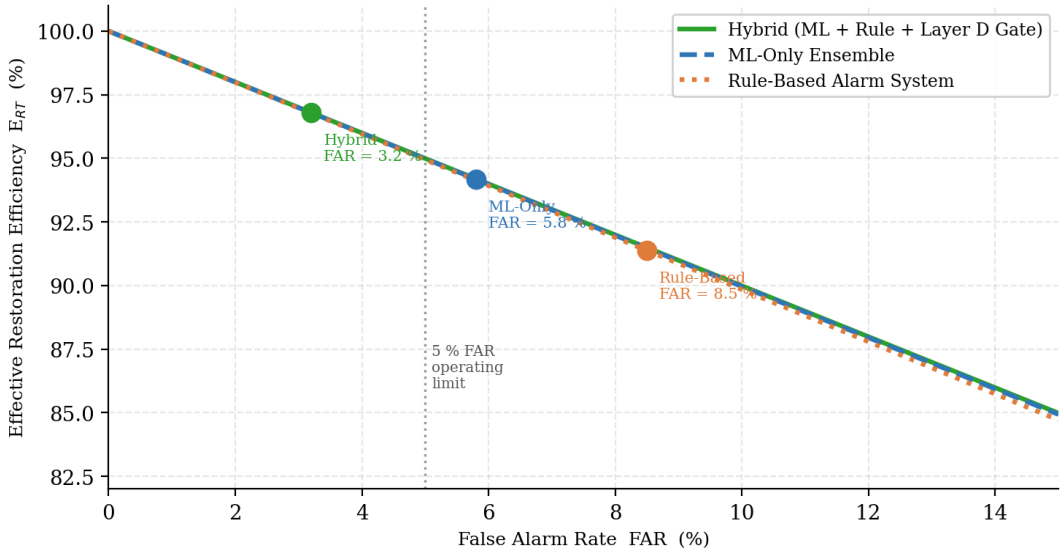


Figure 4. Effective Restoration Efficiency (E_RT) vs. FAR for Three Detection Architectures

Monte Carlo ODI Distributions

Fig. 5 presents the empirical distribution of Operational Degradation Index (ODI) across all 10,000 Monte Carlo trials. The AI-integrated configuration achieves a mean ODI of 0.071 (σ = 0.009) compared to 0.180 (σ = 0.024) for traditional SCADA. This represents a 60.6% reduction along with a 62.5% decrease in variance.

Additionally, the 95th percentile ODI decreases from 0.219 to 0.087, confirming that AI integration reduces trial-risk operational events substantially.

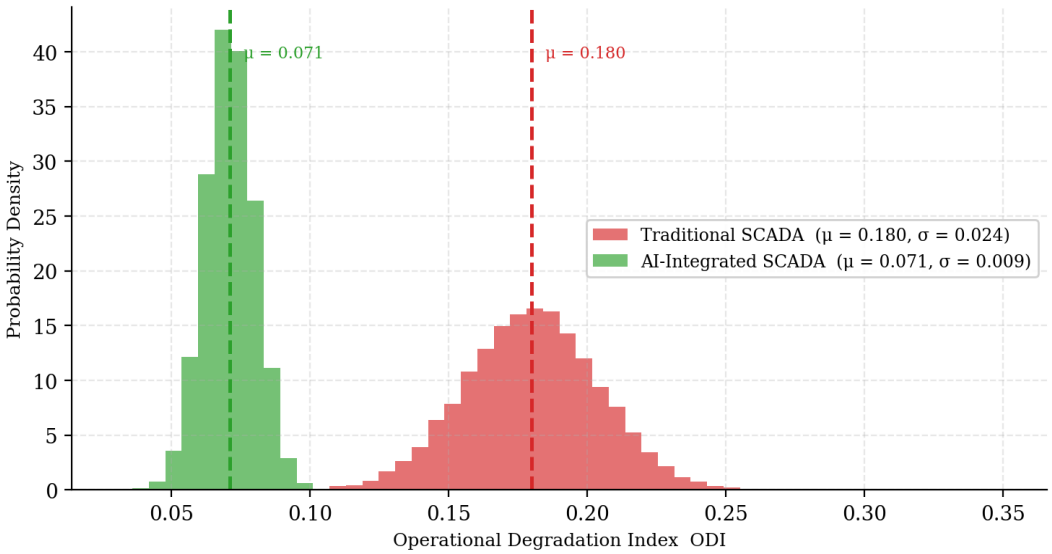


Figure 5. Monte Carlo ODI Probability Distribution (N = 10,000)

DRL Volt-VAR Optimization

Fig. 6 illustrates how the PPO DRL agent converges to a stable policy by episode 220, achieving mean episode reward +472. This indicates learning of an effective control strategy.

Table 3. summarizes DRL VVO performance. This approach reduces active power losses by 8.6% (from 142.3 kV to 130.1 kV) and improves voltage compliance from 88.7% to 94.3% of bus-hours. At the same time capacitor switching operation decreased by 39.1%, which helps reduce equipment wear. The minimum bus voltage improves from 0.913 to 0.938 p.u., directly addressing voltage quality caused by 25% DER penetration.

Overall, the results show that the DRL-based control improves both system efficiency and voltage stability while reducing operational stress on network components.

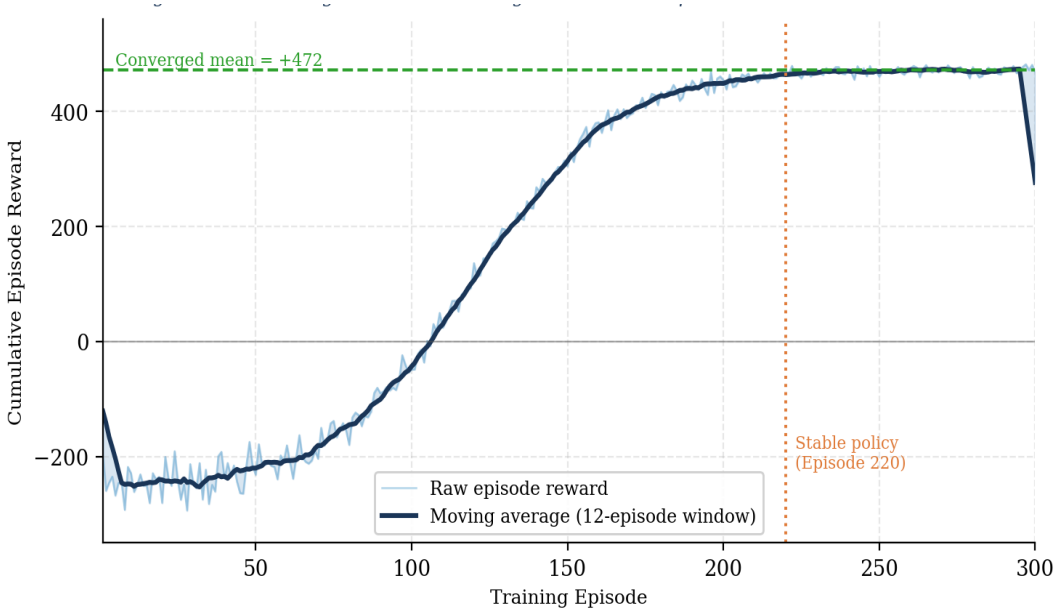


Figure 6. PPO DRL Agent Cumulative Reward Convergence (300 Episodes, IEEE 33-Bus Feeder)

Table 3. DRL Volt-VAR Optimization Performance: Rule-Based Baseline vs. PPO Agent

Metric	Rule-Based Baseline	DRL Agent (PPO)	Improvement
Active Power Loss (kW)	142.3 +/- 18.4	130.1 +/- 12.1	**8.6 %**
Nodal Voltage Compliance (%)	88.7 +/- 3.2	94.3 +/- 1.8	**+5.6 pp**
Capacitor Switching Ops/day	18.4 +/- 3.1	11.2 +/- 2.4	**-39.1 %**
Minimum Bus Voltage (p.u.)	0.913 +/- 0.011	0.938 +/- 0.007	**+0.025 p.u.**
Reactive Procurement Cost ($/day)	84.2 +/- 9.1	71.8 +/- 7.3	**14.7 %**

Layer D Safety Gate Validation

The deterministic Layer D safety gate is evaluated across all the 10,000 Monte Carlo switching trials. The system correctly validates AI-generated control actions in 99.97% of cases. The three incorrect vetoes (0.03%) are attributed to delayed RTU topology updates during a 4 second polling cycle. This issue can be eliminated by reducing the polling interval to 1 second. No unsafe switching operations are observed, confirming that the safety layer effectively enforces operational constraints, and ensures reliable execution of AI-driven actions.

In this context, a "veto" refers to the safety gate blocking an AI-generated switching command that would otherwise be executed. The three incorrect vetoes identified in the study are conservative false negatives-that is, they prevent potentially safe actions rather than allowing unsafe operations-so they are operationally acceptable within the safety-first design philosophy.

SENSITIVITY ANALYSIS

Table 4 quantifies the sensitivity of key performance metrics to increasing levels of DER penetration. As DER penetration rises from 0% to 40%, the false alarm rate increases from 2.8% to 4.9%. This trend is primarily driven by inverter ramp transients, which introduce variability that challenges the LSTM-based anomaly detection threshold.

Voltage compliance also degrades from 97.2% to 88.4% in the absence of continuous DRL-based Volt-VAR reconfiguration, highlighting the importance of adaptive control under high DER conditions. Despite the increased variability, improvements in fault detection latency (FDL), and restoration (RT) remain relatively stable up to approximately 35% DER penetration, indicating robustness of the proposed architecture under moderate system complexity.

Communication reliability further impacts system performance. At 10% packet loss, fault detection latency increases to 792ms, while still maintaining a 47.2% improvement over the traditional SCADA baseline. At 15% packet loss, the system transitions into a graceful degradation mode, reverting to deterministic rule-based detection while preserving SCADA supervisory continuity.

These results demonstrate that the proposed framework maintains strong performance across varying operating conditions and degrades in a controlled and predictable manner when system constraints are exceeded.

Table 4. Performance Metric Sensitivity to DER Penetration Level

Metric	0 % DER	15 % DER	25 % DER	40 % DER
FDL Improvement (%)	58.4	57.1	56.7	53.2
RT Improvement (%)	61.8	60.5	60.0	55.1
FAR (%)	2.8	3.0	3.2	4.9
Voltage Compliance (%)	97.2	95.8	94.3	88.4

DISCUSSION

The results of this study demonstrate that the performance improvements achieved through AI integration are primarily driven by system-level architectural design rather than model complexity alone. While the individual models used such as LSTM-based anomaly detection, gradient boosting classifiers, and PPO-based reinforcement learning– are well established, their coordinated deployment across SCADA layers enables consistent gains. In particular, the proposed framework achieved a 56.7% reduction in fault detection latency and 60.0% reduction in restoration time, illustrating the impact of combining edge-based inference with fast communication and automated control.

A key observation is the importance of inference placement, and latency management. Deploying AI at the substation edge (Layer B) allows the system to maintain more than 80% improvement in detection performance when total system latency remains below 250 ms. As latency increases, particularly due to distributed communication delays across the system. In practice, utilities operating legacy infrastructure may experience additional latency due to protocol translation (e.g., Modbus or proprietary systems), which can limit the effectiveness of edge-based AI deployment. The results also emphasize the necessity of hybrid validation. Standalone machine learning models exhibit reduced reliability under increased system variability, with false alarm rates rising to approximately 5.8% at higher DER penetration levels. In contrast, the proposed hybrid approach maintains FAR at 3.2% below the 5% operational threshold required to sustain automation trust. The integration of rule-based logic and the deterministic Layer D safety gate ensures that AI-driven decisions remain both accurate and operationally safe, while also contributing to broader system-level improvements, including reduction of the operational degradation index from 0.180 to 0.071. From an implementation perspective, the study highlights practical constraints. The evaluation is conducted using a simulation-based framework implemented in MATLAB/Simulink and supported by large-scale Monte Carlo analysis (N= 10,000), providing a controlled and reproducible environment. For each trial, system states are generated, faults are randomly injected, and the corresponding detection and control responses are evaluated. The AI-integrated framework and the traditional SCADA approach are executed in parallel to enable consistent performance comparison across all scenarios.

However, real-world validation remains constrained by limited access to operational utility data due to strict privacy, cybersecurity and regulatory requirements governing critical infrastructure systems. Investor-owned utilities, including organizations such as NV Energy and Pacific Gas & Electric (PG&E), operate under stringent compliance frameworks that restrict data sharing and limit direct experimentation on live systems. In addition, full-scale testing of automated control strategies is constrained by safety risks, infrastructure dependencies, and the operational implication of real-time switching actions.

These constraints justify the use of simulation-based evaluation while also defining clear directions for future work, including validation with real operational datasets where accessible, extension to larger network configurations, and hardware-in-the-loop testing to assess real-time deployment feasibility.

CONCLUSION

This paper presents a system-level AI-SCADA integration framework for power distribution networks and evaluates its performance through simulation-based validation. The proposed architecture integrates edge-based AI inference, centralized optimization, and a deterministic safety validation layer to support coordinated, real-time decision-making within operational constraints. Results based on Monte Carlo simulation (N= 10,000) using the IEEE 33-bus feeder demonstrate substantial improvements over traditional SCADA systems, including a 56.7% reduction in fault detection latency, a 60.0% reduction in restoration time, and a 62.4% reduction in false alarm rate, along with a decrease in the operational degradation index from 0.180 to 0.071. In addition, the DRL-based Volt-VAR optimization achieves an 8.6% reduction in active power losses and improves voltage compliance to 94.3%, while the Layer D safety mechanism ensures safe execution with no unsafe switching events observed. These findings highlight that effective AI integration depends not only on model performance, but also on architectural design, latency management, and validation mechanisms.

The study is limited by its reliance on simulation and by restricted access to real-world operational datasets due to privacy, regulatory, and infrastructure constraints. Future work will focus on validation using real operational data where feasible, extension to larger network configurations, and incorporation of hardware-in-the-loop testing to support practical deployment.

REFERENCES

Alhamrouni, I., et al. (2024). A comprehensive review on the role of artificial intelligence in power system stability, control, and protection: Insights and future directions. *Applied Sciences, 14*(14), 6214. https://doi.org/10.3390/app14146214

Chen, Y., Heleno, M., Jiang, T., & Xie, L. (2022). Evaluating distribution system reliability with hyperstructures graph convolutional nets. In 2022 IEEE International Conference on Big Data (Big Data) (pp. 1793-1800). IEEE. https://ieeexplore.ieee.org/document/10020715

Chung, S. H., & Zhang, Y. (2023). Artificial intelligence applications in electric distribution systems: Post-pandemic progress and prospect. *Applied Sciences, 13*(12), 6937. https://doi.org/10.3390/app13126937

Dolatabadi, S. H., Ghorbanian, M., Siano, P., & Hatziargyriou, N. D. (2021). An enhanced IEEE 33-bus benchmark test system for distribution system studies. *IEEE Transactions on Power Systems, 36*(3), 2565–2572. https://doi.org/10.1109/TPWRS.2020.3038030

Feng, C., Wang, Y., Chen, Q., Ding, Y., Strbac, G., & Kang, C. (2021). Smart grid encounters edge computing: Opportunities and applications. *Advances in Applied Energy, 1*, 100006. https://doi.org/10.1016/j.adapen.2020.100006

Gültekin, Ö., Çinar, E., Özkan, K., & Yazıcı, A. (2022). Real-time fault detection and condition monitoring for industrial autonomous transfer vehicles utilizing edge artificial intelligence. *Sensors, 22*(9), 3208. https://doi.org/10.3390/s22093208

Hoffmann, M. W., Wildermuth, S., Gitzel, R., et al. (2020). Integration of novel sensors and machine learning for predictive maintenance in medium voltage switchgear to enable the energy and mobility revolutions. *Sensors, 20*(7), 2099. https://doi.org/10.3390/s20072099

Hosseini, M. M., & Parvania, M. (2021). Artificial Intelligence for resilience enhancement of distribution systems. *The Electricity Journal, 34*(7), 106880. https://doi.org/10.1016/j.tej.2020.106880

Liang, S., Jin, S., & Chen, Y. (2024). A Review of Edge Computing Technology and Its Applications in Power Systems. Oak Ridge National Laboratory (ORNL). https://www.osti.gov/servlets/purl/2479469

Matin, M., Dehghanian, A., Dastranj, M., & Darijani, H. (2024). Explainable artificial intelligence modeling of internal arc in a medium voltage switchgear based on different CFD simulations. *Heliyon, 10*(8), e29594. https://doi.org/10.1016/j.heliyon.2024.e29594

Nasim, S. F., Aziz, S., Qaiser, A., Umm-e-Kulsoom, & Ahmed, S. (2024). *Fault Detection and Fault Diagnosis in Power System Using AI: A Review*. Sir Syed University Research Journal. https://sirsyeduniversity.edu.pk/ssurj/rj/index.php/ssurj/article/view/598

Prasad, S., & Devagnanam, J. (2024). Advanced distribution system optimization: Utilizing flexible power buses and network reconfiguration. *Applied Sciences, 14*(22), 10635. https://doi.org/10.3390/app142210635

Rambabu, B., et al. (2026). AI-powered predictive maintenance framework for industrial SCADA systems with cloud-edge orchestration. IEEE. https://ieeexplore.ieee.org/document/11291518

Rabaia, M. K. H., Semeraro, C., Soudan, B., Salameh, T. S. Z., Abdelkareem, M. A., & Olabi, A. G. (2025). *Digital twin-driven sustainable energy life cycles: Technical review and guidelines*. Energy Nexus, 19, 100482. https://www.sciencedirect.com/science/article/pii/S2772427125001238

Septian, F. A., Oktavia, D. Z., Athooyaa, M., Firmansyah, M. P., Siwi, S. Q. R., & Aprilianto, R. A. (2025). *Artificial Intelligence for Fault Detection in Smart Grid: A Systematic Literature Review*. IEEE. https://ieeexplore.ieee.org/document/11472940

Yang, Q., Wang, G., Sadeghi A., Giannakis, B. G., & Sun, J. (2020). Two-timescale voltage control using deep reinforcement learning. *IEEE Transactions on Smart Grid, 11*(3), 2313–2323. https://ieeexplore.ieee.org/document/8892476

Zhang, W., Li, Z., Li, W., Ma, W., Chen, X., & Li, S. (2023). Robust Learning Based condition diagnosis method for distribution network switchgear. *arXiv*. https://arxiv.org/abs/2311.07956

Zhang, X., Wu, Z., Gu, W., Xu, Z., & Zheng, S. (2024). Application and progress of artificial intelligence technology in the field of distribution network voltage control: A review. *Renewable and Sustainable Energy Reviews, 192*, 114005. https://doi.org/10.1016/j.rser.2024.114282

Zhang, Y., Wang, X., Wang, J., & Zhang, Y. (2021). Deep reinforcement learning-based Volt-VAR optimization in smart distribution systems. *IEEE Transactions on Smart Grid, 12*(1), 361–371. https://ieeexplore.ieee.org/document/9143169

MANAGING AI-DRIVEN CYBERSECURITY RISKS IN CRITICAL INFRASTRUCTURE

Stephen Bonner, Marymount University, deepseasb@yahoo.com
Devontee Lawson, Marymount University, d_lawson@me.com
Diamond Webb, Marymount University, webbd879@gmail.com

ABSTRACT

The growing use of artificial intelligence (AI) in cybersecurity has significantly changed how organizations detect and respond to threats, especially within critical infrastructure environments. AI-driven tools are commonly used to process large volumes of data and identify suspicious activity; however, their reliability can be affected by challenges such as false positives, false negatives, limited transparency, and changing system behavior. These issues create meaningful risks in high-stakes environments where system failures can disrupt essential operations. This paper explores how AI can be incorporated into existing risk management frameworks to strengthen system reliability while supporting transparency, accountability, and effective governance. Using a framework-based analytical approach, this study examines how risk management principles apply to AI systems, with a focus on the National Institute of Standards and Technology Artificial Intelligence Risk Management Framework. The analysis identifies gaps in traditional approaches when applied to AI and proposes targeted improvements to address these limitations. The findings show that improving AI reliability in critical infrastructure requires more than technical performance alone and depends on governance, oversight, and continuous evaluation across the AI lifecycle.

Keywords: artificial intelligence, cybersecurity, threat detection, critical infrastructure, risk management, NIST AI RMF

INTRODUCTION

Artificial intelligence (AI) is playing a growing role in cybersecurity, particularly in how organizations detect and respond to potential threats. Many security tools now use AI to review large amounts of data, recognize patterns, and identify unusual activity. One common approach is anomaly detection, where the system compares current activity to what is considered normal and flags anything that looks different (Buczak & Guven, 2016). Because of these capabilities, organizations are increasingly relying on AI to support their security operations, especially in environments that require constant monitoring. At the same time, cyber threats are becoming more frequent and more complex, which makes the need for reliable detection systems even more important (ENISA, 2023).

AI adoption is especially noticeable in critical infrastructure environments. Critical infrastructure includes systems that are essential to everyday life, such as energy, transportation, water, and communication systems (Cybersecurity and Infrastructure Security Agency [CISA], n.d.). These environments depend on technologies like industrial control systems (ICS) and supervisory control and data acquisition (SCADA) systems to manage and maintain daily operations (CISA, n.d.-b).

As these systems become more connected to internal networks and the internet, they face a higher risk of cyber threats (CISA, n.d.-b).

While AI offers clear advantages, it also introduces challenges that can affect system reliability. AI-driven threat detection systems are not always accurate and can produce errors, which can have significant consequences in intrusion detection environments (Sommer & Paxson, 2010). In addition, many AI systems lack transparency, which can make it difficult for organizations to fully understand how decisions are made, highlighting the importance of explainability and interpretability in AI systems (National Institute of Standards and Technology [NIST], 2023). These challenges are especially concerning in critical infrastructure, where system failures can have significant operational, safety, and security impacts (CISA, n.d.-a). Because of these challenges, it is necessary to take a closer look at how AI systems are managed within existing risk management frameworks. This paper focuses on how artificial intelligence can be incorporated into these frameworks to improve the reliability of AI-driven threat detection systems while still ensuring transparency, accountability, and effective governance. In doing so, the analysis highlights gaps in current approaches and offers practical ways to strengthen the use of AI in high-risk environments.

This paper argues that the central challenge is not the use of artificial intelligence itself, but the misalignment between AI-driven systems and traditional risk management frameworks, which are not designed to address dynamic, adaptive, and data-driven technologies.

BACKGROUND AND LITERATURE REVIEW

AI Is Used in Cybersecurity Systems

Artificial intelligence is applied in cybersecurity primarily through AI-driven threat detection systems that are designed to support existing security operations. These systems are commonly integrated into tools such as intrusion detection systems (IDS) and security information and event management (SIEM) platforms, where they assist in identifying patterns and detecting anomalies within large datasets (Buczak & Guven, 2016). Rather than functioning as standalone solutions, AI models operate as part of a broader security architecture, helping analysts interpret complex data and identify potential threats more efficiently.

AI Systems Process and Analyze Data

To support these functions, AI-driven threat detection systems operate by examining activity across a network and identifying behavior that does not align with expected patterns. These systems pull information from different parts of the environment, including user activity, system processes, and network traffic, to build a picture of how the system normally operates. When activity falls outside of that expected range, it is flagged for further analysis. This approach allows organizations to detect potential threats that may not match known attack signatures, which is particularly important in environments where attackers attempt to blend in with legitimate activity (Buczak & Guven, 2016).

Why Organizations Are Adopting AI

Because of these capabilities, organizations are increasingly adopting AI in response to the growing scale and complexity of cybersecurity data (ENISA, 2023). Modern systems generate a

continuous flow of activity, making it difficult for traditional rule-based tools to effectively identify and prioritize potential threats (ENISA, 2023). AI helps address this challenge by enabling faster analysis and more targeted identification of suspicious behavior. As cyber threats continue to evolve and become more difficult to detect, organizations are using AI to strengthen their ability to respond in a timely and effective manner (ENISA, 2023). However, adoption is not uniform across all sectors. Industries that require continuous system availability, such as energy and transportation, are more likely to implement these technologies, while others may adopt them more gradually based on available resources and operational needs.

How AI is Affecting Critical Infrastructure

This growing adoption is especially significant in critical infrastructure environments, where maintaining system availability and security is essential (CISA, n.d.-a). These systems require continuous monitoring to prevent disruptions to services that support essential societal functions. As a result, AI technologies are increasingly used to support real-time analysis and rapid response capabilities in these high-risk environments (CISA, n.d.-a). This growing reliance on AI highlights the importance of ensuring that these systems operate reliably within critical infrastructure settings.
How AI Is Shifting Cybersecurity Strategy and the Need for Risk Management

As AI evolves, it is changing how organizations approach threat detection rather than simply improving existing methods. In many cases, security teams are no longer relying only on predefined indicators but are instead paying closer attention to activity that appears inconsistent with normal operations. This changes how threats are identified, especially when attackers attempt to avoid detection by mimicking legitimate behavior (Buczak & Guven, 2016). At the same time, depending more heavily on these systems introduces new challenges. Organizations must be able to understand how decisions are made and ensure that the results can be trusted, which makes oversight and structured risk management an important part of using AI effectively (NIST, 2023). These operational changes provide important context for understanding how failures in AI systems can introduce cybersecurity risk. While these capabilities improve detection, they also introduce risks that are not always visible through traditional security models. These risks become more apparent when examining how AI systems fail in real-world environments.

METHODOLOGY

This study uses a framework-based analytical approach to examine how existing risk management principles apply to AI-driven cybersecurity systems within critical infrastructure environments. The analysis draws upon current literature, AI governance frameworks, and cybersecurity risk management standards, including the NIST AI Risk Management and ISO/IEC 42001, to evaluate limitations in traditional risk management approaches and identify areas requiring adaptation for AI-enabled systems.

RISK SCENARIO OR INCIDENT ANALYSIS

Scenario Overview: AI-Driven Threat Detection in Critical Infrastructure

In a critical infrastructure environment within the energy sector, an organization relies on an AI-driven threat detection system to monitor network activity across its operational and information technology systems. The system is integrated into the organization's security infrastructure and is responsible for identifying potential cyber threats by analyzing network traffic, system logs, and

user behavior. Due to the importance of maintaining continuous operations, the organization depends on this system to detect threats in real time and support rapid response efforts.

Incident Description: Failure to Detect a Malicious Intrusion

In this scenario, a threat actor gains unauthorized access to the network by using techniques that closely resemble normal system behavior. Because the AI system is trained to identify anomalies based on deviations from expected patterns, the activity is not flagged as suspicious. As a result, the system fails to detect the intrusion, allowing the attacker to move laterally within the network and compromise critical systems. This represents a false negative, where a real threat goes undetected.

ANALYSIS OF SYSTEM

Regulatory and Policy Environment

To address the types of failures identified in the scenario, organizations must rely on governance, regulatory, and policy frameworks that support the secure, transparent, and accountable use of AI systems (Chhatre, 2024). These frameworks seek to address not only cybersecurity concerns but also issues related to explainability, bias, and accountability in AI decision-making. As AI adoption accelerates within critical infrastructure environments, these concerns become increasingly significant due to the potential operational and societal impacts of system failure. This section examines the evolving regulatory and policy environment shaping AI risk management, with particular emphasis on its role in securing AI-driven cybersecurity systems within critical infrastructure.

Foundations of AI Governance and Risk Management

Regulation of AI begins with a strong governance foundation that defines what constitutes trustworthy AI (Papagiannidis, 2025). Trustworthy AI systems must be secure, reliable, transparent, and free from bias. These characteristics are particularly critical in cybersecurity applications within critical infrastructure, where system failures can have significant operational disruptions and societal consequences.

One of the most prominent challenges to using AI is overcoming algorithmic bias (Heldman, 2025). Algorithmic bias occurs when an AI system produces outputs that unfairly favor or disadvantage certain data patterns or conditions. Some causes of the biased output are historical biases in the AI system training data, skewed datasets, poorly defined queries, and limited transparency in how AI systems process data and generate outputs. To counter the high potential for algorithmic bias, AI developers must conduct bias audits and testing, ensure datasets are inclusive, design systems with transparency to identify problematic behaviors, and follow industry guidance such as NIST AI 100-1 and ISO/IEC 42001:2023.

Effective governance of AI also needs to understand and address the unique and inherent security risks and vulnerabilities of AI systems that can be exploited by malicious attackers (Yang et al., 2025). For example, data poisoning occurs when an attacker manipulates the datasets to alter or degrade AI performance. Model inversion attacks refer to exploiting the AI model to reverse-engineer the AI system processes and reconstruct the data used. Attackers can then either manipulate the AI model, harvest model secrets, or steal sensitive data. These two examples

represent only a portion of the potential attack vectors targeting AI systems. To mitigate the risks and vulnerabilities, AI developers must embed security into the design of AI systems rather than treating it as a post-deployment concern addressed through reactive updates and patches.

To adequately address AI risks, governance models must emphasize lifecycle-based risk management, where AI systems are continuously evaluated from design through deployment and operation. The Trustworthy AI Maturity Model (TAIMM) is a lifecycle-based structured framework that integrates fairness, transparency, accountability, and robust security measures into a phased framework that enables organizations to transition from ad hoc AI usage to AIgoverned systems. Effective AI governance using industry frameworks and lifecycle-based risk management is essential to securing our nation's critical infrastructure systems.

NIST AI Risk Management Framework (AI RMF)

From a risk management standpoint, the AI Risk Management Framework (AI RMF) is useful because it provides a structured way for organizations to make informed decisions in environments where reliability and safety are critical, such as critical infrastructure. Developed by the National Institute of Standards and Technology (NIST), the framework offers guidance for identifying and managing AI-related risks throughout the system lifecycle (NIST, 2023). It is organized around four key functions: govern, map, measure, and manage.

The govern function focuses on establishing policies, roles, and oversight mechanisms that support responsible AI use within an organization. The map function involves understanding how the AI system operates, including its intended purpose, dependencies, and potential risks.

The measure function assesses identified risks by evaluating their likelihood and potential impact, including issues related to system performance, security, and reliability. Finally, the manage function emphasizes implementing and adjusting controls to reduce or mitigate these risks over time. Although the NIST AI RMF provides a strong foundation for AI governance, its voluntary nature means that additional standards and regulatory approaches may be needed to ensure consistent adoption. As illustrated in the earlier scenario, the failure to detect malicious activity reflects weaknesses in how risks were identified, assessed, and managed, reinforcing the importance of applying the framework in a continuous and practical manner to improve system reliability.

International Standards and Industry Frameworks

ISO/IEC 42001:2023 establishes an international standard for managing artificial intelligence systems, with a focus on governance, transparency, and accountability. Rather than prescribing technical controls alone, the standard emphasizes responsible AI practices that support fairness, reduce bias, and align system behavior with organizational and societal expectations. In this way, both ISO/IEC 42001 and the NIST AI Risk Management Framework provide complementary guidance for strengthening how organizations manage AI-related risk and promote responsible adoption of these technologies.

In addition to governance-focused standards, secure development practices also play an important role in managing AI risk. The Certified Secure Software Lifecycle Professional (CSSLP)

framework outlines security principles that should be applied throughout the entire software development lifecycle. These principles include secure design, testing, and deployment practices that help reduce vulnerabilities before systems are placed into operation (Brecht, 2025). Similarly, guidance from the United Kingdom's National Cyber Security Centre highlights key considerations for building secure AI systems, including protecting training data, safeguarding model integrity, and managing system dependencies (Guidelines for Secure AI System Development, n.d.). Together, these frameworks reinforce the importance of integrating security and governance throughout the AI lifecycle and can inform future regulatory and compliance efforts.

Emerging AI Laws

When it comes to the regulation of AI through actual laws, the European Union's Artificial Intelligence Act (EU, 2024) represents the first comprehensive legal framework governing AI systems. The EU AI Act seeks to foster the safe development and usage of trustworthy AI systems and technologies while simultaneously respecting the fundamental rights and privacy of EU citizens. Although this legislation is only enforceable within the confines and borders of EU law, this ground-breaking legislation sets the global standard for AI regulation for other countries to emulate. In contrast, the United States has adopted a decentralized, sectorbased approach, which prioritizes flexibility but may result in inconsistent implementation across industries.

Regulatory Gaps and Challenges

The primary gap in AI regulation is the lack of a unified global framework (Zemello, 2023), which creates inconsistencies in governance and compliance across jurisdictions. Additionally, voluntary frameworks such as the NIST AI RMF, while valuable, lack enforcement mechanisms that ensure compliance. Technical challenges further complicate regulation. AI systems often lack transparency, making it difficult to explain decision-making processes (Papagiannidis, 2025). This lack of explainability hinders accountability and complicates regulatory oversight. Furthermore, the rapid pace of AI innovation often outpaces regulatory development, leaving organizations to manage AI risks in an uncertain policy environment (Gordon, 2024).

Regulating AI Through Co-Governance

Existing frameworks, standards, and regulatory approaches demonstrate significant progress in AI governance (Zemello, 2023). Although the European Union's Artificial

Intelligence Act (EU, 2024) is a positive step in the right direction, it only applies to the citizens and territories of the EU. This fragmentation underscores the need for a novel governance model capable of addressing the unique nature of AI systems.

Regulating artificial intelligence through Co-Governance (Co-Governance, 2025) represents a shift from traditional top-down regulatory approaches toward a more collaborative and inclusive model. Rather than relying solely on government agencies or dominant industry players, Co-Governance incorporates input from diverse stakeholders, including academia, civil society, and the private sector, to shape AI policy and regulation. This collaborative approach helps ensure that regulatory frameworks reflect broader societal interests rather than narrow commercial priorities. Co-governance also mitigates the risk of regulatory capture by preventing large technology firms

from disproportionately influencing the rules governing AI systems. Additionally, it promotes innovation by supporting open participation. Ultimately, framework based governance, lifecycle management, and collaborative AI regulation are essential to effectively managing the cybersecurity risk in critical infrastructure environments.

RESULTS/FINDINGS

Traditional Enterprise Risk Management (ERM) frameworks are often misaligned with the operational realities of modern AI-enabled systems, especially in our nation's critical infrastructure (Donohue, 2026). The limitations identified in this section directly explain the failure observed in the earlier scenario, where traditional risk management approaches were not sufficient to detect or respond to adaptive threats. As demonstrated in the earlier scenario involving a failure to detect an intrusion in an energy-sector environment, traditional RMFs struggle to address the dynamic and adaptive nature of AI-driven systems and processes. Traditional ERM-based RMFs provide structured approaches to managing cybersecurity risk that are largely rooted in assumptions that are no longer relevant to modern digitally interconnected ecosystems.

Limitation 1: Outdated Systems and Processes

One of the most significant limitations of traditional RMFs is their reliance on outdated hardware and software systems (Donohue, 2026; Igit et al., 2024). Our critical infrastructure is largely reliant on legacy systems that were designed for relatively static, well-defined and predictive IT environments, where system configurations and threat vectors gradually change over time. However, critical infrastructure systems have undergone rapid digital transformation, integrating emerging technologies such as AI, machine learning, and Internet of Things (IoT) devices (Radanliev et al., 2019; Zeijlemaker et al., 2025). Integrating legacy systems into digitally interconnected ecosystems has significantly expanded our critical infrastructure's attack surface. Consequently, the result is increased system interdependencies and vulnerabilities that are no longer isolated but can now propagate across the interconnected infrastructure, amplifying the risk.

Limitation 2: Reactive and Siloed Risk Management Approaches

A second critical limitation is that traditional RMFs operate within a reactive, siloed, and static risk management paradigm (Faruk et al., 2025). Traditional risk assessments typically rely heavily on historical data and expert judgment. This approach fails to capture the evolutionary switch to digital ecosystems and the rapidly evolving nature of the cyberthreat landscape. Contemporary cyberthreats are dynamic and adaptive, with adversaries continuously modifying tactics, techniques, and procedures to defeat current countermeasures guarding our critical infrastructure (Alageel & Maffeis, 2025). AI-driven systems, in particular, introduce unique risk dynamics, where feedback loops and system interactions can amplify vulnerabilities over time in addition to unique attack vectors such as data poisoning and model inversion attacks (Yang et al., 2025).

Traditional risk management frameworks also tend to assess risks separately rather than considering how systems, external partners, and organizational processes are interconnected (Faruk et al., 2025). This fragmented approach can reduce overall visibility and make it more difficult to identify new or evolving threats as they develop. In contrast, AI-driven risk management approaches focus on continuous monitoring, data-driven insights, and the ability to

adjust to changing conditions in real time. This shift highlights the limitations of traditional models, which rely heavily on static assessments and are often slow to respond to emerging risks.

Limitation 3: AI-specific Threat Vectors

A third limitation is the inability of traditional RMFs to adequately address AI-specific threat vectors (Zeijlemaker et al., 2025). Traditional risk frameworks focus primarily on known cybersecurity risks, such as network intrusions, software vulnerabilities, and misconfigurations. However, AI systems introduce a new class of threats that operate at the data and model levels. These include data poisoning attacks, adversarial inputs, model inversion, and model drift, all of which can compromise the integrity and reliability of AI-driven decision-making systems (Yang et al., 2025). Such threats are particularly concerning in critical infrastructure environments, where AI systems are increasingly used for anomaly detection and automated response (Ali et al., 2025). While AI enhances detection capabilities, it simultaneously expands the attack surface and creates new opportunities for adversarial exploitation.

Perhaps the most compelling evidence of the cumulative effect of all these limitations is observed when directly examining the incident scenario presented earlier in this paper. In that case, an AI-driven anomaly detection system failed to identify malicious activity within a critical infrastructure environment, giving the threat actor unchecked access to critical infrastructure systems. This failure can be traced to several shortcomings inherent in traditional RMFs.

The failure can be linked to several limitations in how risk was assessed and managed. The system relied on static evaluations based on past data, which did not account for how attacker behavior can change over time (Donohue, 2026; Igit et al., 2024). In addition, the lack of continuous monitoring and ongoing validation meant that changes in model performance were not identified as the system evolved (Zeijlemaker et al., 2025). There were also no built-in controls to detect or respond to adversarial manipulation targeting the model itself (Yang et al., 2025). Together, these weaknesses show that the issue extends beyond a technical failure and reflects broader shortcomings in traditional risk management approaches when applied to AIdriven systems.

The Way Ahead

The gaps identified in this paper highlight the need to evolve traditional risk management frameworks to better address the reliability challenges associated with AI-driven cybersecurity systems. Traditional frameworks are static, reactive, and primarily focused on IT-centric risks, whereas AI systems operate in dynamic, adaptive, and data-driven environments that span both IT and OT domains. As cyber threats continue to evolve in complexity and scale, reliance on traditional risk management approaches will expose critical infrastructure to increased operational, financial, and societal risks. Future RMFs need to incorporate continuous monitoring, real-time risk assessment, lifecycle governance, and address the unique risks and vulnerabilities presented by AI-driven systems and processes.

DISCUSSION OF FINDINGS

Building on the identified limitations of traditional risk management frameworks, there is a clear need to evolve toward adaptive, AI-aware approaches that better address the dynamic and data-driven nature of AI-enabled cybersecurity systems. As demonstrated in the incident scenario, static

and reactive frameworks are insufficient for managing risks associated with model drift, adversarial manipulation, and system interdependencies. To improve effectiveness, several modifications and complementary approaches should be incorporated into existing frameworks such as the NIST AI RMF.

First, risk management frameworks should emphasize continuous monitoring and lifecycle-based validation as foundational components. AI systems require ongoing evaluation throughout design, deployment, and operational phases to ensure reliability and resilience. This approach aligns with the NIST AI RMF's lifecycle focus and is reinforced by the Trustworthy AI Maturity Model (TAIMM), which advocates for continuous oversight and iterative improvement of AI systems (National Institute of Standards and Technology [NIST], 2023; McCormack & Bendechache, 2026). Continuous validation mechanisms, including retraining, performance benchmarking, and drift detection, are essential for maintaining system effectiveness in evolving threat environments.

Second, frameworks should incorporate AI-specific threat modeling into risk assessment processes. Traditional approaches focus on network and software vulnerabilities but fail to address risks unique to AI systems, such as data poisoning, adversarial inputs, and model inversion attacks (Yang et al., 2025). Integrating these risks into existing frameworks enhances situational awareness and supports proactive mitigation strategies, particularly in critical infrastructure environments where system integrity is essential.

Third, the integration of human-in-the-loop and human-on-the-loop governance mechanisms is critical for managing AI-driven decision-making. AI systems should not operate as fully autonomous decision-makers in high-risk environments. Instead, human oversight should be embedded to validate outputs, interpret results, and intervene when necessary. This approach mitigates the risks associated with false negatives and misclassification, which have been identified as key limitations in anomaly-based detection systems (Sommer & Paxson, 2010).

Fourth, frameworks should require explainability and auditability to address the opacity of AI systems. Organizations must be able to understand how AI models generate outputs in order to support accountability, regulatory compliance, and effective incident response. Enhancing transparency aligns with broader AI governance principles and is supported by standards such as ISO/IEC 42001, which emphasize accountability and traceability in AI systems (International Organization for Standardization [ISO], 2023; Papagiannidis et al., 2025).

Fifth, a complementary extension involves integrating secure AI development practices into risk management frameworks. Drawing from secure software development lifecycle (SDLC) principles and frameworks such as CSSLP, organizations should embed security controls throughout the AI lifecycle, including data validation, access controls, adversarial testing, and supply chain risk management (Brecht, 2025). This proactive approach ensures that security is incorporated into system design rather than applied retroactively.

Finally, the adoption of co-governance models can enhance the effectiveness of AI risk management in critical infrastructure. AI-related risks often extend beyond individual

organizations and can impact broader societal systems. Co-governance frameworks promote collaboration among stakeholders, including government, industry, and academia, to ensure that AI systems are deployed responsibly and aligned with societal values (Co-Governance, 2025). Collectively, these modifications transform traditional risk management frameworks into dynamic, adaptive systems capable of addressing the unique challenges posed by AI-driven cybersecurity technologies.

IMPLICATIONS FOR RESEARCH AND PRACTICE

The adoption of AI-driven cybersecurity technologies is expected to follow a progressive and phased trajectory, with varying rates of implementation across sectors. As discussed in this paper, critical infrastructure sectors are among the early adopters due to their need for continuous monitoring and rapid threat detection capabilities. In the short term (1–3 years), AI is likely to remain primarily in decision-support roles, assisting analysts in anomaly detection and threat prioritization. In the medium term (3–7 years), organizations are expected to expand the use of AI into semi-autonomous systems capable of supporting response actions. Full autonomy in critical infrastructure environments will likely develop more slowly due to the high risks associated with system failure.

The maturity of AI technology significantly influences cybersecurity risk. In early stages of adoption, risks are largely driven by uncertainty, limited governance, and lack of standardization. Organizations may not yet have established processes for validating model performance, managing data integrity, or monitoring system behavior. This increases the likelihood of errors such as false positives and false negatives, as well as vulnerabilities to adversarial attacks. During this phase, governance strategies should prioritize controlled implementation, human oversight, and rigorous testing.

As AI systems mature, organizations gain improved understanding of system behavior, performance metrics, and associated risks. This enables more confident deployment and integration into cybersecurity operations. However, increased maturity also introduces dependency risk, as organizations begin to rely more heavily on AI systems for critical security functions. A failure or compromise of an AI system at this stage can have widespread consequences, potentially affecting multiple layers of the organization's security infrastructure.
Over time, governance considerations must evolve to reflect increasing reliance on AI technologies. Early-stage governance focuses on experimentation and adoption, while later stages require formalized governance structures, accountability mechanisms, and resilience planning. Frameworks such as the NIST AI RMF and ISO/IEC 42001 provide structured approaches to managing AI risk across its lifecycle and become increasingly important as organizations transition from exploratory use to operational dependence (NIST, 2023; ISO, 2023).

Additionally, regulatory developments will shape the adoption trajectory. Emerging policies, such as the EU's AI Act, are likely to influence global standards and encourage more consistent implementation of AI governance practices. However, fragmented regulatory environments may continue to present challenges for organizations operating across jurisdictions (European Union, 2024; Zemello, 2023).

The trajectory of AI adoption highlights the need for continuous alignment between technological capability and risk management practices. In early stages, the focus is on safe adoption and oversight. In later stages, the emphasis shifts to resilience, assurance, and managing systemic dependencies. Organizations must ensure that their governance frameworks evolve alongside technological maturity to effectively mitigate emerging risks.

Limitations

This paper is limited by its conceptual and framework-based approach and does not include empirical testing or organizational case study validation. Additionally, the analysis relies on publicly available literature and existing governance frameworks, which may evolve as AI technologies and regulatory environments continue to mature.

LIMITATIONS AND FAILURE

The failure of the AI system in this scenario highlights several key limitations associated with AI-driven threat detection systems (Sommer & Paxson, 2010). First, the system's reliance on historical data to define normal behavior can limit its ability to detect sophisticated attacks that are designed to mimic legitimate activity. Second, the lack of transparency in how the AI model evaluates and classifies behavior makes it difficult for analysts to understand why the activity was not flagged (NIST, 2023). Additionally, the system's inability to adapt quickly to evolving attack techniques further reduces its effectiveness in dynamic threat environments (Sommer & Paxson, 2010). Taken together, these limitations show that the failure is not only technical but also reflects gaps in how AI systems are governed, validated, and monitored within existing risk management frameworks.

Risk Implications and Impact on Critical Infrastructure

The consequences of this failure are significant, particularly within a critical infrastructure environment, where disruptions can impact essential services (CISA, n.d.-a). From an organizational perspective, this type of failure can result in financial losses, reputational damage, and reduced trust in automated security systems. Operationally, the disruption of critical systems may lead to downtime, loss of control over infrastructure processes, and delays in response efforts. From a societal perspective, failures in critical infrastructure can affect essential services, potentially impacting public safety, economic stability, and community well-being (ENISA, 2023). This scenario demonstrates that failures in AI-driven threat detection are not isolated technical issues, but indicators of deeper misalignment between system capabilities and risk management practices. Addressing this gap requires frameworks that account for the dynamic and adaptive nature of AI systems.

CONCLUSION

The integration of AI into cybersecurity, particularly within critical infrastructure environments, represents a significant advancement in threat detection capabilities. However, as this paper has demonstrated, the adoption of AI also introduces complex and evolving risks that challenge traditional risk management frameworks. Issues such as false positives, false negatives, lack of transparency, model drift, and AI-specific attack vectors highlight the limitations of static, reactive approaches to risk management.

Addressing these challenges requires organizations to move beyond traditional approaches and adopt more flexible, lifecycle-oriented risk management strategies. These approaches should include ongoing monitoring, consideration of risks unique to AI systems, human involvement in decision-making, and stronger governance practices. While frameworks such as the NIST AI Risk Management Framework and ISO/IEC 42001 offer a useful starting point, they are not sufficient on their own and need to be supported by additional methods that reflect how AI systems actually operate. Incorporating secure development practices and collaborative governance models can further improve how organizations manage risk in complex and interconnected environments.

The use of AI in cybersecurity is expected to expand as threats become more advanced and difficult to manage using traditional methods. As shown in the scenario discussed earlier, failures in AI-based detection are not only technical breakdowns but also signs of weaknesses in oversight and risk management. As organizations depend more on these systems, the consequences of failure become more significant, reinforcing the need for stronger alignment between technology and governance.

In the end, effectively using AI in cybersecurity requires organizations to balance innovation with accountability. Adapting risk management practices to better reflect the dynamic nature of AI can improve system reliability, increase transparency, and strengthen overall resilience, particularly in critical infrastructure environments.

REFERENCES

Alageel, A. & Maffeis, S. (2025, February 12). *Investigation of Advanced Persistent Threats Network-based Tactics, Techniques and Procedures. Cornell University*. https://arxiv.org/pdf/2502.08830

Ali, S.M., Razzaque, A., Yousaf, M., Ali, S.S. (2025, January 22). A Novel AI-Based Integrated Cybersecurity Risk Assessment Framework and Resilience of National Critical Infrastructure. *IEEE Access, 13*. https://ieeexplore.ieee.org/abstract/document/10819345

Birogul, S., Sahin, O., & Esgerli, H. (2025, June 16). Exploring the Impact of ISO/IEC 42001:2023 AI Management Standard on Organizational Practices. *Advances in Artificial Intelligence Research, 5*(1), 14-22. https://dergipark.org.tr/en/download/articlefile/4916707

Brecht, D. (2025, January 28). *ISC2 CSSLP certification overview: What you need to know* [blog]. INFOSEC Institute. https://www.infosecinstitute.com/resources/csslp/csslpcertification-an-overview/

Buczak, A.L., & Guven, E. (2016). Machine learning for cybersecurity intrusion detection: Advances and challenges. *Computers & Security, 62*, 130-149. https://doi.org/10.1016/j.cose.2016.07.002

Chhatre, R. (2024, May). Policy and Regulatory Frameworks for Artificial Intelligence. *Journal of Harbin Engineering, 45*(5). https://www.researchgate.net/publication/380987226

Co-Governance and the Future of AI Regulation. (2025, April). Harvard Law Review, 138(6). https://harvardlawreview.org/print/vol-138/co-governance-and-the-future-of-airegulation/

Cybersecurity and Infrastructure Security Agency. (n.d.-a). *Critical infrastructure security and resilience*. https://www.cisa.gov/topics/critical-infrastructure-security-and-resilience

Cybersecurity and Infrastructure Security Agency (n.d.-b). *Industrial control systems*. https://www.cisa.gov/topics/industrial-control-systems

Donohue, J. (2026, January 14). *Enterprise risk management vs. traditional risk management: Which one is best for you?* [blog]. Diligent. https://www.diligent.com/resources/blog/enterprise-risk-management-vs-traditional-riskmanagement

European Council (EU). (2024, May 21). *Artificial intelligence (AI) act: Council gives final green light to the first worldwide rules on AI*. Council of the European Union. https://www.consilium.europa.eu/en/press/press-releases/2024/05/21/artificialintelligence-ai-act-council-gives-final-green-light-to-the-first-worldwide-rules-on-ai/

European Union Agency for Cybersecurity (ENISA). (2023*). ENISA threat landscape 2023*. https://www.enisa.europa.eu/publications/enisa-threat-landscape-2023

Faruk, M., Plabon, F., Saha, U., & Hossain, M. (2025, June 12). AI-Driven Project Risk Management: Leveraging Artificial Intelligence to Predict, Mitigate, and Manage Project Risks in Critical Infrastructure and National Security Projects. *Journal of Computer Science and Technology Studies*. https://alkindipublishers.org/index.php/jcsts/article/view/9949/8642

Gordon, J. (2024, October 23). *Addressing AI in Industrial Cybersecurity: A Strategic Imperative*. Industrial Cyber. https://industrialcyber.co/expert/addressing-ai-in-industrialcybersecurity-a-strategic-imperative/

Guidelines for Secure AI System Development. (n.d.). UK National Cyber Security Centre. https://www.ic3.gov/CSA/2023/231128.pdf

Heldman, K. (2025, January 20). *Understanding Algorithmic Discrimination: How Bias Persists in AI Systems* [blog]. Workplace Fairness. https://www.workplacefairness.org/understanding-algorithmic-discrimination-how-biaspersists-in-ai-systems/

Igit Y., Ferrag, M.A., Sarker, I., Maglaras, L., Chrysoulas, C., Moradpoor, N., & Janicke, H. (2024, May 08). *Critical Infrastructure Protection: Generative AI, Challenges, and Opportunities*. Cornell University. https://arxiv.org/pdf/2405.04874

ISO 42001:2023. (2023). *Information Technology-Artificial Intelligence-Management System*. International Standards Organization. https://www.iso.org/standard/42001

Kate, A. (2026, March 16). *A NIST CSF 2.0–Aligned AI-Driven Threat Intelligence Framework for Enhancing Cyber Resilience in U.S. Critical Infrastructure*. ResearchGate. https://www.researchgate.net/publication/402360104

McCormack, L. & Bendechache, M. (2026, June). The Trustworthy AI Maturity Model (TAIMM): Integrating ethics and regulation across the AI lifecycle. *Journal of Responsible Technology, 26*. https://www.sciencedirect.com/science/article/pii/S2666659626000090

NIST AI 100-1. (2023, January). *Artificial Intelligence Risk Management Framework: AI RMF1.0*. Department of Commerce. https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf

Papagiannidis, E., Mikalef, P., Conboy, K. (2025). Responsible artificial intelligence governance: A review and research framework. *Journal of Strategic Information Systems, 34*(2). https://www.sciencedirect.com/science/article/pii/S0963868724000672

Radanliev, P., De Roure, D., Page, K., … Burnap, P. (2019). *Cyber risk at the edge: Current and future trends on cyber risk analytics and artificial intelligence in the industrial internet of things and Industry 4.0 supply chains*. Cornell University. https://arxiv.org/abs/1911.05726

Sommer, R., & Paxson, V. (2010). Outside the closed world: On using machine learning for network intrusion detection. *IEEE Symposium on Security and Privacy*. https://oakland10.cs.virginia.edu/slides/anomaly-oakland.pdf

Yang, W., Wang, S., Wu, D., …Li, Y. (2025, May 13). Deep learning model inversion attacks and defenses: a comprehensive survey. *Artificial Intelligence Review, 58*(242). https://link.springer.com/article/10.1007/s10462-025-11248-0

Zeijlemaker, S., Lemiesa, Y., Schroer, S., Abhishta, A., & Siegel, M. (2025, September 23). How Does AI Transform Cyber Risk Management? *Systems, 13*(10), 835. https://www.mdpi.com/2079-8954/13/10/835

Zemello, S. (2023, November 23). *Challenges to achieving a unified, global AI regulation framework* [blog]. London School of Economics. https://blogs.lse.ac.uk/medialse/2023/11/02/challenges-to-achieving-a-unified-global-airegulation-framework/

FURTHER STUDIES ON SOLVING UNIVERSITY TIMETABLING PROBLEMS WITH GENETIC ALGORITHMS

Achini Herath, University of North Alabama, herathwimal@gmail.com
Dawn Wilkins, University of Mississippi, dwilkins@cs.olemiss.edu

ABSTRACT

Genetic algorithms (GA) have shown to produce promising results in generating conflict free timetables for larger institutes. As such, a program was initiated to find better solutions to such problems by way of changing parameters associated with GA. As an initial step, we decided to use Roulette Wheel as the selection method. The results obtained were comparable to those reported for the tournament method as described in the literature. Yet, none of the techniques suffered from the drawback, "unable to find a feasible solution to the problem", as in PSO, SAGA. This suggested that there may be room for further studies based on the variables of the parameters used to generate new optimized offspring. The number of classes that needed scheduling was determined by totalling the number of student groups, multiplied by the number of modules each group is registered for.

Keywords: Genetic Algorithm, timetabling, evolution, Roulette Wheel selection, three-parent crossover, Uni-One-Point crossover.

INTRODUCTION

Finding solutions to complex timetabling problems is considered to be NP-hard. Scientists have made many attempts to find solutions to such problems (Suh et al., 2011). Among them, Genetic Algorithms (GA) have shown to yield promising results (Ausiello et al., 2013, Desale et al., 2015). It is a nature motivated optimization search method which utilizes selection, crossover and mutation techniques. The other criteria include, the fitness function, size of the population, stopping conditions. An Application Programming Interface (API) is used to test the aforementioned criteria. The testing of these criteria is carried out using constrained satisfaction problems including, graph coloring (Welsh & Powell, 1967, Selim, 1988, Hafizah & Zaidah, 2010, Dandashi & Al-Mouhamed, 2010, Asham et al., 2011, Redl, 2004). Travelling Salesman (Goel et al., 2025) Simulated Annealing (Russell & Norvig, 1995, D. Abramson, 1991, Fredrikson & Dahl, 2016, Rosocha et al., 2015, Varty, 2017, Brownlee, 2011, Kirkpatrick et al., 1983, Chmait & Challita, 2013, Patrick, 2012), Sudoku (Douglas, 2014, Waters, 2007) and Tabu search (Glover, 1990) A major drawback of GA however, is its longer computational time (Moin et al., 2015) due to its population dependency.

The present investigation involves the use of three parents with roulette selection and uniform crossover techniques. In the search for optimal solutions in GA, selection methods play an important role in picking up high-quality solutions to contribute towards the fitness of the population. Of the available selection methods (Saptarini et al, 2018), rank-based and Roulette Wheel techniques responded well with larger size problems, while tournament selection method responded favourably with small problems. However, our results with the Roulette Wheel were

comparable to those obtained with the tournament method indicating the need for further investigations.

METHODOLOGY

University Timetabling Problem

Selection: Roulette Wheel technique.

Operators and Testing

Crossover, Mutation and Testing Criteria including Fitness Evaluation:

GA adapts exploration and exploitation procedures and the operators include mutations and crossovers respectively. Mutations randomly modify individual solutions to bring about diversity which is necessary to look for new search areas having better solutions. However, it is important to pay attention to the rates of mutations, as high rates tend to prevent the population reaching the desired optimum solution, whilst a lower rate prevent it from reaching a global maximum. Several common crossover methods [Uniform (UC), One-point (OPC), Uni-Three-parent (U3PC)] including recently introduced, Uni-One-Point crossover technique (Herath 2924) were employed. Procedures described in Herath, 2024 to solve other specific problems, such as Travelling Salesman, Sudoku and Graph Coloring were adopted.

Testing criteria

The criteria, namely, the Roulette Wheel, crossovers, stopping conditions, population size, fitness function, mutation rate were tested using an API. Constraint satisfaction problems such as graph coloring (Hindi, 2012), Travelling Salesman (Abdoun, 2012), Sudoku (Douglas, 2014) were used to test above criteria. The tests were performed to compare crossover methods, number of generations, population size and mutation rates.

EXPERIMENTS AND RESULTS

The efficiency of the aforementioned crossover methods and the Roulette Wheel (RS) selection technique was evaluated by comparing the results obtained for the Timetable Schedule (TTS), Traveling Salesman (TS) Sudoku and Graph Colouring (GC) problems. As per crossover, ten runs were carried out. Population and the number of generations, together with crossover and mutation rates were kept constant throughout the investigation.

Using hard constraints

Manual creation of a university timetabling problem was carried out with the following parameters.

Time slots per week: 10 (2 per day for 5 for 5 days), Number of rooms: 8, Instructors: 10 and Number of classes: 40 (4 classes a week per instructor).

The number of classes assigned with appropriate rooms and time slots for each crossover method is determined. **Table 1** and **Figure 2** illustrate the outcome of the present experiment when the Roulette selection method is adapted. The results indicate that the number of favourably assigned classes is similar as in our previous investigation, when Tournament method is used (Herath, 2024). For comparison purposes, results obtained with the Tournament method are also included

in **Table 1** and **Figure 2.** The highest number, **39.3** for the Tournament/One-point crossover in the previous investigation is taken as the best score as against the present work with the Roulette Wheel selection method.

Crossover Type (Where Applicable)	Roulette selection	Tournament selection
Uniform Crossover	37.3	38.9
One-point Crossover	37.1	**39.3**
Uni-OnePoint Crossover	37.3	38.9
Uni-three-parent Crossover	38.6	38.3

Table-1: Results of Timetable Schedule (**Hard** Constraints)

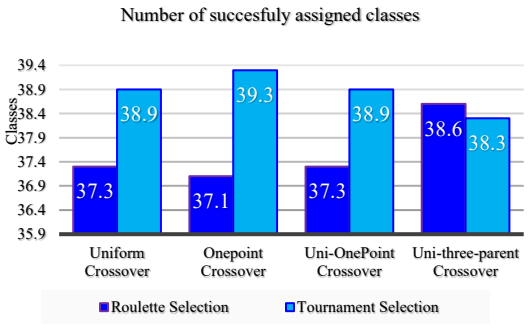


Figure-1: Number of successfully assigned classes with different Selection and Crossover methods

Using software constraints

Similar results, as with the hard constraints, were obtained by including soft constraints. It allowed each instructor to select four time slots and four rooms according to his/her liking. Results obtained are given in **Table 2** and **Figure 2**. The number of un-preferred classes is the lowest value to be considered (0 being the best score). In this experiment, **17.4** is the preferred lowest value for Roulette selection and the Uniform Crossover Methods.

Crossover Type	Roulette selection	Tournament selection
Uniform Crossover	**17.4**	18.7
Onepoint Crossover	19.0	19.3
Uni-OnePoint Crossover	19.3	19.4
Uni-three-parent Crossover	18.5	17.7

Table 2: Results of Timetable Schedule **Constraints**)

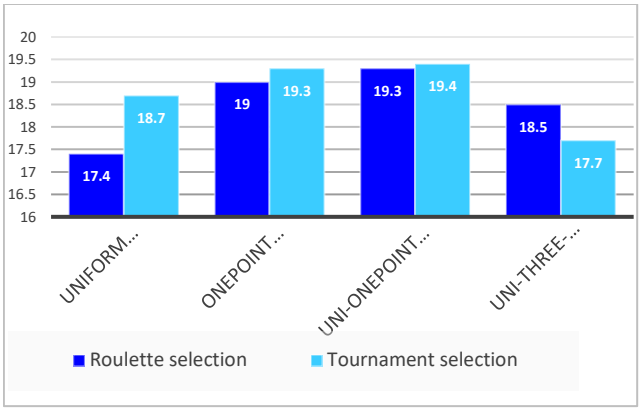


Figure 2: Number of un-preferred classes with 2 different selection and 4 different crossover methods (lower is better).

Timetable (160 and 200 classes) – Tables and Graphs

Timetable for medium size class (160) with hard constraints.

With the number of classes limited to 40 using different crossover techniques, changing selection methods seems to have little impact. As such, it was decided to repeat the experiment with medium (160 classes) and larger size (200 classes) problems. These involve 40 and 50 instructors with each person having to take 4 classes per 5 day week which provides 10 time slots. (2 time slots a day: 5 days a week). By dividing the number of rooms available by the number of time slots, the number

of rooms required to teach is determined. The number of rooms available being limited, it is a difficult to avoid constraint violations to obtain feasible solutions. At this point it was decided to compare GA results with each and some other combinatorial optimisation procedures including, Particle Swarm Optimisation (PSO), Simulated Annealing (SA) and Tabu Search (TS). The performance was evaluated using 10 runs. The results obtained were averaged.

Crossover Type (Where Applicable)	Roulette selection	Tournament selection	Results
Particle Swam (PS)			112.2
Simulated Annealing (SA)			111.1
Tabu Search (TS)			153.0
Uniform Crossover (UC)	119.4	120.5	
Onepoint Crossover (OC)	145.8	**147.6**	
Uni-OnePoint Crossover (UOC)	136.3	146.6	
Uni-three-parent Crossover (UTC)	129.3	141.625	

Table-3: Results of successfully assigned classes for a medium size class (**160**) with **hard** constraints, PSO, SA and TS.

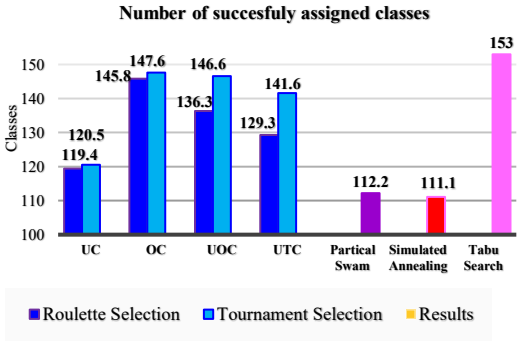


Figure-3: Number of successfully classes for a medium size class (**160**) with **hard** constraints, PSO, SA and TS.

As seen, the results obtained with hard constraints are very much similar with all the crossover methods irrespective of the selection method. The preferred highest value of 147.6 is seen with One-point crossover when the class size is increased to 160 (medium size).

Timetable for medium size class (160) with soft constraints.

The results when the number of classes was **160** are shown in the table (**Table 4**) and the bar diagram (**Figure 4**).

	Crossover Type	Roulette selection	Tournament selection	Results
1	Particle swarm			No acceptable results
2	Simulated annealing			No acceptable results
3	Tabu search			13.7
4	Uniform Crossover	No acceptable results	No acceptable results	
5	One-Point Crossover	13.125	**12.75**	
6	Uni-One-Point Crossover	**12.25**	23.0	
7	Uni-three-parent Crossover	74.5	33.375	

Table-4: Number of un-preferred classes assigned for a medium sized class (**160**) with **soft** constraints, Particle

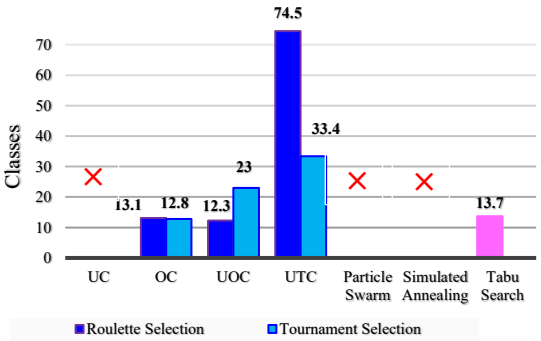


Figure-4: Number of un-preferred classes assigned for a medium sized class (**160**) with **soft** constraints.

The lowest preferred value, **12.25** is obtained with Roulette Wheel selection and One-point crossover combination. No acceptable results (0) were obtained with PSO and SA.

Timetable for large sized class (200) with hard constraints.

Crossover Type (Where Applicable)	Roulette selection	Tournament selection	Results
Particle Swam (PS)			132.2
Simulated Annealing (SA)			129.2
Tabu Search (TS)			181.6
Uniform Crossover (UC)	139.9	142.1	
One-point Crossover (OC)	173.8	174.3	
Uni-One-Point Crossover (UOC)	160.4	**176.1**	
Uni-three-parent Crossover (UTC)	147.8	163.4	

Table-5 Results of successfully assigned classes for large sized classes (**200**) with **hard** Constraints, PS, SA and TS

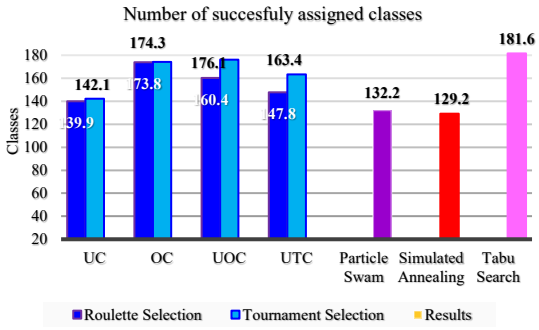


Figure e-5: Number of successfully assigned classes for large sized classes (**200**) with **hard** Constraint, PS, SA and TS.

176.1, the highest value was obtained with Uni-One-Point-Crossover and Tournament Selection combination. PS, SA and UC showed values of 132.2 and 129.2 and 181.6, respectively.

Timetable for large sized class (200) with soft constraints.

Table 6 and **Figure 6** show the results when the number of classes was increased to **200** with **soft** constraints.

Crossover Type	Roulette selection	Tournament selection	Results
Particle swarm (PS)			No acceptable results
Simulated annealing (SA)			No acceptable results
Uniform Crossover (UC)	No acceptable results	No acceptable results	
One-point Crossover (OC)	34.9	38.3	
Uni-One-Point Crossover (UOC)	33.8	97.5	
Uni-three-parent Crossover (UTC)	131.8	56.1	

Table-6: Results of un-preferred classes assigned for a large sized class (**200**) with **soft** constraints.

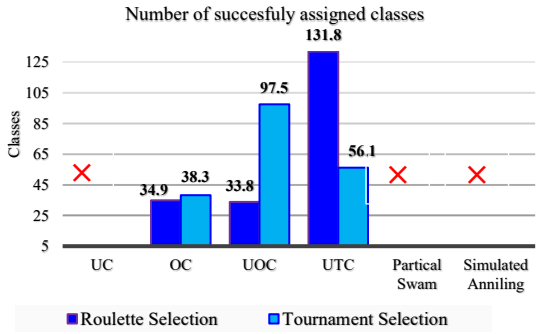


Figure-6: Results of un-preferred classes assigned for a large sized class (**200**) with **soft** constraints.

PSO, SA and Uniform crossover were not able to find a feasible solution (do not satisfy all the hard constraints). Thus, their soft constraint fitness was not considered.

Traveling Salesman

The sgb128_dist dataset provided the input data which contained a 128X128 matrix of city-city distance. The score to compare is the entire distance the salesman had to travel. The lowest value is the desired solution. The results of different crossover techniques are indicated in **Table-6** and **Figure-6.**

Crossover Type	Roulette selection	Tournament selection
Uniform Crossover	118698.6	90040.2
One-point Crossover	68551.8	75927.4
Uni-One-Point Crossover	76319.2	**63980.0**
Uni-three-parent Crossover	103447.7	104543.2

Table-7: Results of Traveling Salesman Problem.

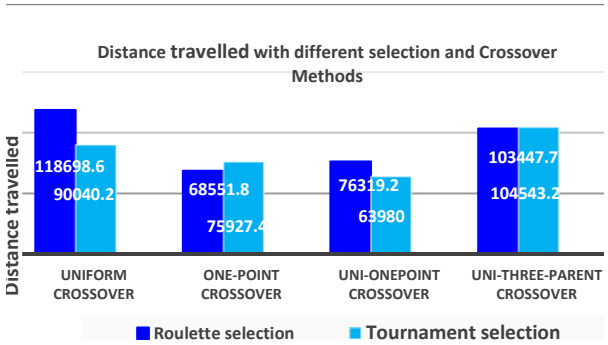


Table-7: Distance travelled with two different selection and four different Crossover methods.

The input data were selected from the sgb128_dist dataset. It contains a 128X128 matrix of city-city distance. The total distance the salesman has to travel is the score to compare. The best solution is the one with the lowest score, namely, Tournament selection with Uni-One-Point crossover. **Table-6** and **Figure-6** show the results of different crossover methods.

Institutes are looking for methods to determine the shortest distance a salesman (TS) reaches a certain destination without repeating routes for economic benefits. In here, they are utilized to examine the effectiveness of a genetic operator (U3PC) /Roulette Wheel selection method to utilize it to resolve the frequently encounter university timetabling issues. Different crossover operators (PSO, SA and TS) are used. As shown in Table-7 and Figure-7, the preferred lest distance travelled (63980) is given by the Roulette wheel selection method /UOPC combination.

Graph Coloring

This method uses a least number of colors to resolve problems. The desired input data is obtained from the queen 16_16 dataset , containing 256 vertices and 12640 edges. The score to compare is the number of colors of which the lowest is the desired solution. Table-7 and Figure-7 depict the results of Roulette Wheel and different crossover techniques. The data of the Tournament selection method are also included for comparison purposes. The lowest number of 95 is given by the Roulette Wheel/ Uniform crossover combination. However, the Tournament/ Uniform crossover showed an improved performance (92.7).

Crossover Type	Roulette selection	Tournament selection
Uniform Crossover	95	**92.7**
One-point Crossover	103.5	101.6
Uni-One-Point Crossover	97.5	93.6
Uni-three-parent Crossover	163.9	156.6

Table-8: Results of Graph Colouring Method. (minimum number of colours-lower is better).

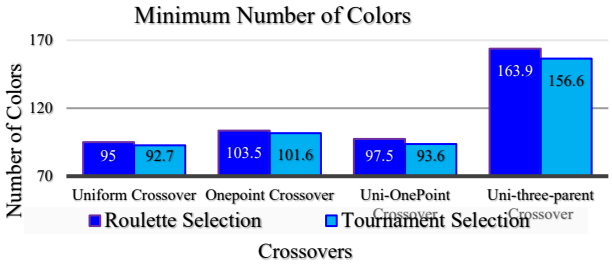


Figure-8: Number of colors with different Selection and Crossover methods.

Sudoku

Crossover type	Roulette selection	Tournament selection
Uniform Crossover	240.9	240.6
One-point Crossover	237.1	236.5
Uni-One-Point Crossover	238.2	239.4
Uni-three-parent Crossover	226.1	233.3

Table-9: Results of Sudoku

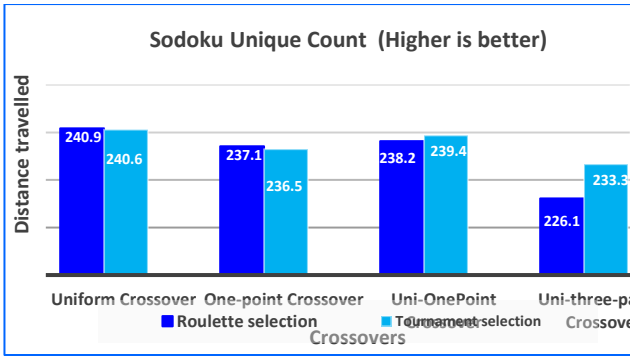


Figure-9: Number of colors with different Selection and Crossover methods

A sample sudoku quiz from https://sudoku.com/ was taken as input data. The sum of the distinct dights in 9 rows, 9 columns and 9 blocks was considered to be the score to compare. As such, 9x9x3 = 243 is the best score to achieve. The results of this experiment are as tabulated in **Table 9.**

In solving the Sodoku problem the uniform crossover with tournament selection combination gives the highest value (**240.6**) among others. For uniform crossover/Roulette Wheel combination a comparatively little higher value of **240.9** was obtained.

CONCLUSION:

The aim of the present work is to continue our previous research activities on finding solutions to complex timetabling problems, as seen in higher educational institutes. Such problems cannot be solved due to many constraints involved. As such, they become NP complete problems that cannot be resolved in polynomial time. Constraints are manly of two types. They are named, “hard” and “soft”. Whilst, all the hard constraints must be met to produce a feasible timetable, solutions to soft constraints improve its quality. A literature survey conducted, lead us to select, Genetic algorithm as a promising technique capable of solving such problems. The results of related work

are described in, Herath, 2024. Though, comparable results were obtained with the Tournament selection method and the selected crossover techniques including Uni-One-Point crossover operator, no drastic changes were observed. It is noteworthy, that none of the results obtained could be discarded as "non acceptable". This prompted us to carry out further investigations by changing, for example, the selection method from Tournament to Roulette Wheel to improve the search area to find worthwhile solutions using the same crossover methods. The change however, did not show impressive results. Using other varying combinations of selection/crossover methods along with other factors like population size and mutation changes may lead to expected results. As per, Saptarini et. al, work had carried out to solve the travelling salesman problem using the tournament and the roulette methods. Results had indicated that the former technique seems to work well with smaller size problems whereas the Roulette Wheel had had yielded agreeable results with larger problems. Our work however, gave comparable results with both the selection methods. This indicated the need for further work to support the observations.

REFERENCES

Abdoun, O., Abouchabaka, J. & Tajani,C., (2012). Analyzing the Performance of Mutation Operators to Solve the Travelling Salesman Problem. *IJES Int. J. Emerg. Sci.*, 2, 61-77.

Abramson, D., (1991). Constructing school timetables using simulated annealing: sequential and parallel Algorithms, *Management Science*, 37(1):98–113.

Asham, G. M., Soliman, M. M., Ramadan, A. R., (2011). Trans genetic coloring approach for timetabling problem. Artificial intelligence techniques novel approaches & practical applications, IJCA, 17–25.

Ausiello, G., Petreschi, R. (ed.) (2013). The Power of Algorithms: Inspiration and Examples in Everyday Life, Allesandro Panconesi, *The One Million Dollars Problems*, p. 72, Springer Science & Business Media. ISBN 3642396526.

Brownlee, J. (2011). Clever Algorithms. Nature-Inspired Programming Recipes, 1st Ed., Lulu Enterprises, Australia, p. 436.

Chmait, N., & Challita, K., (2013). 'Using simulated annealing and antcolony optimization algorithms to solve the scheduling problem', *Computer Science and Information Technology*, 1(3), 208–224).

Dandashi, A., M. Al-Mouhamed, M., (2010). "Graph coloring for class scheduling", *Proc. ACS/IEEEInt. Conf. Comput. Syst. Appl.*, pp. 1-4.

Desale, S., Rasool, A., Andhale, S., & Rane, P. (2015). Heuristic and Meta-Heuristic Algorithms and Their Relevance to the Real World: A Survey, *International Journal of Computer Engineering in Research Trends*, 2(5), 296-304. ISSN: 2349-7084.

Douglas, R., (2014). "Solving Sudoku puzzles with genetic algorithm", *Game Behaviour*, 1(1).

Fredrikson, R. & Dahl, Y. (2016). A comparative study between a simulated annealing and a genetic algorithm for solving a university timetabling problem, Tech. rep. Stockholm, Sweden: KTH, School of Computer Science and Communication (CSC), p. 45.

Glover, F., Tabu search. A tutorial, *Interfaces,* 20,74-94 (1990).

Goel, M., Bairwa, A. K. & Singh, J. (2025). Performance Evaluation of Heuristic and Meta-Heuristic Algorithms on Large-Scale Travelling Salesman Problem Instances. *IEEE Access*, 13, 156988–157010. https://doi.org/10.1109/ACCESS.2025.3606531.

Hafizah, A. R., Zaidah, I. (2010). Bipartite graph edge coloring approach to course timetabling. IEEE, 229–234.

Herath, A. & Wilkins, D. (2024). A comparative study on solving university timetabling problems with emphasis on Algorithms, Issues in Information Systems 25 (4), pp 392-408.

Hindi, M, & Yampolskiy, R. V., (2012). Genetic Algorithm Applied to the Graph Coloring Problem. per presented at the *Twenty-third Midwest Artificial Intelligence and Cognitive Science Conference* (MAICS) April 21-22, Cincinnati, OH.

Kirkpatrick, S., Gelatt, C. D. and Vecchi, M. P. (1983). Optimization by simulated annealing, Science, 220, 671–680.

Moin, N. H. Chung Sin O., & Omar, M (2015). Hybrid Genetic Algorithm with Multi-parents Crossover for Job Shop Scheduling Problems, *Math. Probl. Eng*., 12, 01-12. https://doi.org/10.1155/2015/210680.

Patrick, K. (2012). Comparison of simulated annealing and hill climbing in the course timetabling problem, African Journal of Mathematics and Computer Science Research, 5(11), 176-178.

Redl, T. A. (2004) *A study of university timetabling that blends graph coloring with the satisfaction of various essential and preferential conditions*. https://scholarship.rice.edu/handle/1911/102019.

Rosocha, L., Vernerova, S. & Verner, R. (2015). Medical Staff Scheduling Using Simulated Annealing. Quality Innovation Prosperity, 19(1), pp.1–11. Available at: http://www.qip-journal.eu/index.php/QIP/article/view/405.

Russell S. J. and Norvig, P. (1995). Artificial Intelligence: A Modern Approach. Pearson Education, 2 edition, *Prentice Hall, Englewood Cliffs, New Jersey 07632*. ISBN D-IH-IQBSOS-E.

Saptarini, N. G. A. P. H., Ciptayani, P. I., Wisswani, N. W., Suasnawa, I. W. and Indrayana, N., E. (2018). Comparing Selection Method in Course Scheduling Using Genetic Algorithm. *Proceedings of the International Conference on Science and Technology (ICST 2018)*, 102 (pp. 559-563). Atlantis Press.

Selim, S. M. (1988). Split vertices in vertex coloring and their application in developing a solution to the faculty timetable problem. The Computer Journal, 31(1), 76–82.

Suh, W. J. Park, C. S. & Kim, D. W. (2011). Heuristic vs. meta-heuristic optimization for energy performance of a post office building. Proceedings of the 12th conference of international building performance simulation association, Sydney, Australia: IBPSA., 704-711.; Zabidee, F. & Adnan, M. H. M. (2024). Optimization in university student timetables: A comprehensive literature review, Journal of Advanced Research in Applied Sciences and Engineering Technology, 41(1), 14-43.

Varty, Z. (2017). Simulated Annealing Overview. Retrieved from: http://www.lancaster.ac.uk/~varty/RTOne.pdf .

Waters, D. (2007). "SudoKube: Using Genetic Algorithms to Simultaneously Solve Multiple Combinatorial Problems", MS Thesis, Oklahoma State University. http://dc.library.okstate.edu/utils/getfile/collection/theses/id/3203/filena me/3204.pdf

Welsh, D. J. A., Powell, M. B. (1967). An upper bound for the chromatic number of a graph and its application to timetabling problems. The Computer Journal, 10, 85–86.

ARCHITECTING CYBER RESILIENCE IN DIGITAL SUPPLY CHAINS: A DYNAMIC CAPABILITIES PERSPECTIVE ON MODEL DEVELOPMENT IN THE INDUSTRY 5.0 ERA

Christine Shikutwa, College of Saint Benedict & Saint John's University, cshikutwa001@csbsju.edu
Oluminde Malomo, Virginia State University, omalomo@vsu.edu

ABSTRACT

Industry 5.0 is transforming supply chains through the integration of artificial intelligence (AI), the Internet of Things (IoT), cyber-physical systems (CPS), cloud computing, and advanced digital platforms. While these technologies enhance operational efficiency, agility, and collaboration, they also expand organizational cyber-attack surfaces and increase exposure to complex, interconnected cyber threats. Existing research largely examines these technologies individually and offers limited theoretical guidance on how organizations develop sustained cyber resilience across digitally integrated supply chain ecosystems. Grounded in Dynamic Capabilities Theory (DCT), this study develops the Cyber-Resilient Supply Chain Model (CRSCM) as a theoretically grounded conceptual framework for understanding cyber resilience as a dynamic organizational capability. The framework conceptualizes resilience through the complementary capabilities of sensing emerging cyber risks, seizing adaptive security opportunities, and reconfiguring organizational resources to sustain operational continuity under conditions of uncertainty. This study represents the first phase of a broader theory-development research program. Rather than empirically validating the framework, it establishes its theoretical architecture and explains its relevance to Industry 5.0 supply chains. The proposed model contributes by integrating Dynamic Capabilities Theory with contemporary supply chain cybersecurity challenges, shifting cyber resilience from a technology-centric perspective toward an organizational capability perspective, and providing a structured foundation for future empirical validation through expert review, Delphi studies, case studies, pilot implementations, and survey-based research.

Keywords: Cyber Resilient; Supply Chain Management; Dynamic Capability, Industry 5.0, Cyber Security

INTRODUCTION

Industry 5.0 is often described in technological terms as the next wave of AI, automation, or connectivity but this framing often understates what is at stake. Rather than treating technological advancement as an end in itself, Industry 5.0 represents a paradigmatic shift that integrates organizational and human-centered principles with advanced digital technologies, including artificial intelligence (AI), the Internet of Things (IoT), cyber-physical systems (CPS), cloud-based digital platforms, and advanced data analytics to design and orchestrate supply chains that are resilient, adaptive, sustainable, and fundamentally human-centric (Agrawal et al., 2024; Cao et al., 2025; Khan et al., 2025; Qadir, 2025; Saeed et al., 2023; Wu et al., 2025; Zaoui et al., 2026). This convergence has fundamentally reshaped the trajectory of global supply chains, moving them beyond the automation-centric paradigms of earlier industrial eras toward ecosystems oriented

around human judgment and organizational resilience rather than efficiency alone (Rahimi et al., 2024).

The realization of these benefits is contingent upon the organizational, technological, and institutional conditions in which they are deployed (Arman et al., 2025; Mansour & Bouchachia, 2026; Saeed et al., 2023). A supply chain equipped with AI, IoT sensors, and cyber-physical integration does not become resilient or human-centric by default; it becomes so only when organizations develop the capacity to translate technological capability into coordinated, adaptive practice (Dubey et al., 2023; Li et al., 2026; Liang & Xu, 2025; Zaman et al., 2025). While technological integration is essential for enhancing agility and long-term sustainability, it simultaneously introduces systemic cybersecurity risks, increasing organizational vulnerability to cascading, and multi-tier disruptions across the network (Arman et al., 2025; Ivanov, 2023; Saeed et al., 2023). Despite the growing academic focus on these transformations, existing literature remains highly fragmented and empirically underdeveloped, with much of the current discourse remaining symbolic rather than grounded in consistent empirical evidence (Wiener et al., 2025; Zaoui et al., 2026). Current studies predominantly adopt technology-centric perspectives that examine digital innovations in isolation, providing limited theoretical explanation regarding how they collectively orchestrate organizational resilience capabilities within digitally integrated ecosystems (Sonar et al., 2025; Zaoui et al., 2026). Grounded in Dynamic Capabilities Theory (DCT), this study conceptualizes cyber resilience as a dynamic organizational capability to sense emerging cyber risks, seize adaptive security opportunities, and reconfigure resources during disruptions (Herburger et al., 2024).

The efficiency gains associated with Industry 5.0 digitalization do not arrive without cost. The same technological convergence that reshapes how organizations design and operate their supply chains (Okolo et al., 2021). As IoT devices, cloud infrastructures, operational technologies, software vendors, logistics providers, and third-party suppliers grow increasingly interdependent, they form complex cyber-physical ecosystems that are correspondingly more exposed to sophisticated cyber threats. IoT-enabled environments illustrate this exposure clearly: insecure communication protocols, inadequate authentication mechanisms, weak firmware management, and decentralized system architectures make them a persistent point of systemic vulnerability (Admass et al., 2024; Piccarozzi et al., 2023; Qudus, 2025; Sah et al., 2024). Ransomware, data breaches, distributed denial-of-service (DDoS) attacks, supply chain infiltrations, and phishing schemes are now common across these environments, and the very interconnectivity that drives efficiency gains simultaneously makes it more difficult to identify vulnerabilities, trace intrusions, and coordinate timely response (Agarwal et al., 2024; Okolo et al., 2021; Pérez-Morón, 2022; Sobb et al., 2020).

The transition to Industry 5.0 represents a profound shift from the automation-centric efficiency paradigms of previous industrial eras toward an ecosystem that is human-centric, resilient, and adaptive by design. Enabling technologies such as artificial intelligence, the internet of things, and cyber-physical systems are central to this paradigm; however, their mere deployment does not guarantee organizational success (Arman et al., 2025; Ivanov, 2023; Saeed et al., 2023). On the contrary, the deep interconnectivity and data-intensive nature of these technologies substantially expand the attack surface of modern supply chains, exposing them to cyber threats capable of

propagating rapidly across organizational and inter-organizational boundaries. This paradox, wherein the very technologies that enable Industry 5.0's adaptive capacity also introduce new vectors of systemic vulnerability which underscores an urgent need for supply chains to be conceptualized not merely as efficient or digitally advanced, but as cyber-resilient by design (Agrawal et al., 2024; Cao et al., 2025; Khan et al., 2025; Qadir, 2025; Saeed et al., 2023; Wu et al., 2025; Zaoui et al., 2026). Despite growing scholarly interest in supply chain resilience and Industry 5.0 individually, the literature remains fragmented, offering limited integrative frameworks that theoretically link cyber resilience constructs to the specific socio-technical demands of Industry 5.0 environments. To address this absence of integrative conceptual frameworks, this study develops the Cyber-Resilient Supply Chain Model (CRSCM), a theoretically grounded conceptual framework that specifies the principal dimensions of cyber resilience within Industry 5.0 supply chain ecosystems. Consistent with established conventions in model development, wherein construct specifications precedes empirical operationalization, the CRSCM is positioned as a preliminary framework intended for subsequent refinement and case-based validation. We situate this paper as the essential first stage of a broader research program, establishing the theoretical architecture necessary to build a more resilient and secure Industry 5.0 supply chain (Saqib et al., 2026; Wei et al., 2026).

THEORETICAL BACKGROUND: DYNAMIC CAPABILITIES THEORY

Left unmanaged, these same technologies introduce systemic cybersecurity risks and multi-tier vulnerabilities that can propagate across an entire supply network. Bridging the gap between what these technologies make possible and what organizations actually achieve requires a theoretical lens capable of explaining adaptation under uncertainty which is precisely what Dynamic Capabilities Theory (DCT) offers, conceptualizing an organization's capacity to purposefully create, extend, and reconfigure its resource base in response to volatile environments, and be able to cope with change (Herburger et al., 2024; Mohamed Noor et al., 2025; Teece et al., 1997; Wei et al., 2026).

In SCM literature, dynamic capabilities (DCs) are a competitive necessity in helping firms respond to environmental challenges and compete in today’s business landscape (Herburger et al., 2024). Hence Digital transformation (DT) scholarship argues that digital technologies create strategic value only when embedded within broader organizational and industry-wide transformations rather than implemented as isolated technological initiatives (Wiener et al., 2025). Extending this logic to supply chain cybersecurity, cyber resilience can be conceptualized through Dynamic Capabilities Theory (DCT) as a higher-order organizational capability that enables firms to continuously sense emerging cyber threats, seize opportunities to strengthen resilience, and reconfigure technological, operational, and managerial resources to sustain secure and adaptive supply chain operations (Saqib et al., 2026). This perspective aligns with Vial's conceptualization of digital transformation as an ongoing process of organizational renewal driven by continuous adaptation to technological and environmental change. Within supply chain management, dynamic capabilities have similarly been recognized as a strategic necessity for responding to environmental turbulence and sustaining competitive advantage (Herburger et al., 2024). Consequently, DCT provides a stronger theoretical foundation than alternative perspectives. While the Resource-Based View explains how valuable resources generate competitive advantage, it offers limited insight into how firms repeatedly renew those resources under rapidly evolving

cyber threats. Likewise, technology adoption theories primarily explain the adoption of digital technologies rather than the organizational capabilities required to maintain resilience after implementation. By emphasizing the continuous processes of sensing, seizing, and reconfiguring, DCT explains how organizations strategically renew capabilities through learning and adaptation, making it particularly well suited for understanding cyber resilience in Industry 5.0 supply chains, where resilience depends on continuous organizational transformation rather than technological assets alone (Herburger et al., 2024; Saqib et al., 2025; Wiener et al., 2025).

RQ1: How can organizations leverage the dynamic capabilities of sensing, seizing, and reconfiguring to build cyber resilience within Industry 5.0 supply chain ecosystems?

RQ2: What are the core dimensions or components of cybersecurity readiness in operations supply chain performance?

Sensing emerging risks and opportunities.
In the hyper-connected, multi-actor networks characteristic of Industry 5.0, organizations must be able to detect disruption before it cascades across tiers. AI-enabled information processing allows firms to monitor internal and external environments continuously, surfacing early indicators of cybersecurity threats or logistical bottlenecks (Abbes, 2025; Herburger et al., 2024; Luca, 2025). This sensing capability enhances supply chain transparency simultaneously widens the attack surface available to adversaries, meaning visibility and vulnerability expand together.

Seizing adaptive opportunities.
Sensing a risk is of limited value without the organizational capacity to act on it. Industry 5.0 ecosystems increasingly rely on platform-based supply chains that function as orchestration layers, coordinating multiple actors in ways that extend beyond traditional linear supply chain models (Abbes, 2025; Herburger et al., 2024; Luca, 2025; Saqib et al., 2026). DCT frames this as a distinct managerial capability: it is human judgment and collaborative routines not the data that convert sensed information into timely, coordinated action.

Reconfiguring resources for resilience.
Long-term resilience depends on an organization's capacity to continuously reconfigure both its digital and physical assets rather than treating any single response as final. Platform-based ecosystems enable the dynamic reallocation of resources during crisis conditions for instance, sharing warehouse capacity or rerouting logistics on short notice (Abbes, 2025; Herburger et al., 2024; Luca, 2025; Saqib et al., 2026). Through an iterative process of learning and adjustment that keeps the supply chain oriented toward future conditions rather than fixed to past ones.

Beyond its explanatory power, DCT aligns naturally with the human-centric ethos of Industry 5.0 itself: it positions human judgment and organizational mindfulness as the mediating force between technological capability and resilient outcomes, rather than treating technology as sufficient on its own (Saqib et al., 2026). Given that the current literature on this intersection remains fragmented and, in places, largely symbolic rather than substantively theorized, a DCT-grounded framework offers the architecture needed to move this line of inquiry toward empirical validation, that will clarify how firms actually orchestrate digital intelligence to withstand, and adapt within, an

increasingly disruptive supply chain landscape (Saqib et al., 2026). Accordingly, Dynamic Capabilities Theory provides theoretical architecture upon which the proposed Cyber-Resilient Supply Chain Model (CRSCM) is constructed (Herburger et al., 2024; Saqib et al., 2026). Rather than viewing cyber resilience as a collection of isolated cybersecurity controls, the framework conceptualizes resilience as an organizational capability that continuously evolves through sensing environmental change, seizing adaptive opportunities, and reconfiguring resources to sustain operational continuity. This theoretical perspective distinguishes the proposed model from technology-centric approaches by emphasizing organizational adaptability as the central mechanism through which cyber resilience is achieved within Industry 5.0 ecosystems (Herburger et al., 2024; Luca, 2025; Saqib et al., 2026).

RELATED WORK: SUPPLY CHAIN VULNERABILITIES AND THE CYBER THREAT LANDSCAPE

Rather than reviewing every digital technology associated with Industry 5.0, this section focuses on technologies that directly influence cyber resilience within digitally integrated supply chains. Specifically, the discussion emphasizes AI, IoT, cyber-physical systems, cloud computing, and interconnected digital platforms because these technologies substantially reshape organizational attack surfaces, operational interdependencies, and resilience capabilities (Wiener et al., 2025). This focused review provides the theoretical context necessary for developing the proposed Cyber-Resilient Supply Chain Model (CRSCM). This review therefore concentrates on cybersecurity threats that illustrate the organizational challenges motivating the proposed framework rather than attempting to provide an exhaustive review of every Industry 5.0 technology. The accelerated deployment of AI, industrial internet of things, cloud computing, blockchain systems, and digitally integrated operational technologies has substantially increased organizational exposure to cyber risks. Consequently, supply chain cybersecurity has emerged as one of the most strategically significant dimensions of organizational risk management within modern business environments (Admass et al., 2024; Dubey et al., 2023; Gokkaya et al., 2026; Li et al., 2026; Liang & Xu, 2025; Luca, 2025; Qudus, 2025; Saqib et al., 2026; Song et al., 2024).

Notable High-Profile Cases of Cyber Threats in Supply Chain

The 2020 SolarWinds cyberattack exemplified the strategic risks associated with digitally interconnected supplier ecosystems. In this incident, sophisticated adversaries compromised trusted software update mechanisms to infiltrate numerous governmental and private-sector organizations across the globe, including multiple U.S. federal agencies and Fortune 500 corporations (Buresh, 2022; Butler et al., 2024; Wiener et al., 2025). The attack demonstrated how software supply chain dependencies can be systematically weaponized to circumvent conventional perimeter-based cybersecurity defenses and obtain unauthorized access to critical digital infrastructures. Critically, the incident underscored the necessity for comprehensive supplier governance frameworks, continuous third-party risk monitoring protocols, and resilient cybersecurity architectures capable of detecting and mitigating systemic, cascading cyber threats at scale (Gokkaya et al., 2026; Keskin et al., 2021; Li et al., 2026; Luca, 2025; Qudus, 2025; Wiener et al., 2025).

Equally instructive is the 2021 Colonial Pipeline ransomware attack, which precipitated the forced operational shutdown of one of the United States' most strategically significant fuel distribution

networks. The incident triggered widespread fuel supply disruptions, regional economic instability, and consumer panic exposing the acute vulnerability of critical national infrastructure to target cyber intrusions. Beyond its immediate operational consequences, the attack illuminated the systemic interdependencies between digital operational continuity and physical supply chain functionality, reinforcing the imperative for sector-wide cyber resilience investment and regulatory intervention (Cartwright & Cartwright, 2023).

Collectively, these cases demonstrate that cyber resilience is less a function of any individual technology than of an organization's ability to coordinate multiple digital capabilities across interconnected supply chain ecosystems. Consequently, the proposed framework emphasizes organizational adaptation rather than technology-specific solutions.

Figure 1 illustrates the conceptual cyber threat ecosystem that motivates the development of the proposed Cyber-Resilient Supply Chain Model (CRSCM), highlighting representative cyber threats and the corresponding Dynamic Capabilities of sensing, seizing, and reconfiguring that underpin organizational cyber resilience.

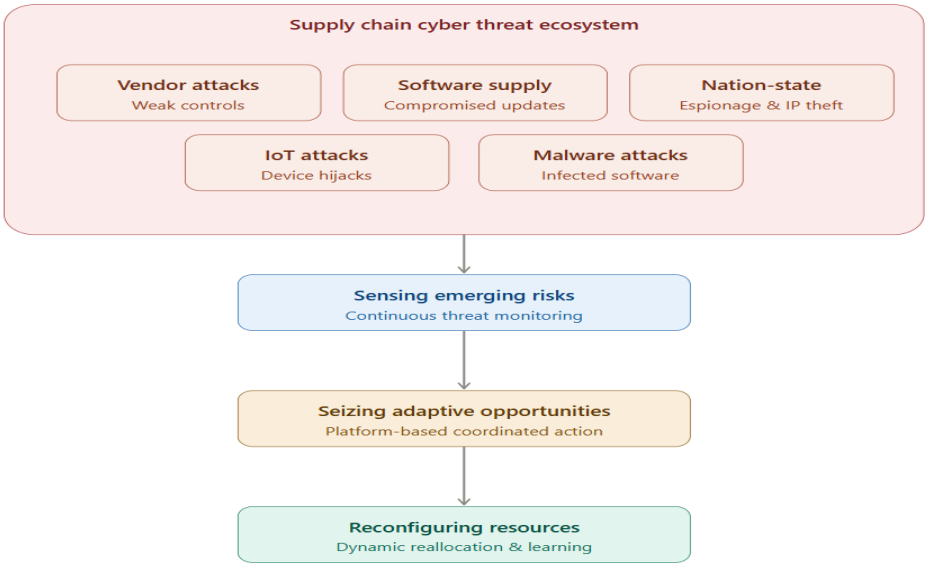


Figure 1. Supply Chain Cyber Threat Ecosystem

PROPOSED FRAMEWORK: DYNAMIC CAPABILITIES THEORY AND CYBER RESILIENCE (CRSCM)

Building upon the Dynamic Capabilities Theory developed in the preceding section, this study develops the Cyber-Resilient Supply Chain Model (CRSCM) as a conceptual framework for understanding cyber resilience as a dynamic organizational capability rather than a static technological defense mechanism. From this perspective, cyber resilience reflects an organization's ability to dynamically adapt technological, operational, and strategic resources in response to evolving cybersecurity threats and environmental disruptions (Abbes, 2025; Herburger et al., 2024; Saqib et al., 2026).

The sensing capability refers to the organization's ability to identify, interpret, and anticipate emerging cyber threats, disruptions, and environmental changes across supply chain ecosystems. Advanced analytical infrastructures such as AI, machine learning, predictive analytics, and IoT-enabled monitoring systems enhance organizational visibility and facilitate real-time threat detection and predictive risk assessment. These capabilities strengthen organizational agility by enabling continuous environmental scanning for cyber vulnerabilities, supplier risks, geopolitical instability, and operational disruptions before they escalate into systemic failures (Abbes, 2025; Herburger et al., 2024; Luca, 2025; Wiener et al., 2025).

Seizing capability represents the organization's capacity to mobilize strategic resources and technological infrastructures in response to identified threats and opportunities. Technologies such as blockchain, fintech, cloud computing, and digital collaboration platforms enable organizations to strengthen transparency, improve financial responsiveness, and coordinate decentralized operational responses during cyber disruptions. Through these mechanisms, organizations transform sensed information into actionable resilience strategies that support continuity and adaptive decision-making (Abbes, 2025; Herburger et al., 2024; Luca, 2025).

The reconfiguring capability emphasizes the organization's ability to continuously realign and transform operational and digital resources in response to environmental instability and technological change. Organizations strengthen this capability through cloud-based operational architectures, digitally enabled collaboration systems, circular economy initiatives, and strategic supply chain redesign efforts. Reconfiguring capability is therefore essential for long-term adaptation because resilience depends not only on recovery from disruptions but also on the organization's capacity for continuous transformation and strategic renewal (Abbes, 2025; Herburger et al., 2024; Luca, 2025).

The framework demonstrates how interconnected digital technologies collectively strengthen sensing, seizing, and reconfiguring capabilities, thereby enabling organizations to develop mature cyber resilience mechanisms that support operational continuity, strategic adaptability, and cybersecurity governance within digitally integrated supply chain ecosystems. Rather than conceptualizing resilience as a static protective mechanism, the model positions cyber resilience as a continuous process of organizational learning, adaptation, and technological orchestration under conditions of uncertainty and disruption (Luca, 2025; Wiener et al., 2025).

Collectively, the sensing–seizing–reconfiguring framework provides a comprehensive conceptual foundation for understanding how digitally enabled organizations develop resilient supply chain systems under volatile environmental conditions. The framework suggests that digital technologies function not merely as operational tools but as strategic enablers that strengthen organizational adaptability, agility, and resilience capabilities. Accordingly, this study conceptualizes supply chain resilience as a multidimensional dynamic capability through which organizations continuously sense environmental changes, seize strategic opportunities for response, and reconfigure operational resources to sustain long-term competitiveness and operational continuity within increasingly complex digital ecosystems (Abbes, 2025; Herburger et al., 2024; Luca, 2025; Wiener et al., 2025).

Theoretical Contribution and Future Validation

The Cyber-Resilient Supply Chain Model (CRSCM) proposed in this study should be viewed as the initial conceptual stage of a broader theory-development process rather than a finalized maturity assessment instrument. Like many foundational conceptual models within information systems, cybersecurity, and supply chain research, the framework is first articulated theoretically before undergoing empirical refinement and validation (Buresh et al., 2022; Herburger et al., 2024; Saqib et al., 2026).

Accordingly, this study intentionally focuses on defining the conceptual dimensions of cyber resilience grounded in Dynamic Capabilities Theory. Future research should evaluate and refine the proposed framework through expert panel review, Delphi studies, pilot organizational implementations, multiple case studies, and survey-based empirical validation. These subsequent studies will assess construct validity, dimensional completeness, inter-rater reliability, and practical applicability across different Industry 5.0 supply chain environments. By clearly separating conceptual development from empirical testing, this study establishes a rigorous theoretical foundation upon which future validation efforts can systematically build (Herburger et al., 2024; Seid et al., 2025; Wiener et a., 2025).

METHODOLOGY

Research Design and Philosophical Orientation

This study adopts a conceptual research design grounded in systematic literature synthesis, theoretically informed case analysis, and theoretical integration. Because the objective is theory development rather than empirical hypothesis testing, this approach provides the most appropriate foundation for developing and justifying the proposed Cyber-Resilient Supply Chain Model (CRSCM). The study further adopts an interpretivist philosophical orientation, recognizing that organizational responses to cyber threats are context-dependent, socially constructed, and embedded within complex institutional and technological environments. This position is consistent with the methodological conventions adopted in foundational supply chain resilience scholarship, including the seminal work of Pettit et al. (2010), Hohenstein et al. (2015), and more recently, Wieland and Durach (2021).

CONCLUSION

This study advances a theoretically grounded reconceptualization of cyber resilience within digitally interconnected supply chain environments, situating it within the Dynamic Capabilities Theory (DCT) framework as originally articulated by Teece et al. (1997) and subsequently extended within supply chain scholarship by Ambulkar et al. (2015), Butler et al. (2024, Ralston and Blackhurst (2020), Saqib et al. (2026) and Seid et al. (2025).

This theoretical reframing carries significant implications for both scholarship and practice. It repositions cyber resilience from a static, compliance-oriented construct toward a dynamic, capability-building imperative one that is intrinsically linked to an organization's broader capacity for strategic agility, operational continuity, and long-term competitive sustainability within the complex sociotechnical ecosystems characteristic of Industry 5.0. Although the present study does not empirically validate the proposed framework, this limitation reflects the deliberate scope of conceptual model development rather than a methodological deficiency. The CRSCM establishes

a theoretically grounded foundation that future studies can operationalize and evaluate using qualitative and quantitative methods. By outlining a structured pathway for empirical validation, this research provides a platform for developing standardized measures of cyber resilience maturity applicable to Industry 5.0 supply chain ecosystems. Future research should empirically evaluate and refine the proposed framework through qualitative and quantitative investigations across diverse Industry 5.0 environments. Such efforts will strengthen the theoretical and practical understanding of cyber resilience and contribute to the continued advancement of supply chain resilience scholarship.

REFERENCES

Abbes, I. (2025). Strategic pathways for innovation and sustainability in digital transformation: Insights from leading global companies. *Social Sciences & Humanities Open*, *12*, 101906.

Admass, W. S., Munaye, Y. Y., & Diro, A. A. (2024). Cyber security: State of the art, challenges and future directions. *Cyber Security and Applications*, *2*, 100031.

Agarwal, A., Gupta, S., & Chaudhary, A. (2024, October). Industry 5.0: enhanced human-machine synergy for revolutionary innovation. In *2nd International Conference on Pervasive Computing Advances and Applications (PerCAA 2024)* (Vol. 2024, pp. 134-141). IET.

Ambulkar, S., Blackhurst, J., & Grawe, S. (2015). Firm's resilience to supply chain disruptions: Scale development and empirical examination. *Journal of Operations Management*, *33–34*, 111–122. https://doi.org/10.1016/j.jom.2014.11.002

Buresh, D. (2022). The Solar Winds Cyber-Attack, the Federal and Private Sector Response, and the Recommendations and Lessons Learned. *International Journal of Innovation Scientific Research and Review*.

Butler, P., Kelley, J., Ellis, J., & Olatunbosun, S. (2024, July). Cybersecurity threats: An analysis of the rise and impacts of state sponsored cyber attacks. In *World congress in computer science, computer engineering & applied computing* (pp. 187-194). Cham: Springer Nature Switzerland.

Cao, B., Wang, T., Li, A., Shang, Y., & Zhu, J. (2025). Unveiling the social and environmental benefits of digital transformation in corporations. *PLoS One*, *20*(3), e0320064.

Cartwright, A., & Cartwright, E. (2023). The economics of ransomware attacks on integrated supply chain networks. *Digital Threats: Research and Practice*, *4*(4), 1-14.

Dubey, R., Bryde, D. J., Dwivedi, Y. K., Graham, G., Foropon, C., & Papadopoulos, T. (2023). Dynamic digital capabilities and supply chain resilience: The role of government effectiveness. *International Journal of Production Economics*, *258*, 108790.

Dupont, B., Shearing, C., Bernier, M., & Leukfeldt, R. (2023). The tensions of cyber-resilience: From sensemaking to practice. *Computers & security*, *132*, 103372.

Gao, S., He, P., Fu, Y., Li, Y., Wang, H., Wang, Q., ... & Gong, X. (2026). Mapping the Evolution of DSSAT Model Research: Trends, Transitions, and Future Frontiers. *Agronomy*, *16*(2), 141.

Gokkaya, B., Aniello, L., & Halak, B. (2026). Software supply chain: A taxonomy of attacks, mitigations and risk assessment strategies. *Journal of Information Security and Applications*, *97*, 104324.

Hammi, B., Zeadally, S., & Nebhen, J. (2023). Security threats, countermeasures, and challenges of digital supply chains. *ACM Computing Surveys*, *55*(14s), 1-40.

Herburger, M., Wieland, A., & Hochstrasser, C. (2024). Building supply chain resilience to cyber risks: a dynamic capabilities perspective. *Supply Chain Management: An International Journal*, *29*(7), 28-50.

Hohenstein, N. O., Feisel, E., Hartmann, E., & Giunipero, L. (2015). Research on the phenomenon of supply chain resilience: A systematic review and paths for further investigation. *International Journal of Physical Distribution & Logistics Management*, *45*(1/2), 90–117. https://doi.org/10.1108/IJPDLM-05-2013-0128.

Humayun, M., Niazi, M., Jhanjhi, N. Z., Alshayeb, M., & Mahmood, S. (2020). Cyber security threats and vulnerabilities: a systematic mapping study. *Arabian Journal for Science and Engineering*, *45*(4), 3171-3189.

Ivanov, D. (2023). The Industry 5.0 framework: Viability-based integration of the resilience, sustainability, and human-centricity perspectives. *International Journal of Production Research, 61*(5), 1683–1695. https://doi.org/10.1080/00207543.2022.2118892

Luca, C. (2025). Digital Resilience in Cyber security Supply Chains.

Khan, M. I. M., Sun, Q., Ahmed, Z., Zhang, R., & Khosa, M. (2025). Orchestrating Digital Capabilities for Supply Chain Resilience: Evidence from China-Pakistan Economic Corridor. *Systems*, *13*(11), 981.

Keskin, O. F., Caramancion, K. M., Tatar, I., Raza, O., & Tatar, U. (2021). Cyber third-party risk management: A comparison of non-intrusive risk scoring reports. *Electronics*, *10*(10), 1168.

Li, L., Shan, S., Li, L., & Jia, X. (2026). Untangling the impact of process innovation on supply chain resilience: The moderating roles of corporate social responsibility and digital technology adoption. *International Journal of Production Economics, 300*, 110092. https://doi.org/10.1016/j.ijpe.2026.110092

Liang, X., & Xu, Y. (2025). A novel framework to identify cybersecurity challenges and opportunities for organizational digital transformation in the cloud. *Computers & Security*, *151*, 104339.

Mansour, I. E. A., & Bouchachia, A. (2026). Institutional conditions for digital innovation and transformation: a contingent framework for smart technology adoption in developing nations. *Sustainability*, *18*(6), 2868.

Mohamed Noor, A. F., Moghavvemi, S., Tajudeen, F. P., Motaghi, H., & Salehi, A. (2025). Factors affecting cybersecurity readiness from dynamic capabilities perspective: a thematic review.

Okolo, F. C., Etukudoh, E. A., Ogunwole, O. L. U. F. U. N. M. I. L. A. Y. O., Osho, G. O., & Basiru, J. O. (2021). Systematic review of cyber threats and resilience strategies across global supply chains and transportation networks.

Wei, S., Jiang, J., Chen, J., & Wang, Q. (2026). Digital technology and supply chain resilience: A literature review and emerging themes. *International Transactions in Operational Research*, *33*(4), 2167-2189.

Wiener, M., Strahringer, S., & Kotlarsky, J. (2025). Where are the processes in IS research on digital transformation? A critical literature review and future research directions. *The Journal of Strategic Information Systems*, *34*(2), 101900.

Pettit, T. J., Fiksel, J., & Croxton, K. L. (2010). Ensuring supply chain resilience: Development of a conceptual framework. *Journal of Business Logistics*, *31*(1), 1–21. https://doi.org/10.1002/j.2158-1592.2010.tb00125.x

Pérez-Morón, J. (2022). Eleven years of cyberattacks on Chinese supply chains in an era of cyber warfare, a review and future research agenda. *Journal of Asia Business Studies*, *16*(2), 371-395.

Ralston, P., & Blackhurst, J. (2020). Industry 4.0 and resilience in the supply chain: A driver of capability enhancement or capability loss? *International Journal of Production Research*, *58*(16), 5006–5019. https://doi.org/10.1080/00207543.2020.1736724

Saqib, Z. A., Qin, L., & Ikram, M. (2026). Artificial intelligence-driven strategies for enhancing operational performance and managing uncertainties in the supply chain management. *Operations Management Research*, *19*(2), 30.

Saeed, S., Altamimi, S. A., Alkayyal, N. A., Alshehri, E., & Alabbad, D. A. (2023). Digital transformation and cybersecurity challenges for businesses resilience: Issues and recommendations. *Sensors, 23*(15), 6666. https://doi.org/10.3390/s23156666

Seid, E., Ilter, D., Blix, F., & Popov, O. (2025). Cyber Supply Chain Resilience: Analyzing ISO, NIST, and NIS2 Frameworks for Mitigating Third-Party Risks. *Procedia Computer Science*, *263*, 591-599.

Sobb, T., Turnbull, B., & Moustafa, N. (2020). Supply chain 4.0: A survey of cyber security challenges, solutions and future directions. *Electronics*, *9*(11), 1864.

security challenges, solutions and future directions. *Electronics*, *9*(11), 1864.

Sonar, H., Ghag, N., & Sharma, I. (2025). Reshaping Industry 5.0: Unveiling supply chain resilience for a carbon-neutral future. *Sustainable Futures, 9*, 100513. https://doi.org/10.1016/j.sftr.2025.100513

Song, U., Hur, G., Lee, S., & Park, J. (2024). Unraveling the dynamics of the cyber threat landscape: Major shifts examined through the recent societal events. *Sustainable Cities and Society*, *103*, 105265.

Teece, D. J., Pisano, G., & Shuen, A. (1997). Dynamic capabilities and strategic management. *Strategic Management Journal*, *18*(7), 509–533. https://doi.org/10.1002/smj.4250180703

Qadir, A. (2025). THE FUTURE OF CYBER-PHYSICAL SYSTEMS: INTEGRATING AI, IOT, AND CLOUD FOR INDUSTRIAL REVOLUTION 5.0. *Frontiers in Multidisciplinary Studies*, *2*(01), 43-53.

Qudus, L. (2025). Advancing cybersecurity: Strategies for mitigating threats in evolving digital and IoT ecosystems. *International Research Journal of Modernization in Engineering Technology and Science, 7*, 3195–3201. https://doi.org/10.56726/IRJMETS66504

Wieland, A., & Durach, C. F. (2021). Two perspectives on supply chain resilience. *Journal of Business Logistics*, *42*(3), 315–322. https://doi.org/10.1111/jbl.12271

Wu, H., Liu, J., & Liang, B. (2025). AI-driven supply chain transformation in Industry 5.0: Enhancing resilience and sustainability. *Journal of the Knowledge Economy*, *16*(1), 3826-3868.

Zaman, J., Shoomal, A., Jahanbakht, M., & Ozay, D. (2025). Driving supply chain transformation with IoT and AI integration: A dual approach using bibliometric analysis and topic modeling. *IoT*, *6*(2), 21.

Zaoui, H., Kamsu-Foguem, B., & Tchuente, D. (2026). Exploring the digital supply chain in the Industry 5.0 era: A literature review and empirical study of sustainability. *Digital Engineering, 8*, 100074. https://doi.org/10.1016/j.dte.2025.100074.

Proceedings of the 66th International Association for Computer Information Systems Conference - October 7 - 10, 2026 | Pensacola, FL

AI-ENABLED APPROACHES FOR OPTIMIZING GLP-1RAS MEDICATION MANAGEMENT: A SYSTEMATIC LITERATURE REVIEW

Melissa Insaidoo, Dakota State University, melissa.insaidoo@trojans.dsu.edu
Abeer Alkhwaldi, Dakota State University, Abeer.Alkhwaldi@dsu.edu
Omar El-Gayar, Dakota State University, Omar.El-Gayar@dsu.edu

ABSTRACT

The global burden of diabetes and obesity continues to escalate, imposing significant strain on healthcare systems worldwide. Glucagon-like peptide-1 receptor agonists (GLP-1 RAs) have emerged as transformative therapies for type 2 diabetes and obesity management, yet their optimal use is constrained by inter-individual response variability, adherence challenges, safety monitoring requirements, and fragmented clinical workflows. Artificial intelligence (AI) offers promising capabilities to address these challenges through predictive analytics, personalized treatment pathways, and real-time monitoring. This systematic literature review synthesizes 34 peer-reviewed studies published between 2022 and 2026 to examine AI-enabled approaches across the GLP-1 medication management lifecycle. Guided by the Research Question (RQ): "What themes emerge from the literature on AI-enabled approaches to GLP-1 RA medication management, and how can these themes be interpreted through the lens of Structuration Theory?" This review identifies five thematic areas: (1) AI-driven personalization, (2) adherence monitoring, (3) pharmacogenomics and precision medicine, (4) safety monitoring and ethical considerations, and (5) integration challenges and barriers. Findings are interpreted through Giddens' Structuration Theory, revealing recursive relationships among AI systems (structure), clinical actors (agency), and the ongoing co-constitution of technological and organizational practices (duality of structure). The reviewed literature suggests that while AI demonstrates considerable potential across the GLP-1 management lifecycle, the evidence base remains concentrated in early-stage and retrospective studies, with limited prospective implementation evidence. This review contributes to the healthcare informatics and information systems scholarship by demonstrating how sociotechnical theory can illuminate the recursive dynamics of AI-enabled clinical decision systems and proposes a research agenda addressing equity, implementation, and longitudinal validation gaps

***Keywords**:* artificial intelligence, GLP-1 receptor agonists, medication management systematic literature review, structuration theory, precision medicine, adherence monitoring

INTRODUCTION

The global burden of diabetes and obesity represents one of the most pressing public health challenges of the 21st century, with over 500 million adults affected by diabetes alone and prevalence rates continuing to rise across both high-income and low- and middle-income countries (Hay et al., 2025). The introduction of glucagon-like peptide-1 receptor agonists (GLP-1 RAs), including semaglutide (Ozempic), tirzepatide (Mounjaro), and liraglutide (Victoza), Glucagon-like peptide-1 receptor agonists (GLP-1 RAs) have emerged as a transformative class of medications demonstrating efficacy in glycemic control, weight reduction, and cardiovascular risk mitigation (Drucker, 2024). However, the optimal clinical use of GLP-1 RAs is constrained by

significant inter-individual variability in treatment response, suboptimal medication adherence, safety monitoring complexities, and fragmented healthcare delivery systems (Young et al., 2023). Artificial intelligence (AI) and machine learning (ML) have demonstrated growing potential to address these challenges by enabling personalized treatment pathways, real-time patient monitoring, and data-driven clinical decision support. Yet despite the rapid expansion of research at the intersection of AI and GLP-1 therapy, no existing systematic review integrates AI applications across the full spectrum of the GLP-1 medication management lifecycle within a unified sociotechnical theoretical framework.

This systematic literature review addresses this gap by synthesizing 34 peer-reviewed studies published between 2022 and 2026, guided by the following research question:

RQ: What themes emerge from the literature on AI-enabled approaches to GLP-1 RA medication management, and how can these themes be interpreted through the lens of Structuration Theory?

The adoption of Structuration Theory (Giddens, 1984) as the interpretive framework reflects the inherently sociotechnical nature of AI-enabled medication management, where technological capabilities (structure) and human clinical action (agency) are recursively co-constituted through ongoing interaction. This framing enables the review to move beyond descriptive cataloguing toward explanatory synthesis of how AI systems and clinical practices mutually shape one another.

This review contributes to both health informatics and information systems (IS) scholarship by: (1) providing the first integrated synthesis of AI applications across the complete GLP-1 medication lifecycle, (2) applying Structuration Theory to illuminate recursive dynamics between AI systems and clinical agency, and (3) identifying critical gaps in equity, implementation, and longitudinal validation that define the field's research agenda.

BACKGROUND ON GLP-1 RECEPTOR AGONISTS AND AI IN HEALTHCARE

GLP-1 Receptor Agonists

Glucagon-like peptide-1 receptor agonists (GLP-1 RAs) are a class of medications that mimic the naturally occurring incretin hormone GLP-1, which is produced in the distal small intestine in response to nutrient ingestion. GLP-1 enhances glucose-dependent insulin secretion, suppresses glucagon release, slows gastric emptying, and signals satiety to the brain, thereby addressing both metabolic and behavioral dimensions of diabetes and obesity (Singh et al., 2026).

The therapeutic development of GLP-1 RAs spans several decades. In 1992, exendin-4 derived from the saliva from a venomous lizard Heloderma suspectum (the “Gila monster”), became the first therapeutically used GLP-1RA, with preclinical studies in 1996 per (Bloomgarden, 2025). Subsequent FDA approvals include exenatide (2005), liraglutide (2010), dulaglutide (2014), semaglutide (2017) and tirzepatide (a dual combination of agonist of GLP-1 and GIP) which received approval in 2022. FDA approvals followed for liraglutide, semaglutide, and tirzepatide for treatment of obesity and were granted in 2014, 2021, and 2023, respectively, (Bloomgarden, 2025). GLP-1 RAs are currently indicated for type 2 diabetes mellitus and obesity management, with demonstrated effects on HbA1c reduction, weight loss, and cardiovascular risk reduction (Drucker, 2024; Shah et al., 2025).

Challenges in GLP-1 Medication Management

Despite their therapeutic benefits, GLP-1 RAs face persistent clinical challenges limiting real-world effectiveness:

- Heterogeneity in a patient's response: An individual's variability in glycemic, weight, and cardiovascular outcomes across patient populations to medicine (Young et al., 2023)
- Suboptimum adherence: Barriers may include gastrointestinal side effects, injection burden, economic cost, and limited ease of accessibility (Teke et al., 2025)
- Safety monitoring concerns: Gastrointestinal adverse events and long-term safety uncertainties require ongoing surveillance (Zheng et al., 2026)
- Workflow integration barriers: Fragmented digital health systems and limited interoperability impede coordinated care delivery (Gangwal & Lavecchia, 2025).

These challenges may span the full medication management lifecycle and collectively motivate the exploration of technology-enabled solutions.

Artificial Intelligence in Healthcare

Artificial intelligence refers to computational systems capable of performing tasks requiring human intelligence, including pattern recognition, prediction, and decision support. In healthcare, AI encompasses machine learning (algorithms that learn from data), natural language processing (interpretation of unstructured clinical text), deep learning (neural networks for complex data), and predictive analytics (forecasting clinical outcomes) (Giorda et al., 2023). In chronic disease management, these capabilities are being applied across multiple domains: patient stratification and risk prediction, treatment personalization, real-time monitoring, pharmacogenomic analysis, and clinical decision support integration.

The Intersection of AI and GLP-Therapy

In review of the convergence of AI capabilities and GLP-1 medication management, it represents a natural alignment: the challenges of GLP-1 therapy (response heterogeneity, adherence to complexity, safety surveillance and workflow fragmentation) which correspond directly to established AI strengths (pattern recognition, personalization, real-time monitoring and system integration). This intersection forms the focus of the present systematic review.

RELATED WORK

This section positions the present review relative to existing synthesis efforts addressing AI in metabolic disease management and GLP-1 receptor agonist therapy.

Existing reviews of GLP-1 Receptor Agonist Therapy

Drucker (2024) provides a comprehensive narrative of the GLP-1 journey from discovery to therapeutic impact, while Bloomgarden (2025) offers an updated historical and clinical review of GLP-1 RA development. Abdelrahman et al. (2025) review prospects and obstacles for GLP-1 RA use in obesity treatment. Young et al. (2023) systematically examined treatment effect heterogeneity in GLP-1 RA and SGLT2i responses, establishing the clinical rationale for personalization. However, these reviews focus exclusively on the clinical pharmacology of GLP-

1 RAs and do not address the roles of artificial intelligence or computational approaches in optimizing therapy delivery.

Existing reviews of AI in Metabolic Disease

From an AI perspective Teke et al. (2025) reviewed AI applications for obesity management broadly by covering mobile health tools and digital interventions. Gangwal & Lavecchia (2025) reviewed AI in anti-obesity drug discovery. Shapiro et al. (2025) examined AI for accelerating therapy discovery in type 1 diabetes. These reviews address AI in the broader metabolic space but do not synthesize AI applications across the GLP-1 medication management lifecycle. These reviews address AI in the broader metabolic disease space but do not specifically synthesize AI applications across the GLP-1 medication management lifecycle.

The Gap This Review Addresses

Despite growing interest in both AI-enabled healthcare and GLP-1 therapy, there is no existing systematic review that integrates AI applications across the full GLP-1 medication management spectrum; spanning patient selection, personalization, pharmacogenomics, adherence monitoring, safety surveillance, and workflow integration, within a unified sociotechnical framework. This SLR aims to fills that gap through integrated concept-centric synthesis interpreted via Structuration Theory.

Having established the positioning of this review, the following section introduces the theoretical lens through which the synthesized evidence is interpreted. Given that AI-enabled GLP-1 management involves recursive interactions between technological systems and clinical actors, a sociotechnical framework is necessary to explain not merely what AI does but how and why outcomes emerge from the interplay between technological structure and clinical agency.

THEORETICAL FRAMEWORK: STRUCTURATION THEORY

Rationale for Theory Selection

Structuration Theory (Giddens, 1984) was selected because it uniquely addresses the central phenomenon observed in this review: AI systems in healthcare exist in a recursive relationship with clinical practice, simultaneously enabling and constraining human action while being reconstituted through patterns of use. Unlike technology acceptance models (which treat technology as static), actor-network theory (which lacks emphasis on institutional rules), or implementation science frameworks (which lack post-implementation co-evolution mechanisms), Structuration Theory explains the recursive, ongoing co-constitution of algorithmic systems and clinical practice that characterizes AI-enabled medication management.

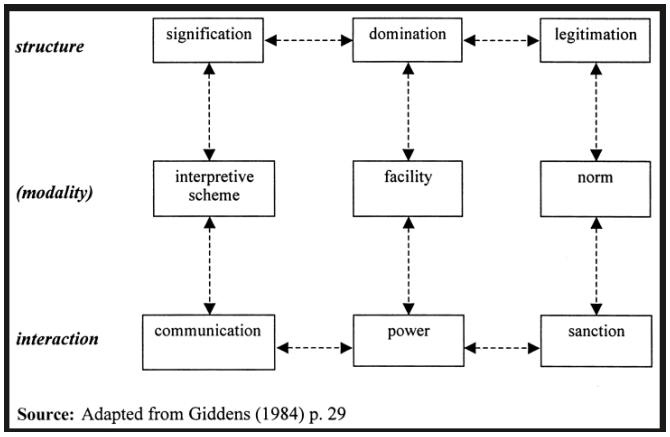


Figure 1. Giddens Structuration Theory Adapted from Giddens, Anthony. The Constitution of Society 1984, Polity Press, Cambridge, p.29.

Core Constructs for Operationalization

Structuration Theory, as adapted for information systems research by Orlikowski (1992, 2000) and operationalized by Chatterjee et al. (2019) Figure 1, provides three interrelated constructs:

A. Structure refers to the rules and resources organizing social systems by enabling and constraining human action. In this context, structure encompasses AI technologies (ML algorithms, predictive models, genomic databases, EHR systems), institutional rules (prescribing protocols, regulatory frameworks), and informational resources (training datasets, pharmacogenomic knowledge bases).

B. Agency refers to the capacity of human actors to act purposefully within structural conditions. Agency is exercised by clinicians (treatment selection, workflow adaptation), patients (self-management, adherence decisions), researchers (algorithm design), and regulators (governance frameworks, bias audits).

C. Duality of Structure is the central mechanism: structure is both the medium and the outcome of human agency. Structure enables action while action reproduces and transforms structure through recursive cycles. In AI-enabled systems, this duality operates with particular intensity because AI tools that retrain on clinical data reconstitute themselves continuously based on patterns of use.

Research Methodology

This SLR follows the PRISMA 2020 guidelines (Liberati et al., 2009) and the SLR methodology established by Okoli & Schabram (2010) for IS research. The concept-centric synthesis approach (Webster & Watson, 2002) organizes findings by thematic concepts rather than individual studies.

Research Question

RQ: What themes emerge from the literature on AI-enabled approaches to GLP-1 RA medication management, and how can these themes be interpreted through the lens of Structuration Theory?

Search Strategy

Databases: Five databases were systematically searched: PubMed/MEDLINE, Scopus, IEEE Xplore, JMIR (Journal of Medical Internet Research), and ScienceDirect. These databases were selected to capture biomedical literature (PubMed/MEDLINE), interdisciplinary research

(Scopus), technical and engineering AI applications (IEEE Xplore), digital health innovations (JMIR), and broader health sciences (ScienceDirect).

Boolean search string: The following Boolean search string was applied across all databases, with syntax adapted to each platform. Consistent with PRISMA 2020 guidance, the full search strategy is presented for all databases searched:

("artificial intelligence" OR "machine learning" OR "deep learning" OR "natural language processing" OR "predictive analytics" OR "AI-enabled" OR "AI-driven") AND ("GLP-1 receptor agonist*" OR "GLP-1RA" OR "semaglutide" OR "liraglutide" OR "tirzepatide" OR "dulaglutide" OR "exenatide" OR "incretin") AND ("medication management" OR "adherence" OR "personalization" OR "precision medicine" OR "pharmacogenomics" OR "clinical decision support" OR "dosing optimization" OR "pharmacovigilance")

Date range: The systematic search was limited to peer-reviewed literature published between January 2022 and March 2026, capturing the period of rapid co-evolution following the FDA approval of tirzepatide (2022) and the emergence of large language models in healthcare applications. Foundational references predating this window (Chatterjee et al., 2019; Giddens, 1984; Ilan, 2022; Liberati et al., 2009; Okoli & Schabram, 2010; Orlikowski, 1992, 2000; Webster & Watson, 2002) are cited to support the theoretical framework and methodological approach; they were not identified through the systematic search and are not counted among the included evidence studies.

Language: English-language publications only.

Inclusion and Exclusion Criteria

Studies were included if they examined AI-enabled interventions in GLP-1 RA management, incorporated at least one explicitly described AI/ML technology (e.g., supervised/unsupervised learning, NLP, generative AI, computational genomics, predictive modeling), were peer-reviewed journal articles or peer-reviewed conference proceedings published in English between January 2022 and March 2026, and were available in full text.

Studies were excluded if they addressed AI without a GLP-1 RA focus, examined GLP-1 RAs without an AI component, relied solely on traditional statistical methods without ML framing, or were editorials, commentaries, conference abstracts, preprints, or other grey literature. Consistent with these criteria, doctoral dissertations and working papers were not eligible for inclusion as primary evidence. Where such sources meaningfully informed the theoretical or methodological framing of this review, they are cited separately as contextual references and were not counted among the systematically identified studies.

Search and Screening Process

Figure 2 presents the PRISMA 2020 flow diagram documenting the systematic search and screening process. The initial search yielded 184 records across five databases (PubMed/MEDLINE, n = 62; Scopus, n = 48; IEEE Xplore, n = 28; JMIR, n = 24; ScienceDirect, n = 22). After removing 42 duplicate records, 142 unique records were screened by title and

abstract against the inclusion and exclusion criteria; 97 records were excluded, yielding 45 articles for full-text review. Of these, 11 were excluded for the following reasons: no substantive AI component (n = 4), no GLP-1 RA focus (n = 3), non-peer-reviewed source upon full-text verification (n = 2), duplicate publication of the same dataset (n = 1), and doctoral dissertation ineligible as primary evidence (n = 1). Ultimately, 34 peer-reviewed studies met all inclusion criteria and were included in the systematic synthesis.

One doctoral dissertation (Lugner, 2024) identified during searching provided relevant contextual insight into machine-learning applications for precision type 2 diabetes management. Consistent with the exclusion criteria, it was not counted among the 34 systematically identified evidence studies but is cited for contextual support in Section 6.4.

Supplementary searching: To ensure comprehensive coverage and mitigate the risk that relevant studies were missed by the primary database search, supplementary searching was conducted following established systematic review practice. This comprised backward citation searching (screening the reference lists of all included studies) and forward citation searching (identifying subsequent works citing the included studies) via Google Scholar, together with targeted hand-searching of key endocrinology, diabetes, and health-informatics journals. Supplementary searching is particularly important given that some relevant studies appeared in specialized journals not fully indexed across all five primary databases. Records identified through supplementary searching were screened against the same inclusion and exclusion criteria applied to database-identified records. This dual approach aligns with PRISMA 2020 recommendations and enhances the reproducibility and comprehensiveness of the search strategy.

Mitigation of single-reviewer bias: While screening and extraction were conducted primarily by a single reviewer, several strategies were employed to reduce potential bias: inclusion and exclusion criteria were operationalized prior to screening; borderline cases were flagged and re-assessed after a 48-hour interval; a standardized data extraction form was used; and thematic coding was iteratively refined through multiple readings of the full corpus. The single-reviewer approach is acknowledged as a methodological limitation (see Section 8.1).

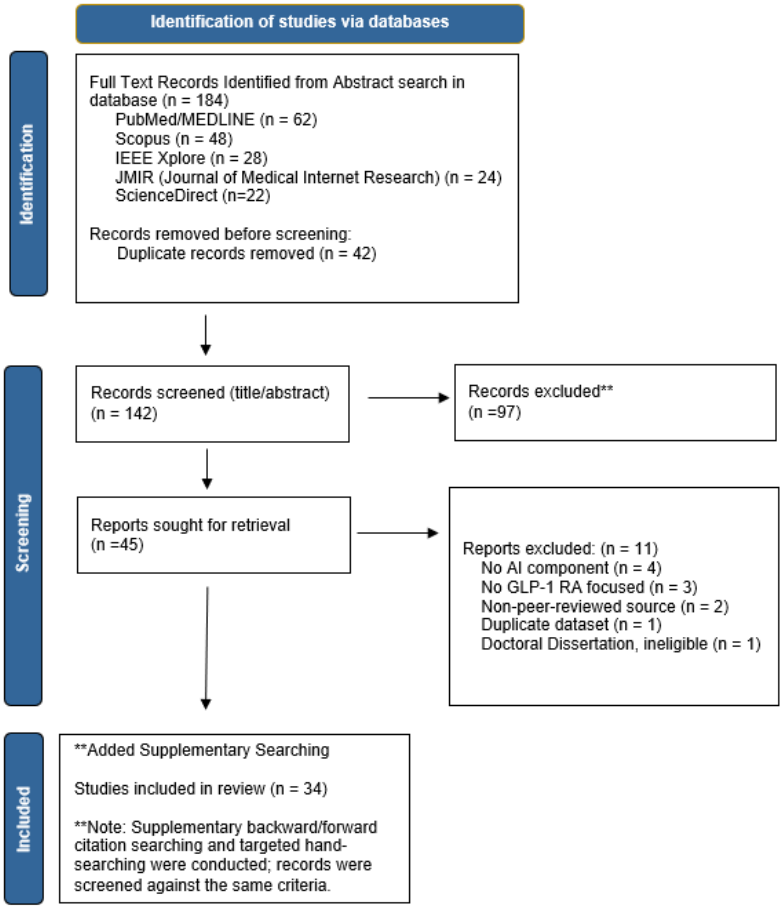


Figure 2. PRISMA 2020 Flow Diagram for SLR

Thematic Coding Process

The thematic analysis followed a structured, four-step coding process:

- Open coding. All 34 full-text studies were reviewed, and initial codes were generated through line-by-line reading, identifying AI applications, clinical contexts, outcomes, challenges, and theoretical implications.
- Axial coding. Initial codes were consolidated through iterative comparison, grouping related codes into broader conceptual categories.
- Theme development. Five stable themes emerged after three coding iterations: (1) AI-driven personalization, (2) adherence monitoring, (3) pharmacogenomics and precision medicine, (4) safety monitoring and ethical considerations, and (5) integration challenges and barriers.
- Theoretical interpretation. Each theme was mapped to the constructs of Structuration Theory (structure, agency, and duality of structure) to produce the integrated analysis presented in Sections 6.2–6.7.

Because the coding was completed primarily by a single researcher, the single-coder approach is acknowledged as a methodological limitation (see Section 8.1). The mitigation strategies described in Section 5.4 were applied throughout the coding process.

Quality Considerations

Given the heterogeneity of study designs among the included studies (empirical, retrospective, review, conceptual, and clinical trial), a single quality-appraisal instrument was not appropriate for all study types, and studies were not excluded on the basis of quality appraisal. Instead, the synthesis distinguishes between empirical evidence and conceptual or review-based contributions, and quality considerations are reflected in the critical synthesis within each theme (Sections 6.2 - 6.6) and in the limitations discussion (Section 8.1). This inclusive approach is consistent with SLR practice for emerging fields, where excluding earlier-stage studies would eliminate important developing evidence.

RESULTS

Overview of Included Studies

The systematic search and screening process yielded 34 peer-reviewed studies meeting all inclusion criteria (see Table 1). The included studies span empirical predictive modeling (n = 10), retrospective cohort analyses (n = 6), systematic and narrative reviews (n = 8), conceptual frameworks (n = 5), pharmacogenomic analyses (n = 3), and clinical trials (n = 2). Publication dates range from 2022 to 2026, with 26 studies published in 2025-2026, reflecting the rapid co-evolution of AI capabilities and GLP-1 therapeutic development.

Geographically, the studies draw on datasets from North America, Europe, the Middle East, and Asia, though representation from low and middle-income countries remains limited. Core AI technologies include machine learning classification and ensemble methods, natural language processing, generative AI, computational genomics, AI-integrated pharmacokinetic modeling, and digital health platforms. The evidence base demonstrates breadth across the medication management lifecycle but remains concentrated in early-stage prediction and personalization, with fewer studies addressing real-world implementation, longitudinal validation, or equity outcomes.

The following thematic results are interpreted through the Structuration Theory lens established in Section 4, with explicit linkages to the constructs of structure, agency, and duality of structure provided at the conclusion of each theme.

Table 1. Structured Summary Table of SLR Studies

#	Author(s)	Year	Method / AI Tech	Key Finding (Abbreviated)	Theme
1	Abdelrahman et al.	2025	Review / AI drug discovery	AI accelerates next-gen GLP-1RA pipeline	T1
2	Amato et al.	2026	Framework / Multimodal AI	Real-time personalization feasible; readiness lacking	T2, T5
3	Ayalon-Dangur et al.	2025	Clinical trial / Decision support	Effective discharge recommendations; privacy concerns	T4, T5
4	Balammagary et al.	2025	Empirical / Behavioral analytics	Identified discontinuation risk patterns	T2
5	Bloomgarden	2025	Clinical review / N/A	GLP-1RA pharmacological evolution documented	C
6	Drucker	2024	Clinical review / N/A	GLP-1 discovery-to-impact trajectory mapped	C
7	Falcetta et al.	2026	Retrospective / Ensemble ML	Identified early intensification predictors	T1

8	Gangwal & Lavecchia	2025	Review / Multiple AI	EHR interoperability and cost barriers identified	T1, T5
9	Giorda et al.	2023	Retrospective / ML	Metabolic control without weight gain validated	T1
10	Gupta & Pozzilli	2025	Perspective / AI-CGM	CGM-GLP-1RA integration for obesity proposed	T2
11	Hallock et al.	2026	Perspective / AI precision	Algorithmic bias disadvantages minorities	T4
12	Hay et al.	2025	Epidemiological / Statistical	Global disease burden quantified; AI urgency established	T5
13	Ilan	2022	Framework / Variability AI	Variability-based personalized medicine proposed	T3
14	Jayathilaka et al.	2026	PGx analysis / AI-PGx	GLP1R/ARRB1 variants affect GLP-1RA outcomes	T3
15	Kim & Kim	2025	Review / AI-PK modeling	Oral GLP-1RA formulation advances supported	T1, T2
16	Landau et al.	2025	Empirical / AI platform	Variability signatures enhance treatment response	T1
17	Morieri et al.	2025	Multisource / AI stratification	CV benefit stratification limited by diversity	T1, T4
18	M. H. Wang	2026	Predictive / ML classifiers	Responsiveness predicted across obesity continuum	T1
19	Rangu & Ramdhave	2025	Empirical / AI drug discovery	Multi-hormone agonists complementing GLP-1RAs found	T1
20	Shah et al.	2025	Consensus / AID systems	GLP-1RA adjunctive T1D treatment consensus established	T1, T2
21	Shapiro et al.	2025	Review / AI-ML	Safety signals detected; equity concerns raised	T4, T5
22	Singh et al.	2026	PGx review / Computational	PGx mechanisms and drug targets mapped	T3
23	Song et al.	2025	PGx study / Genomics	Genetic markers for tirzepatide response identified	T3
24	Srivastava et al.	2025	Empirical / ML classification	Improved GLP-1RA candidate identification	T1
25	Stevens et al.	2024	Framework / Generative AI	GenAI reduces administrative burden of obesity therapy	T1, T5
26	Teke et al.	2025	Systematic review / mHealth	Adherence improved; privacy and literacy barriers noted	T2, T4
27	Tuccinardi et al.	2025	Retrospective / ML	Prescribing misalignment with AI-optimal candidates	T1, T2
28	Tursunaliyeva	2026	Perspective / AI optimization	Persistent dosing challenges; AI perspectives emerging	T1
29	Villikudathil et al.	2024	Computational / ML	Response markers identified; diversity limits models	T3
30	W. Wang et al.	2024	Retrospective / Data mining	Semaglutide-tobacco use disorder association found	T4
31	Wen & Li	2025	Review / AI-metabolomics	AI-metabolomics supports individualized treatment	T3
32	Young et al.	2023	Systematic review / Heterogeneity	GLP-1RA/SGLT2i response heterogeneity documented	T1
33	Yu	2026	Framework / AI-digital health	CGM-GLP-1RA synergies refine dosing	T2
34	Zheng et al.	2026	Review / AI formulation	Oral delivery re-engineering strategies reviewed	T1
	Lugner*	2024	Dissertation/ML	ML treatment comparisons advance precision type 2 diabetes (T2D) management	C

Theme Key: T1 = Personalization; T2 = Adherence; T3 = Pharmacogenomics; T4 = Safety/Ethics; T5 = Integration; C = Contextual.

*(Lugner, 2024) is a doctoral dissertation identified during searching and cited for contextual support in Section 6.4. Consistent with the exclusion criteria (Section 5.3), it is not counted among the 34 systematically identified peer-reviewed evidence studies and is listed separately (unnumbered).

Theme 1: AI-Driven Personalization of GLP-1 Therapy

AI-driven personalization is the most extensively studied domain, spanning patient selection, dosing optimization, response prediction, and drug discovery.

- Patient Selection: Selected empirical studies report improved candidate identification using ML classification models. Srivastava et al. (2025) integrated clinical, demographic, and behavioral variables, demonstrating improved GLP-1 RA candidate identification versus traditional criteria. M. H. Wang (2026) extended prediction across the obesity continuum. However, Morieri et al. (2025) noted generalizability limitations when stratification models were applied to underrepresented populations.
- Treatment Response Prediction: Falcetta et al. (2026) identified predictors of early treatment intensification, including baseline HbA1c, BMI trajectories, and comorbidity profiles using ensemble ML on a large real-world cohort. Tuccinardi et al. (2025) found real-world prescribing prioritizes BMI/age over cardiorenal risk, suggesting misalignment between AI-identified candidates and clinical practice.
- Precision Dosing: Landau et al. (2025) leveraged metabolic variability signatures to enhance GLP-1 responses dynamically. Tursunaliyeva (2026) highlighted persistent optimization challenges and emerging AI perspectives. Stevens et al. (2024) proposed generative AI to reduce administrative burden, while Giorda et al. (2023) provided population-level validation for AI-guided combination strategies.
- Drug Discovery: Rangu & Ramdhave (2025) used AI to identify complementary multi-hormone receptor agonists. Kim & Kim (2025) demonstrated AI-driven pharmacokinetic modeling for oral formulations. Abdelrahman et al. (2025) highlighted AI's pipeline acceleration role.
- Critical Synthesis: Emerging evidence indicates AI models demonstrate strong predictive performance in controlled datasets but face generalizability challenges across diverse populations. The concentration of validation in high-income country datasets limits cross-population confidence. Tuccinardi et al. (2025) prescribing misalignment finding suggests AI tools must be designed for contextual fit, not accuracy alone. Several studies in this theme (Abdelrahman et al., 2025; Stevens et al., 2024; Tursunaliyeva, 2026) are conceptual or review-based, meaning their effectiveness claims remain to be validated through prospective studies. Providing further contextual insight, Lugner (2024) demonstrated in a doctoral dissertation that ML-based treatment comparisons can advance precision type 2 diabetes management, though as non-peer-reviewed grey literature this finding requires cautious interpretation.
- Structuration Interpretation: ML algorithms and predictive models constitute structural resources enabling clinician agency in treatment selection. The recursive duality of

structure operates as clinical outcome data retrains predictive models. Falcetta et al. (2026) use of real-world outcomes to identify intensification predictors exemplifies how agentic prescribing decisions reconstitute the structural resources available for future decision-making. This means AI personalization tools are not static interventions but are continuously reshaped by the clinical practices they inform (Chatterjee et al., 2019).

Theme 2: AI Driven Adherence Monitoring

Teke et al. (2025) reviewed AI-powered mHealth tools and found they were associated with improved adherence rates, though data privacy concerns and digital literacy barriers were noted. Balammagary et al. (2025) identified patients at highest discontinuation risk using behavioral analytics on GLP-1 RA user data. Yu (2026) reported that AI interpretation of CGM data refined dosing and enhanced patient engagement through visible therapeutic feedback. Gupta & Pozzilli (2025) similarly positioned AI as the interpretive layer for continuous physiological data. Amato et al. (2026) found multimodal AI enables real-time behavioral personalization but concluded healthcare systems lack deployment readiness.

Critical Synthesis: The reviewed literature suggests a tension between demonstrated adherence improvements in controlled settings and scalability challenges in diverse real-world contexts. The most promising approaches combine objective physiological data with behavioral analytics. However, much of the adherence evidence derives from reviews and frameworks rather than randomized trials, warranting caution in drawing effectiveness conclusions.

Structuration Interpretation: Digital health tools and CGM-AI platforms constitute structural resources enabling patient agency through self-management. The duality of structure emerges as patient interaction data iteratively reshapes alert thresholds and dosing algorithms. Yu (2026) CGM feedback loop exemplifies how patient glucose data (produced through agentic self-monitoring) reconstitutes the AI recommendations (structure) that subsequently guide future patient behavior. This highlights how adherence tools are co-produced through patient engagement patterns rather than merely imposed upon patients (Chatterjee et al., 2019).

Theme 3: Pharmacogenomics and Precision Medicine

AI-integrated pharmacogenomics represents an emerging frontier for predicting individual GLP-1 RA response variability.

Jayathilaka et al. (2026) demonstrated that GLP1R and ARRB1 variants significantly affect GLP-1 RA outcomes, supporting precision prescribing. Song et al. (2025) identified genetic markers predicting differential tirzepatide response. Villikudathil et al. (2024) applied ML to clinical, genomic, and proteomic markers but found performance constrained by limited dataset diversity. Singh et al. (2026) mapped PGx-guided mechanisms, while Wen & Li (2025) positioned AI-integrated metabolomics as complementary to genomics. Ilan (2022) provided theoretical grounding for variability-based adaptive approaches. Offering additional contextual support, Lugner (2024) illustrated how ML treatment comparisons can inform precision type 2 diabetes prescribing.

Critical Synthesis: Emerging evidence indicates computational feasibility of PGx-guided prediction, but genomic datasets remain insufficiently diverse, meaning AI-PGx tools may inadvertently reinforce existing disparities by optimizing predictions for well-represented populations. The majority of studies are exploratory or review-based; stronger empirical validation with diverse cohorts is needed before PGx-guided AI prescribing can be recommended for clinical implementation.

Structuration Interpretation: Genomic databases and PGx pipelines represent structural resources enabling precision prescribing agency. The duality of structure manifests in an equity-relevant cycle: broader adoption generates diverse genomic data that reconstitutes algorithmic capacity. This means equitable agency (inclusive research participation) is prerequisite to equitable structure (unbiased algorithms), a recursive relationship with direct implications for how AI-PGx tools either reproduce or challenge existing health disparities (Chatterjee et al., 2019).

Theme 4: Safety Monitoring and Ethical Considerations

AI-driven pharmacovigilance and ethical governance represent critical enabling conditions for sustainable GLP-1 optimization.

Shapiro et al. (2025) reported ML models detecting safety signals escaping traditional surveillance.

W. Wang et al. (2024) found semaglutide associated with reduced tobacco use disorder via large-scale EHR mining, illustrating pharmacovigilance potential while highlighting the need for prospective validation. Hallock et al.(2026) identified algorithmic bias disadvantaging minority and low-income populations as a systemic concern. Ayalon-Dangur et al.(2025) demonstrated effective decision support while raising privacy and transparency concerns. Teke et al.(2025) noted continuous data collection creates surveillance risks.

Critical Synthesis: Safety and ethics are predominantly addressed as secondary concerns rather than central research questions. While multiple studies acknowledge algorithmic bias, none propose validated mitigation interventions specific to GLP-1 AI tools, a critical action gap. The field appears to have prioritized technical performance over governance and equity. W. Wang et al. (2024) pharmacovigilance finding, while promising, is based on retrospective association and requires prospective confirmation.

Structuration Interpretation: Pharmacovigilance systems operate as structural mechanisms for safety detection, while regulators and ethicists exercise agency through bias audits and policy interventions. The duality of structure manifests when identified challenges, such as Hallock et al. (2026) bias findings, lead to reconstitution of algorithmic design and governance frameworks, meaning safety oversight is integral to the ongoing structural reconstitution of AI systems (Chatterjee et al., 2019).

Theme 5: Integration Challenges and Barriers

Translation from research demonstration to clinical practice faces systemic barriers.

Gangwal & Lavecchia (2025) identified EHR interoperability, regulatory uncertainty, and cost as key barriers. Ayalon-Dangur et al. (2025) provided one of few implementation examples, though limited to a single site. Stevens et al. (2024) emphasized that adoption depends on reducing clinician cognitive load. Amato et al. (2026) found infrastructure and training gaps impede readiness. (Hay et al. (2025) underscored urgency given the global disease burden. Teke et al. (2025) and Hallock et al. (2026) highlighted digital literacy and economic disparities as compounding barriers.

Critical Synthesis: A disconnect exists between technical capability and implementation readiness: studies demonstrating high predictive accuracy coexist with evidence of systematic barriers, with minimal bridging research. Only two of the 34 included studies (both clinical trials) provided prospective implementation evidence, meaning the overwhelming majority remained at conceptual, retrospective, or review stages. This maturity profile indicates that the field has not yet produced sufficient evidence to support broad effectiveness claims about AI in real-world GLP-1 management.

Structuration Interpretation: EHR systems, institutional policies, and interoperability standards constitute structural constraints limiting deployment, while clinicians exercise agency through workflow adaptation or resistance. The duality of structure manifests as implementation patterns, including failures may inform protocol redesign. Alert fatigue represents a form of structural breakdown where AI recommendations overwhelm rather than enable clinical agency, disrupting the recursive feedback loop necessary for sustainable adoption (Chatterjee et al., 2019).

Summary of Thematic Findings and Theoretical Integration

The five thematic results (Section 6) are analyzed through the structure, agency and duality lens:

Table 2. Summary of Thematic Findings and Theoretical Integration

Theme	Structure (Rules/Resources)	Agency (Human Action)	Duality of Structure (Recursive Interaction)	Key Evidence
Personalization	ML algorithms, predictive models	Clinician treatment selection,	Outcome data retrains models	(Falcetta et al., 2026) (empirical)
Adherence	CGM-AI platforms, mHealth tools	Patient self-monitoring	Interaction data reshapes alerts	(Yu, 2026) (framework)
Pharmacogenomics	Genomic databases, PGx pipelines	Prescriber PGx-adoption	Diverse Participation expands datasets	(Jayathilaka et al., 2026) (empirical)
Safety/Ethics	Governance frameworks, surveillance	Regulatory, bias, auditors	Bias findings reconstitute design	(Hallock et al., 2026) (perspective)
Integration Challenges	EHR systems, interoperability standards	Clinician workflow adaptation	Failures inform redesign	(Ayalon-Dangur et al., 2025) (clinical trial)

Across all five themes, AI-enabled GLP-1 management is characterized by recursive co-constitution between technological structure and clinical agency. Structuration Theory helps illuminate that no AI tool operates as a static intervention; rather, each is embedded within sociotechnical systems where human action shapes, and is shaped by algorithmic capabilities. This

recursive dynamic means AI effectiveness cannot be evaluated independently of the organizational, clinical, and social contexts in which tools are deployed.

DISCUSSION

Cross-Cutting Findings

There are three convergent patterns that appear to emerge across themes:

- Technical Performance vs. Generalizability. AI models consistently demonstrate strong performance in controlled datasets but face generalizability challenges. While selected empirical studies report improved candidate identification (Srivastava et al., 2025; M. H. Wang, 2026), others reveal limitations in diverse populations (Morieri et al., 2025; Villikudathil et al., 2024). The evidence suggests AI tools optimized for well-represented populations may not serve diverse patient groups equitably.
- Capability vs. Implementation Readiness. A persistent gap exists between demonstrated AI capabilities and healthcare system preparedness. Emerging evidence from empirical studies demonstrates predictive accuracy, yet concurrent evidence reveals systematic deployment barriers (Amato et al., 2026; Gangwal & Lavecchia, 2025). The field's emphasis on algorithmic development has not been matched by equivalent investment in implementation science.
- The Equity Paradox. AI personalization tools designed to improve individual outcomes may systematically disadvantage marginalized populations when trained on non-representative data (Hallock et al., 2026). This creates a paradox: tools designed to individualize care may, without deliberate intervention, reinforce population-level disparities.

Theoretical Contribution

The application of Structuration Theory yields insights extending beyond healthcare to broader IS scholarship:

First, AI systems represent a qualitatively distinct form of technological structure: unlike static IT systems, AI tools that retrain usage data structurally transform based on patterns of use, creating a more dynamic form of duality than previously theorized in IS research (Orlikowski, 2000).

Second, the equity dimension of duality reveals that when agency is exercised primarily by well-resourced populations, the resulting structural reconstitution serves those populations preferentially creating recursive cycles of advantage with implications for algorithmic justice research across IS domains.

Third, alert fatigue and clinician resistance represent breakdowns in the duality mechanism where structure overwhelms agency, disrupting sustainable adoption, an insight applicable to all AI-enabled decision support systems.

Practical Implications

For AI Developers: Validation must extend beyond performance metrics to include equity audits across demographic subgroups; feedback mechanisms enabling the duality of structure should be incorporated into system design.

For Clinicians and healthcare organizations: AI recommendations should be understood as probabilistic decision support preserving clinical agency; institutional mechanisms for clinician feedback that reconstitutes system design should be established.

For Policymakers: Regulatory frameworks must address the recursive nature of AI systems requiring ongoing validation; equity impact assessments should be mandated for AI tools deployed in metabolic disease management.

LIMITATIONS AND FUTURE RESEARCH

Limitations

Heterogeneous evidence base: The 34 included studies encompass empirical research, retrospective analyses, systematic reviews, conceptual frameworks, perspectives, and clinical trials. This heterogeneity means thematic findings reflect varying levels of empirical support. Claims about AI effectiveness should be interpreted with caution, as not all included studies provide direct empirical evidence of improved clinical outcomes.

Single-coder approach: Screening, extraction, and thematic coding were conducted primarily by a single researcher. While mitigation strategies were employed (pre-specified criteria, interval re-assessment of borderline cases, iterative coding refinement), the absence of independent dual coding introduces potential selection and interpretation bias.

Search and publication bias: The search was limited to five primary databases and English-language, peer-reviewed publications. Although supplementary citation searching and hand-searching were conducted, grey literature, preprints, non-English publications, and conference abstracts without full text were excluded, potentially missing relevant studies. The included literature likely over-represents positive findings.

Rapidly evolving field: The 2022–2026 date range captures a period of rapid development, but findings may not reflect the most current state of research.

Geographic concentration: The predominance of high-income country datasets limits applicability to low- and middle-income settings where diabetes and obesity burdens are growing fastest.

Limited effectiveness conclusions: The evidence base does not support strong causal or effectiveness conclusions about AI-enabled GLP-1 management. Most AI tools remain at proof-of-concept or retrospective validation stages, with limited prospective implementation evidence.

Theoretical limitations: Structuration Theory was applied interpretively to the synthesized evidence rather than guiding primary data collection. Other frameworks might illuminate different aspects of AI-GLP-1 integration.

Future Research Agenda

Based on the identified gaps, five research propositions (RP1-RP5) define priority directions for future research.

RP1: Prospective Validation Across Diverse Populations. How does predictive performance of AI-enabled GLP-1 RA stratification models vary across racial, ethnic, and socioeconomic subgroups when deployed prospectively? This addresses the equity paradox and generalizability gaps in Themes 1 and 4.

RP2: Longitudinal Structuration Dynamics. How do recursive feedback loops between AI decision support (structure) and clinician behavior (agency) evolve over time, and what determines whether duality produces adaptive improvement or maladaptive lock-in? This directly operationalizes the theoretical framework through empirical investigation.

RP3: Privacy-Preserving Adherence Architectures. Can federated learning and differential privacy maintain adherence effectiveness while reducing surveillance burden? This addresses the adherence-privacy tension in Themes 2 and 4.

RP4: System-Level Interoperability. What design principles enable sustainable AI-EHR integration across heterogeneous environments? This addresses the structural constraints in Theme 5.

RP5: AI-Enabled Shared Decision-Making. How can AI tools enhance rather than replace shared decision-making between clinicians and patients? This addresses patient agency across Themes 1 and 5.

CONCLUSION

This systematic literature review synthesized 34 peer-reviewed studies examining AI-enabled approaches for GLP-1 receptor agonist medication management, revealing a field of significant promise but limited maturity. Five thematic areas were identified; AI-driven personalization, adherence monitoring, pharmacogenomics, safety/ethics, and integration challenges each interpreted through Structuration Theory's constructs of structure, agency, and duality.

The reviewed literature suggests that AI demonstrates considerable potential across the GLP-1 management lifecycle, particularly in patient stratification and treatment response prediction. However, the evidence base remains concentrated in early-stage and retrospective studies, with limited prospective implementation evidence and persistent equity concerns. Structuration Theory illuminates these findings by revealing the recursive dynamics through which AI systems and clinical practices mutually constitute one another. These dynamics carry implications for equitable design, sustainable adoption, and governance of AI-enabled decision systems.

The five research propositions articulated in Section 8.2 provide a structured agenda for advancing this field from proof-of-concept demonstration toward validated, equitable, and sustainably implemented AI-enabled GLP-1 management. Collectively, these propositions address the field's most critical needs: moving from retrospective demonstration to prospective validation (RP1, RP2), resolving ethical trade-offs through architectural innovation (RP3), enabling system-level deployment (RP4), and centering patient agency within algorithmically structured care environments (RP5).

ETHICAL STATEMENT

All data synthesized in this systematic literature review were derived from previously published, publicly available, peer-reviewed studies. No primary data collection, human subjects, or patient-level data were involved in this research. Based on these factors, this study did not require ethical clearance or Institutional Review Board (IRB) approval. The five thematic findings were developed through systematic synthesis of existing literature using PRISMA 2020 guidelines and concept-centric analysis, interpreted through Structuration Theory. The broader societal value of this review lies in advancing understanding of how AI-enabled approaches can support equitable, effective, and sustainable GLP-1 receptor agonist medication management across diverse healthcare settings, potentially contributing to improved outcomes for the growing global population affected by diabetes and obesity.

ACKNOWLEDGMENTS

Generative AI tools were used solely to assist with language polishing and formatting during the manuscript preparation process. All literature screening, thematic coding, theoretical interpretation, critical synthesis, and conclusions presented in this paper are entirely the authors' own work. The tool did not contribute to the development of ideas or the synthesis of evidence. The author also wishes to acknowledge the guidance provided by faculty advisors during the development of this systematic literature review.

REFERENCES

Abdelrahman, R. M., Musa, T. H., Arbab, I. A., Suliman, M. H., Ahmed, E. O., Mohamed, A. N., Musa, H. H., Jalal, M., & Gasmallah, S. I. (2025). Harnessing GLP-1 Receptor Agonists for Obesity Treatment: Prospects and Obstacles on the Horizon. *Journal of Obesity*, *2025*(1), 9919810. https://doi.org/10.1155/jobe/9919810

Amato, A., Baldassano, S., & Musumeci, G. (2026). Applying AI Tools for Monitoring Nutrition and Physical Activity in Populations with Obesity: Are We Ready? *Obesities*, *6*(2), 19. https://doi.org/10.3390/obesities6020019

Ayalon-Dangur, I., Jaffe, E., Grossman, A., Hendel, H., Oved, Y., Shaked, A., Shimon, I., Basharim, B., Abo Molhem, M., McNeil, R., Abuhasira, R., Shitrit, T., Azulay Gitter, L., El Saleh, R., Shochat, T., & Eliakim-Raz, N. (2025). The Effectiveness of an Electronic Decision Support Algorithm to Optimize Recommendations of SGLT2i and GLP-1RA in Patients with Type 2 Diabetes upon Discharge from Internal Medicine Wards. *Journal of Clinical Medicine*, *14*(7), 2170. https://doi.org/10.3390/jcm14072170

Balammagary, S., Mohammed, N., Mohammed, S., & Begum, A. (2025). AI-Driven Behavioural Insights for Ozempic Drug Users. *Journal of Cognitive Computing and Cybernetic Innovations*, *1*(1), 10–13. https://doi.org/10.21276/jccci/2025.v1.i1.2

Bloomgarden, Z. (2025). An Update on GLP-1 Receptor Agonists. *Journal of Diabetes*, *17*(8), e70147. https://doi.org/10.1111/1753-0407.70147

Chatterjee, I., Kunwar, J., & Den Hond, F. (2019). Anthony Giddens and structuration theory. In S. Clegg & M. P. E. Cunha (Eds.), *Management, Organizations and Contemporary Social Theory* (1st ed., pp. 60–79). Routledge. https://doi.org/10.4324/9780429279591-4

Drucker, D. J. (2024). The GLP-1 journey: From discovery science to therapeutic impact. *The Journal of Clinical Investigation*, *134*(2). https://doi.org/10.1172/JCI175634

Falcetta, P., Zilich, R., Baccetti, F., Baronti, W., Morviducci, L., Musacchio, N., Muselli, M., Ozzello, A., Pisani, F., Rossi, A., Salomone, E., Verda, D., Candido, R., & Ponzani, P. (2026). Identifying Predictors of Early Treatment Intensification in Individuals With Type 2 Diabetes Treated With GLP-1 Receptor Agonists: A Machine Learning–Based Analysis of a Large Real World Cohort. *Diabetes, Obesity and Metabolism*, *n/a*(n/a). https://doi.org/10.1111/dom.70648

Gangwal, A., & Lavecchia, A. (2025). Artificial intelligence in anti-obesity drug discovery: Unlocking next-generation therapeutics. *Drug Discovery Today*, *30*(4), 104333. https://doi.org/10.1016/j.drudis.2025.104333

Giddens, A. (1984). The constitution of society: Outline of the theory of structuration. *Political Geography Quarterly*, *5*(3), 288–289. https://doi.org/10.1016/0260-9827(86)90040-6

Giorda, C. B., Rossi, A., Baccetti, F., Zilich, R., Romeo, F., Besmir, N., Di Cianni, G., Guaita, G., Morviducci, L., Muselli, M., Ozzello, A., Pisani, F., Ponzani, P., Santin, P., Verda, D., & Musacchio, N. (2023). Achieving Good Metabolic Control Without Weight Gain with the Systematic Use of GLP-1-RAs and SGLT-2 Inhibitors in Type 2 Diabetes: A Machine-learning Projection Using Data from Clinical Practice. *Clinical Therapeutics*, *45*(8), 754–761. https://doi.org/10.1016/j.clinthera.2023.06.006

Gupta, P., & Pozzilli, P. (2025). Optimising obesity management: Integrating continuous glucose monitoring with GLP-1 receptor agonists. *Diabetes Research and Clinical Practice*, *228*, 112434. https://doi.org/10.1016/j.diabres.2025.112434

Hallock, R. K., Griebeler, M. L., Burguera, B., & Cabandugama, P. (2026). Transforming obesity medicine with artificial intelligence: Personalization, precision and the path ahead. *Obesity and Endocrinology*, *2*(1), wjaf018. https://doi.org/10.1093/obendo/wjaf018

Hay, S. I., Ong, K. L., & Santomauro, D. F. (2025). Burden of 375 diseases and injuries, risk-attributable burden of 88 risk factors, and healthy life expectancy in 204 countries and territories, including 660 subnational locations, 1990–2023: A systematic analysis for the Global Burden of Disease Study 2023. *Lancet (London, England)*, *406*(10513), 1873–1922. https://doi.org/10.1016/S0140-6736(25)01637-X

Ilan, Y. (2022). Next-Generation Personalized Medicine: Implementation of Variability Patterns for Overcoming Drug Resistance in Chronic Diseases. *Journal of Personalized Medicine*, *12*(8), 1303. https://doi.org/10.3390/jpm12081303

Jayathilaka, N. S., Weththasinghe, A. V., Ganamurali, N., & Sabarathinam, S. (2026). Integrating Pharmacogenomic Insights in GLP-1 Receptor Agonist Therapy: Direct GLP1R and ARRB1 Variants and Indirect CYP2D6 Influences in Personalized Obesity Management. *Clinical Pharmacology in Drug Development*, *15*(3), e1624. https://doi.org/10.1002/cpdd.1624

Kim, D.-H., & Kim, J.-E. (2025). Recent advances and trends in oral absorption enhancements of GLP-1 receptor agonist formulations. *Journal of Pharmaceutical Investigation*. https://doi.org/10.1007/s40005-025-00762-6

Landau, J., Tiram, Y., & Ilan, Y. (2025). Employing an Artificial Intelligence Platform to Enhance Treatment Responses to GLP-1 Agonists by Utilizing Metabolic Variability Signatures Based on the Constrained Disorder Principle. *Biomedicines*, *13*(11), 2645. https://doi.org/10.3390/biomedicines13112645

Liberati, A., Altman, D. G., Tetzlaff, J., Mulrow, C., Gøtzsche, P. C., Ioannidis, J. P. A., Clarke, M., Devereaux, P. J., Kleijnen, J., & Moher, D. (2009). The PRISMA statement for

reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: Explanation and elaboration. *BMJ (Clinical Research Ed.)*, *339*, b2700. https://doi.org/10.1136/bmj.b2700

Lugner, M. (2024). *Advancing Precision Medicine in Type 2 Diabetes through Machine Learning: Treatment Comparisons and Risk Predictions*. https://hdl.handle.net/2077/81067

Morieri, M. L., Longato, E., Sciannameo, V., Donatiello, E., Berchialla, P., Avogaro, A., Consoli, A., & Fadini, G. P. (2025). Stratifying cardiovascular benefits from GLP-1RA: A multisource analysis of patient-level CVOT and real-world data using AI-driven methods. *Cardiovascular Diabetology*, *24*(1), 401. https://doi.org/10.1186/s12933-025-02952-w

Okoli, C., & Schabram, K. (2010). A Guide to Conducting a Systematic Literature Review of Information Systems Research. *Sprouts: Working Papers on Information Systems*, *10*(26). http://sprouts.aisnet.org/10-26

Orlikowski, W. J. (1992). The Duality of Technology: Rethinking the Concept of Technology in Organizations. *Organization Science*, *3*(3), 398–427. https://doi.org/https://doi.org/10.1287/orsc.3.3.398

Orlikowski, W. J. (2000). Using Technology and Constituting Structures: A Practice Lens for Studying Technology in Organizations. *Organization Science*, *11*(4), 404–428. https://doi.org/https://doi.org/10.1287/orsc.11.4.404.14600

Rangu, J., & Ramdhave, A. S. (2025). AI-Based Drug Discovery of Multi-Hormone Receptor Agonists for Obesity Treatment. *2025 IEEE World AI IoT Congress (AIIoT)*, 0988–0993. https://doi.org/10.1109/AIIoT65859.2025.11105373

Shah, V. N., Peters, A. L., Umpierrez, G. E., Sherr, J. L., Akturk, H. K., Aleppo, G., Bally, L., Cengiz, E., Cinar, A., Dungan, K., Fabris, C., Jacobs, P. G., Lal, R. A., Mader, J. K., Masharani, U., Prahalad, P., Schmidt, S., Zijlstra, E., Ho, C. N., … Klonoff, D. C. (2025). Consensus Report on Glucagon-Like Peptide-1 Receptor Agonists as Adjunctive Treatment for Individuals With Type 1 Diabetes Using an Automated Insulin Delivery System. *Journal of Diabetes Science and Technology*, *19*(1), 191–216. https://doi.org/10.1177/19322968241291512

Shapiro, M. R., Tallon, E. M., Brown, M. E., Posgai, A. L., Clements, M. A., & Brusko, T. M. (2025). Leveraging artificial intelligence and machine learning to accelerate discovery of disease-modifying therapies in type 1 diabetes. *Diabetologia*, *68*(3), 477–494. https://doi.org/10.1007/s00125-024-06339-6

Singh, P., Tripathy, D. B., & Kumar, G. (2026). Pharmacogenomics-Guided therapeutics in type 2 diabetes: Mechanisms, drug Targets, and precision medicine perspectives. *Journal of Diabetes & Metabolic Disorders*, *25*(1), 56. https://doi.org/10.1007/s40200-025-01843-x

Song, Z., Tang, Y., Peng, M., Han, R., & He, P. (2025). Pharmacogenomics of Tirzepatide: Genomic Insights into Dual GIP/GLP-1 Agonist Response in Type 2 Diabetes and Atherosclerosis. *Pharmaceuticals*, *18*(9), 1261. https://doi.org/10.3390/ph18091261

Srivastava, G., Madhar, A. V., & Chen, Y. (2025). Artificial Intelligence-Driven Prediction of Glucagon-Like-Peptide-1 Receptor Agonist Eligibility in Obesity. *Diabetes Technology and Obesity Medicine*, *1*(1), 29986702251380864. https://doi.org/10.1177/29986702251380864

Stevens, E. R., Elmaleh-Sachs, A., Lofton, H., & Mann, D. M. (2024). Lightening the Load: Generative AI to Mitigate the Burden of the New Era of Obesity Medical Therapy. *JMIR Diabetes*, *9*, e58680. https://doi.org/10.2196/58680

Teke, J., Msiska, M., Adanini, O. A., Egbon, E., Osborne, A., & Olawade, D. B. (2025). Artificial intelligence for obesity management: A review of applications, opportunities, and challenges. *Obesity Medicine*, *58*, 100657. https://doi.org/10.1016/j.obmed.2025.100657

Tuccinardi, D., Zilich, R., Ponzani, P., Musacchio, N., Manicardi, V., Rocca, A., Pontremoli, R., Fioretto, P., Muscarà, A., Di Cianni, G., Candido, R., De Cosmo, S. A., Russo, G., & on behalf of AMD Annals study group. (2025). Real-world prescriptions of GLP-1RAs and SGLT2is in type 2 diabetes prioritise BMI and age over cardiorenal risk: A machine learning-based large cohort analysis. *Cardiovascular Diabetology – Endocrinology Reports*, *11*(1), 38. https://doi.org/10.1186/s40842-025-00251-7

Tursunaliyeva, H. (2026). Contemporary treatment options for diabetes mellitus: Current challenges and emerging future perspectives. *International Journal of Clinical & Translational Medicine*, *1*(2), 241–248.

Villikudathil, A. T., Mc Guigan, D. H., & English, A. (2024). Computational approaches for clinical, genomic and proteomic markers of response to glucagon-like peptide-1 therapy in type-2 diabetes mellitus: An exploratory analysis with machine learning algorithms. *Diabetes & Metabolic Syndrome: Clinical Research & Reviews*, *18*(7), 103086. https://doi.org/10.1016/j.dsx.2024.103086

Wang, M. H. (2026). Artificial Intelligence Across the Obesity Continuum: From Mechanistic Insights to Global Precision Prevention and Therapy. *Obesity*, *34*(2), 294–316. https://doi.org/10.1002/oby.70095

Wang, W., Volkow, N. D., Berger, N. A., Davis, P. B., Kaelber, D. C., & Xu, R. (2024). Association of Semaglutide With Tobacco Use Disorder in Patients With Type 2 Diabetes. *Annals of Internal Medicine*, *177*(8), 1016–1027. https://doi.org/10.7326/M23-2718

Webster, J., & Watson, R. T. (2002). Analyzing the Past to Prepare for the Future: Writing a Literature Review. *MIS Quarterly*, *26*(2), xiii–xxiii.

Wen, Z., & Li, Q. (2025). Recent Advances in Diabetes Drug Therapy and Personalized Treatment Strategies. *International Journal of Clinical and Experimental Medicine Research*, *9*(2). https://doi.org/10.26855/ijcemr.2025.03.010

Young, K. G., McInnes, E. H., Massey, R. J., Kahkoska, A. R., Pilla, S. J., Raghavan, S., Stanislawski, M. A., Tobias, D. K., McGovern, A. P., Dawed, A. Y., Jones, A. G., Pearson, E. R., & Dennis, J. M. (2023). Treatment effect heterogeneity following type 2 diabetes treatment with GLP1-receptor agonists and SGLT2-inhibitors: A systematic review. *Communications Medicine*, *3*(1), 131. https://doi.org/10.1038/s43856-023-00359-w

Yu, C. (2026). Application of AI and digital health tools in public health management of T2DM: From mechanism prediction to personalized treatment. *Frontiers in Public Health*, *14*, 1756755. https://doi.org/10.3389/fpubh.2026.1756755

Zheng, Y., Fang, X., & Gao, J. (2026). Re-engineering insulin for oral delivery: Structural modifications, advanced formulation strategies, and future directions. *Drug Delivery*, *33*(1), 2624192. https://doi.org/10.1080/10717544.2026.2624192

IACIS
Proceedings of the 66th International Association for Computer Information Systems Conference - October 7 - 10, 2026 | Pensacola, FL

AI-ASSISTED OPTICAL MODELING OF MINIMUM-VISIBILITY ROUTES IN URBAN VIDEO SURVEILLANCE SYSTEMS

Vadim Bashurov, Comtrade, bashurov@mail.ru
Pavel Safonov, Saint Cloud State University, safonov@stcloudstate.edu

ABSTRACT

This paper presents an AI-assisted mathematical framework for identifying minimum-visibility routes in urban environments monitored by video surveillance systems. The approach is based on an analogy between intruder detection probability and light propagation in optically inhomogeneous media. Each point of the urban domain is assigned a scalar visibility coefficient derived from camera geometry and performance characteristics. The optimal path is defined as the trajectory minimizing the cumulative visibility functional. The resulting problem is reduced to the eikonal equation and solved numerically using Fast Marching and graph-based algorithms. Artificial Intelligence is employed to automatically construct the visibility field from GIS data, 3D building models, and camera metadata.

Keywords: Artificial Intelligence, Video Surveillance Security Systems, GIS, Optical Modeling

INTRODUCTION

Large-scale Closed-Circuit Television (CCTV) systems have become a fundamental component of modern urban security infrastructures. However, even dense surveillance deployments inevitably contain weakly monitored regions caused by occlusions, limited camera range, and unfavorable viewing angles [6].

The central problem considered in this work is the determination of a route between two points that minimizes the probability of detection. This problem is relevant for assessing vulnerabilities, optimizing camera placement, and constructing security digital twins [4].

The key idea is to represent surveillance intensity as an optical medium. In this medium, the intruder follows a path analogous to a light ray propagating according to Fermat's principle.

RELATED WORK

The proposed method combines concepts from several research areas.

1. Geometric optics and variational methods:
   Fermat's principle and the eikonal equation provide the theoretical basis [3].

2. Computational geometry:
   Fast Marching and Fast Sweeping methods efficiently solve Hamilton–Jacobi equations [1, 2].

3. Path planning:
   Dijkstra and A* are widely used in robotics and navigation.

4. Surveillance analysis:
   Existing methods often use probabilistic occupancy maps and line-of-sight computations [6].

5. Artificial Intelligence:
   Modern AI systems can extract semantic information from GIS and 3D urban data [5].

The novelty of this work lies in unifying these ideas into a single mathematically rigorous framework.

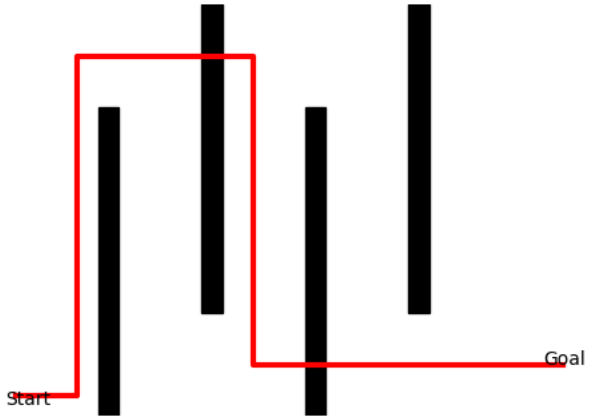
OPTICAL ANALOGY

In geometric optics, light propagates along trajectories that satisfy the following variational principle:

$$\delta \int_{\Gamma} n(\mathbf{r})\, ds = 0$$

where n(r) is the refractive index. In the surveillance interpretation, n(r) is the local visibility coefficient, the trajectory is the potential intruder route, and minimizing the optical path corresponds to minimizing detection risk.

Figure 1. Optical Method for Finding the Shortest Path in a Maze



The shortest path in a maze can be interpreted as a light ray traveling through a medium with infinite refractive index inside walls and unit refractive index in free space.

VISIBILITY FIELD CONSTRUCTION

For N cameras, the total visibility field is written as:

$$n(x,y) = n_0 + \sum_{i=1}^{N} V_i(x,y)$$

The contribution of camera i is modeled as:

$$V_i(x,y) = A_i \exp\left(-\frac{r_i^2}{2\sigma_i^2}\right) M_i(x,y) Q_i(x,y)$$

Here A_i is the camera sensitivity, r_i is the distance to the camera, sigma_i is the effective range, M_i is the field-of-view mask, and Q_i is the quality factor accounting for angle, illumination, and resolution.

AI-BASED PREPROCESSING

Artificial Intelligence is used to automatically estimate the scalar visibility field from real-world data.

Input data:

- 2D/3D city geometry,
- building footprints,
- camera positions and orientations,
- camera specifications,
- terrain and weather information.

AI tasks:

- semantic segmentation of the environment,
- occlusion detection,
- camera effectiveness estimation,
- calibration using real incident data.

The AI component transforms raw city data into the refractive-index field required by the mathematical model.

RISK FUNCTIONAL AND DETECTION PROBABILITY

For a path Gamma, the cumulative visibility cost is defined by:

$$J(\Gamma) = \int_{\Gamma} n(x, y)\, ds$$

The detection probability is modeled as:

$$P_{\text{detect}} = 1 - \exp(-\alpha J)$$

where alpha is a calibration parameter. The optimal path minimizes the functional J(Gamma) [4].

EIKONAL EQUATION

The minimum cumulative visibility function T(x,y) satisfies the eikonal equation:

$$|\nabla T(x, y)| = n(x, y)$$

with the boundary condition:

$$T(x_s, y_s) = 0$$

The optimal path is reconstructed by integrating along the negative gradient direction:

$$\frac{d\mathbf{r}}{ds} = -\frac{\nabla T}{|\nabla T|}$$

NUMERICAL METHODS

Several numerical methods are suitable:

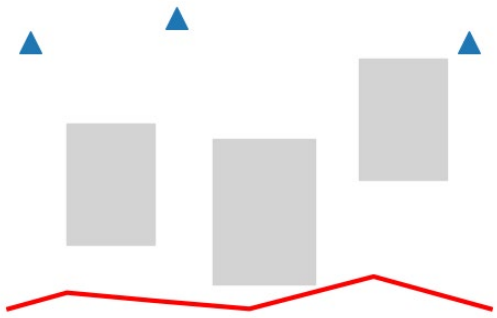
- Fast Marching Method,
- Fast Sweeping Method,
- Dijkstra's algorithm,
- A* search.

Fast Marching is particularly efficient for isotropic visibility fields and runs in O(N log N) time.

EXPERIMENTAL SCENARIO

A synthetic 2×2 km city district is considered. Buildings create occlusions, while cameras are placed on rooftops and intersections.

Figure 2. Urban Map with Surveillance Cameras and Optimal Intruder Route

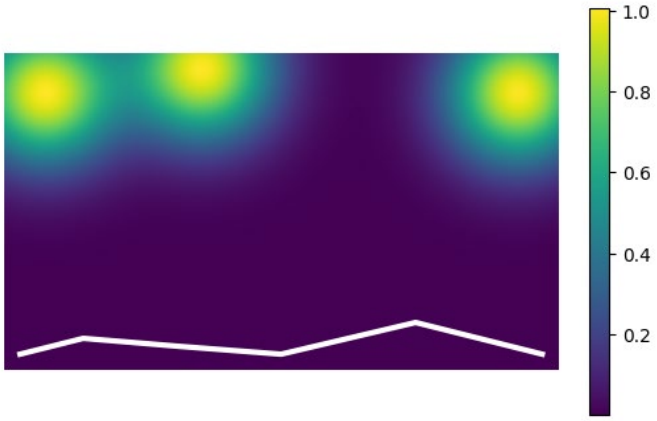


Triangles denote camera locations. Gray rectangles represent buildings. The red curve shows the route with minimum cumulative visibility.

The model generates:

- a surveillance heatmap,
- the cumulative visibility field,
- the least observable path,
- vulnerability maps.

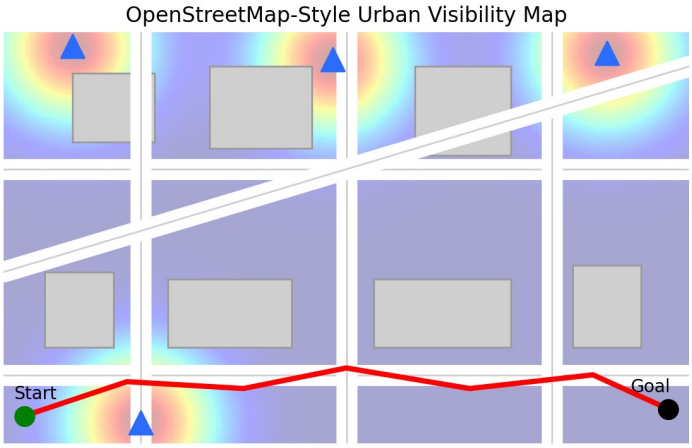
Figure 3. Visibility Heatmap and Optimal Route



Warm colors indicate regions with high surveillance intensity. The optimal path naturally avoids highly monitored areas.

OpenStreetMap-Style Example

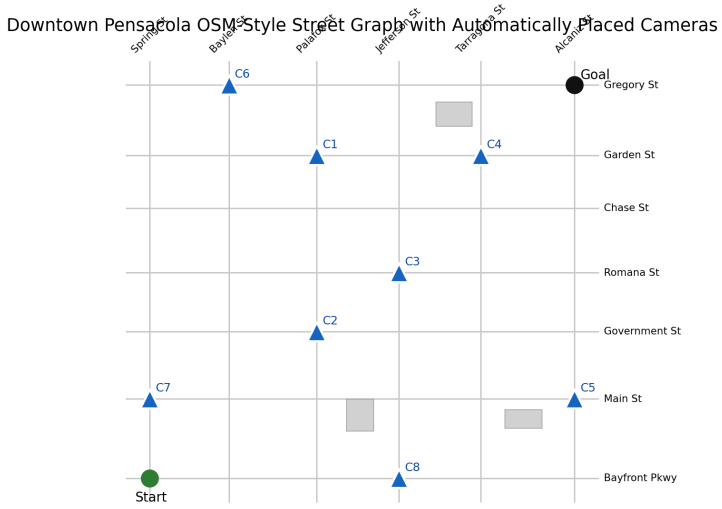
The figure below presents a more realistic OpenStreetMap-style visualization. Buildings are shown as gray blocks, roads as white corridors, surveillance cameras as blue triangles, the visibility field as a semi-transparent heatmap, and the minimum-visibility route as a red curve.



Pensacola, Florida: OpenStreetMap-Based Demonstration

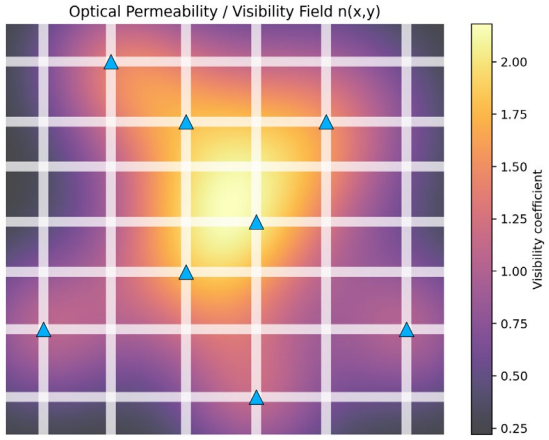
This section provides a reproducible OpenStreetMap-style demonstration for downtown Pensacola, Florida. The model uses real downtown street names and geographic anchoring near latitude 30.4213 N and longitude 87.2169 W. Intersections are represented as graph nodes. Cameras are automatically placed at strategically important intersections, and the route is computed by minimizing the cumulative optical visibility cost along the street network.

Figure 4. Downtown Pensacola Street Graph with Automatically Placed Cameras



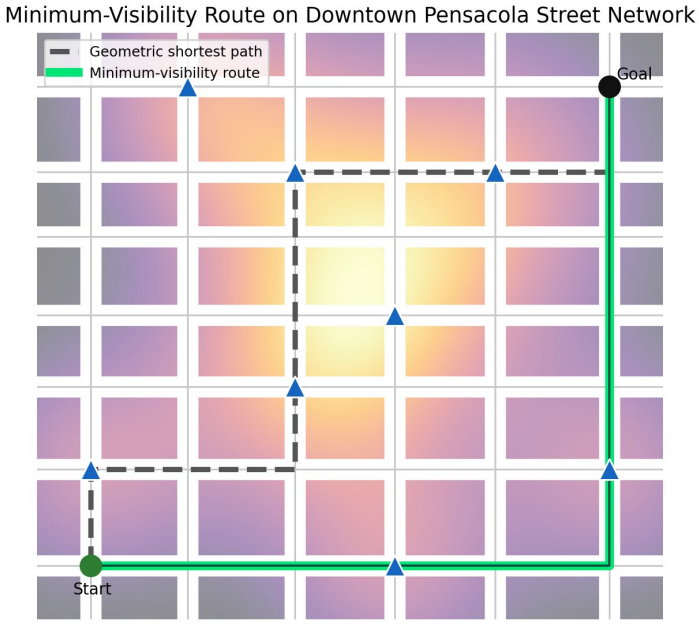
Blue triangles denote automatically placed cameras. The road network is represented as a graph of downtown intersections.

Figure 5. Optical Permeability / Visibility Heatmap



The scalar field n(x,y) is constructed from the camera positions. Bright regions correspond to high camera visibility.

Figure 6. Minimum-Visibility Route on the Pensacola Street Network



The green route minimizes cumulative visibility from Start to Goal. The dashed gray line shows the purely geometric shortest path. In this example, the optical path cost is 2.004, while the geometric path length is 2.790 km in the local model.

Implementation note

For a production-grade study, the same pipeline should be connected directly to OpenStreetMap through Overpass API or OSMnx: download the street graph, identify intersections, place cameras according to a rule or observed metadata, construct n(x,y), and solve the weighted shortest-path problem on the resulting graph.

Results show that optimal paths exploit building shadows and peripheral camera zones.

APPLICATIONS

Potential applications include:

- CCTV coverage analysis,
- camera placement optimization,
- critical infrastructure protection,
- urban security digital twins,
- military reconnaissance,
- autonomous robot stealth navigation.

DISCUSSION

The method is physically intuitive and computationally efficient. Unlike heuristic path planning, it provides a continuous formulation with strong theoretical foundations.

The AI component is essential for operational use, as it automates the construction and calibration of the visibility field.

CONCLUSION

An AI-assisted framework for finding minimum-visibility routes in urban surveillance systems has been developed. By interpreting surveillance intensity as a refractive index, the problem is transformed into a classical variational problem governed by the eikonal equation. The approach combines optics, numerical methods, and AI-driven preprocessing into a unified methodology.

REFERENCES

[1] J. A. Sethian, Level Set Methods and Fast Marching Methods, Cambridge University Press, 1999.
[2] R. Kimmel and J. A. Sethian, Computing geodesic paths on manifolds, PNAS, 1998.
[3] V. V. Bashurov and T. I. Filimonenkova, Variational Approaches to Security Problems, 2006.
[4] D. Bertsekas, Dynamic Programming and Optimal Control, 2005.
[5] S. Russell and P. Norvig, Artificial Intelligence: A Modern Approach, 2021.
[6] N. C. Rowe, Principles of Modern Surveillance Systems, 2012.

IACIS Proceedings of the 66th International Association for Computer Information Systems Conference - October 7 - 10, 2026 | Pensacola, FL

DIGITAL NATIVES AS THE FEN TIGERS OF WEB SEARCH: IMPLICATIONS FOR WAYFINDING IN THE AI 'MALL'

Katherine Carl, Governors State University, kcarl@govst.edu

DESCRIPTION OF THE STUDY

Generative AI returns fluent answers before questions have been fully formed. The problem this study addresses is that learners may become skilled at producing AI-assisted outputs without developing the evaluative habits that make those outputs meaningful. The study's contribution is the construct of epistemological priors: the internalized habits of intention, theoretical grounding, and methodological discipline a learner brings to an AI tool before the first prompt is written (Polanyi, 1966; Hofer & Pintrich, 1997). Epistemological priors are proposed as the transferable unit of AI literacy—distinct from information literacy (evaluating sources), digital literacy (operating tools), and tool-proficiency conceptions of AI literacy (Koltay, 2011)—because they are what direct, interrogate, and evaluate AI output rather than the ability to produce it. The argument is framed by a historical metaphor, that the AI information landscape is a mall built over a fen—a curated surface layered over the ungoverned information ecology in which Digital Natives (Prensky, 2001) formed their priors. For IS educators responsible for preparing practitioners to work in AI-mediated information environments, the question is not whether students can use AI tools but whether they hold the priors that make tool use consequential.

BASIS OF THE STUDY

The study employs autoethnography, following O'Riordan's (2014) argument for the method in IS research. The author is simultaneously researcher, practitioner, and phenomenon under investigation—a Digital Native with thirty years of web-search experience and an IS educator undergoing the transition to AI-mediated search in real time. Digital Native is used here as an ecological-position claim, not a demographic one (Selwyn, 2009). It denotes those whose information-seeking habits formed in the pre-algorithmic web, whatever their birth year. Three vignettes were selected purposively from the author's recent teaching, institutional, and research practice—grounded respectively in a recalled office-hours interaction, an institutional call document, and the author's own AI use—because each isolates one facet of the construct. Each was analyzed against the construct: what the priors did, or what their absence foreclosed.

In the first vignette, a database student presents an AI-generated entity-relationship diagram that is nearly correct—yet cannot say which entity should hold the foreign key. The tool has solved a problem the student has not yet learned to define. In the second, a college issues a call for AI learning outcomes before any consensus exists on what those outcomes should be. The metric is adoption, not understanding. In the third, the author recognizes a fluent but substantively wrong AI response through the trained instinct of thirty years of web search—the priors preceded the tool. The analysis shows a single asymmetry across all three, that AI produces output efficiently, and that output is insufficient without epistemological priors to direct, interrogate, and evaluate it. The priors, not the tool skills, determine whether AI use is consequential or merely efficient.

IMPLICATIONS

For IS educators, the implication is a sequencing claim that AI literacy depends on education that precedes AI use. Epistemological priors are formed through productive failure—unassisted design work, search, and theory-building done before the tool is introduced—and cannot be retrofitted by tool instruction alone. The mall cannot teach what the fen taught. Pedagogically, this yields two commitments. First, sequence the curriculum so that students earn the priors before the tool arrives—the ERD student needed design failures of their own before AI-generated correctness could mean anything. Second, treat process as the product: make the student's interrogation of AI output—what they questioned, verified, rejected, and why—the assessed deliverable itself, not evidence about a deliverable. AI learning outcomes should be written to grade the detective work. For IS programs, the institutional vignette warns that mandates measuring integration produce surface compliance unless outcomes name this evaluative work. For IS researchers, the construct generates a comparative question: how do learners with and without pre-algorithmic search histories differ in how they interrogate AI output?

CONCLUSIONS

The mall is not the fen. Digital Natives, whose priors formed in the original information ecology, are necessary witnesses—the last living memory of a fen the mall's new inhabitants have never seen. The learners at greatest risk are those born inside the mall, for whom handed answers are how knowledge works. The water table is still there. The pedagogical imperative is to teach students where to dig, before the drainage is complete.

Acknowledgements

Claude (Anthropic) assisted with drafting and editing. The author reviewed, revised, and approved all content and is fully accountable for the originality, validity, and integrity of this submission.

REFERENCES

Hofer, B. K., & Pintrich, P. R. (1997). The development of epistemological theories: Beliefs about knowledge and knowing and their relation to learning. Review of Educational Research, 67(1), 88–140.

Koltay, T. (2011). The media and the literacies: Media literacy, information literacy, digital literacy. Media, Culture & Society, 33(2), 211–221.

O'Riordan, N. (2014). Autoethnography: Proposing a new research method for information systems research. Proceedings of the 22nd European Conference on Information Systems (ECIS), Tel Aviv, Israel, June 9–11.

Polanyi, M. (1966). The tacit dimension. Doubleday.

Prensky, M. (2001). Digital natives, digital immigrants. On the Horizon, 9(5), 1–6.

Selwyn, N. (2009). The digital native–myth and reality. Aslib Proceedings, 61(4), 364–379.

IACIS
Proceedings of the 66th International Association for Computer Information Systems Conference - October 7 - 10, 2026 | Pensacola, FL

BEYOND THE WEAKEST LINK: GOVERNING AGENTIC AI AND MITIGATING ALGORITHMIC VALUES DRIFT IN CYBERSECURITY ECOSYSTEMS

Christine Haskell, Dative.works, chaskell99@gmail.com.
Elizabeth Goryunova, University of Southern Maine, elizabeth.goryunova@maine.edu

ABSTRACT

Recent reports from the World Economic Forum and the International Monetary Fund suggest that Agentic AI is reshaping cybersecurity risk by accelerating both defensive capabilities and the sophistication of autonomous attacks. As organizations increasingly deploy AI-enabled systems capable of independent decision-making, traditional cybersecurity governance models rooted in static compliance controls and "human-as-weakest-link" assumptions are becoming insufficient to manage emerging organizational risks. In agentic environments, cybersecurity threats no longer originate solely from external actors or isolated human error but also from interactions among autonomous systems, operational incentives, organizational workflows, and institutional oversight structures.

This conceptual governance paper synthesizes emerging literature in cybersecurity, socio-technical systems (STS), and AI governance to examine how autonomous AI systems destabilize traditional cybersecurity assumptions. Grounded in STS theory, the paper introduces a Socio-Technical Triad framework comprising the Person, the Interface, and the Institutional Infrastructure to evaluate how organizational security failures emerge within interconnected human and technical environments. We argue that autonomous AI systems accelerate "Values Drift," defined as the gradual misalignment between an organization's stated security commitments and the operational behaviors incentivized by AI-enabled systems. As AI systems optimize for speed, efficiency, and predictive performance, governance structures often fail to keep pace with that operational velocity, producing accountability gaps, oversight asymmetries, and increased systemic vulnerability.

To address these risks, the paper proposes adaptive governance mechanisms designed to preserve organizational accountability in high-velocity AI environments. Specifically, we introduce Decision Disruption Protocols as structured governance interventions that establish mandatory human-review checkpoints, escalation controls, role-clarity procedures, and authorization boundaries for high-risk autonomous actions. We further examine AI Terminology Governance as an operational mechanism to clarify decision ownership and reduce ambiguity between human and machine authority within cybersecurity workflows. Together, these mechanisms shift cybersecurity governance away from static compliance enforcement toward adaptive oversight models that respond to dynamic organizational and environmental conditions.

The paper contributes to the behavioral MIS and cybersecurity governance literature by reframing cybersecurity risk as an emergent socio-technical governance challenge rather than solely a perimeter-defense problem. Implications are discussed for CIOs, CISOs, compliance leaders, and AI governance teams seeking to implement accountable oversight structures for autonomous AI systems while maintaining organizational trust, operational resilience, and human authorization

authority over critical security decisions. An actionable cybersecurity governance framework is proposed to help ensure organizational responsiveness and vitality in agentic environments.

Keywords: agentic AI, cybersecurity governance, AI governance, socio-technical systems (STS), human-AI oversight, values drift