DIGITAL 2026

Advances on Societal Digital Transformation

ISBN: 978-1-68558-400-9

July 5 - 9, 2026

Nice, France

DIGITAL 2026 Editors

Sibylle Kunz. IU Internationale Hochschule, Germany

Claudia Heß. IU Internationale Hochschule, Germany

DIGITAL 2026

Forward

The Advances on Societal Digital Transformation 2026 (DIGITAL 2026) held between July 5th -9th, 2026 continued a series of international events covering a large spectrum of topics related to digital transformation of our society.

The society is continuously changing at a rapid pace under digital transformation. Taking advantage of a solid transformation of digital communication and infrastructures, and with great progress in AI (Artificial Intelligence), IoT (Internet of Thinks), ML (Machine Learning), Deep Learning, Big Data, Knowledge acquisition and Cognitive technologies, almost all societal areas are redefined.

Transportation, Buildings, Factories, and Agriculture are now a combination of traditional and advanced technological features. Digital citizen-centric services, including health, well-being, community participation, learning and culture are now well-established and set to advance further on.

As counter-effects of digital transformation, notably fake news, digital identity risks and digital divide are also progressing in a dangerous rhythm, there is a major need for digital education, fake news awareness, and legal aspects mitigating sensitive cases.

We take here the opportunity to warmly thank all the members of the DIGITAL 2026 technical program committee, as well as all the reviewers. The creation of such a high-quality conference program would not have been possible without their involvement. We also kindly thank all the authors who dedicated much of their time and effort to contribute to DIGITAL 2026 We truly believe that, thanks to all these efforts, the final conference program consisted of top-quality contributions. We also thank the members of the DIGITAL 2026 organizing committee for their help in handling the logistics of this event.

We hope that DIGITAL 2026 was a successful international forum for the exchange of ideas and results between academia and industry and for the promotion of progress in the field of societal digital transformation. We also hope that Nice provided a pleasant environment during the conference and everyone saved some time for exploring this beautiful city.

DIGITAL 2026 Chairs

DIGITAL 2026 Steering Committee

Sibylle Kunz, IU Internationale Hochschule, Germany
Fernando Joaquim Lopes Moreira, Universidade Portucalense, Portugal
Júlio Monteiro Teixeira, Universidade Federal de Santa Catarina, Brazil
Claudia Heß, IU Internationale Hochschule, Germany
Christelle Scharff, Pace University, USA

DIGITAL 2026 Publicity Chairs

Laura Garcia, Universidad Politécnica de Cartagena, Spain
Lorena Parra Boronat, Universitat Politecnica de Madrid, Spain
José Miguel Jiménez, Universitat Politecnica de Valencia, Spain

DIGITAL 2026

Committee

DIGITAL 2026 Steering Committee

Sibylle Kunz, IU Internationale Hochschule, Germany
Fernando Joaquim Lopes Moreira, Universidade Portucalense, Portugal
Júlio Monteiro Teixeira, Universidade Federal de Santa Catarina, Brazil
Claudia Heß, IU Internationale Hochschule, Germany
Christelle Scharff, Pace University, USA

DIGITAL 2026 Publicity Chairs

Laura Garcia, Universidad Politécnica de Cartagena, Spain
Lorena Parra Boronat, Universitat Politecnica de Madrid, Spain
José Miguel Jiménez, Universitat Politecnica de Valencia, Spain

DIGITAL 2026 Technical Program Committee

Qammer Abbasi, University of Glasgow, UK
Aaroud Abdessadek, Chouaib Doukkali University, Morocco
Kawiwat Amnatchotiphan, Thai-Nichi Institute of Technology, Thailand
Daniel Amo Filvà, La Salle, Ramon Llull University, Spain
Mariia Andriyiivna Nazarkevych, Lviv Polytechnic National University, Ukraine
Mohamed Basel Almourad, Zayed University, United Arab Emirates
Faouzi Benzarti, Engineering School of Tunis (ENSIT), Tunisia
Louiza Bouallouche-Medjkoune, University of Bejaia, Algeria
An Braeken, Vrije Universiteit Brussel, Belgium
Roberta Calegari, University of Bologna, Italy
Sayali Chavan, Ulster University, UK
Jundong Chen, University of Wisconsin-Parkside, USA
Siming Chen, Fudan University, China
Marta Chinnici, ENEA, Rome, Italy
Javier Cifuentes-Faura, University of Murcia, Spain
María Elicia Cortés Cediel, Universidad Complutense de Madrid, Spain
José Paulo Cravino, Universidade de Trás-os-Montes e Alto Douro, Portugal
Mohamed Dahmane, CRIM – Computer Research Institute of Montreal, Canada
Babu R. Dawadi, Tribhuvan University, Nepal
Burcu Demirdöven, Pamukkale University, Turkey
Noella Edelmann, University for Continuing Education Krems, Austria
Birgitta Dresp-Langley, Centre National de la Recherche Scientifique UMR 7357, France
Zakaria Abou El Houda, University of Montreal, Canada
Zühal Erden, Atilim University, Turkey
Nikos Fakotakis, University of Patras, Greece
Hamed Farbeh, Amirkabir University of Technology (Tehran Polytechnic), Iran

Filipe Portela, University of Minho, Portugal
Rui Quaresma, Universidade de Évora, Portugal
Achim Rettberg, University of Applied Sciences Hamm-Lippstadt / Carl von Ossietzky University Oldenburg, Germany
Manuel Pedro Rodríguez Bolívar, University of Granada, Spain
Amirreza Rouhi, Drillmec SPA / Politecnico di Milano, Italy
John Sabatini, The University of Memphis, USA
Christelle Scharff, Pace University, USA
Inga Schlömer, IU Internationale Hochschule, Germany
Razak Seidu, Norwegian University of Science and Technology (NTNU), Norway
Atriya Sen, University of New Orleans, USA
Ecem Buse Sevinç Çubuk, Adnan Menderes University, Turkey
Pietro Siciliano, Institute for Microelectronics and Microsystems (IMM-CNR), Lecce, Italy
Anthony Simonofski, Namur Digital Institute | Université de Namur, Belgium
Rosario Soria, IAG Finance, New Zealand
Abel Suing, Universidad Técnica Particular de Loja, Ecuador
Do Duy Tan, Ho Chi Minh City University of Technology and Education (HCMUTE), Vietnam
Camel Tanougast, University of Lorraine, Metz, France
Evelyne Tauchnitz, University of Lucerne, Switzerland
Giacomo Valente, University of L'Aquila, Italy
Washington Velasquez Vargas, Escuela Superior Politécnica del Litoral, Ecuador
Massimo Villari, University of Messina, Italy
Ke Wang, University of North Carolina at Charlotte, USA
Li Wang, University of North Carolina at Chapel Hill, USA
Olaf Witkowski, Cross Labs | University of Tokyo | Tokyo Institute of Technology, Japan
Marcin Wozniak, Silesian University of Technology, Poland
Seyed Yahya Nikouei, Kar Global, USA
Guillaume Zambrano, Nimes University, France
Tengchan Zeng, Virginia Tech, USA
Chi Zhang, The University of Glasgow, UK
Zheng Zhao, Synopsys Inc., USA

Table of Contents

When AI Joins the Team: Measuring Perceived Productivity Gains from Generative AI in Project Management

Ole Mehlert, Sibylle Kunz, Claudia Hess
IT & Engineering / Management
IU International University of Applied Sciences
Erfurt, Germany
e-mail: ole.mehlert@iu-study.org; sibylle.kunz@iu.org; claudia.hess@iu.org;

***Abstract*— Amid persistent productivity challenges in Project Management, Generative Artificial Intelligence is emerging as a powerful tool for administrative and information-intensive tasks. However, empirical evidence of its actual productivity impact remains fragmented and lacks a unified measurement framework. This paper addresses this foundational gap by proposing a multidimensional instrument to measure perceived productivity. An initial field test with project management professionals served to examine the instrument's practical usability, provide an initial reliability check, and generate preliminary descriptive findings illustrating its potential value.**

Keywords- project management; generative artificial intelligence; productivity.

I. INTRODUCTION

Much of Project Management (PM) work involves administrative and information-intensive tasks like status reporting, stakeholder communication, risk analysis, and the preparation of decision briefs [1]. Performed under time pressure, resource constraints, and increasing complexity, these duties create persistent productivity challenges, especially in synthesizing and communicating large volumes of information efficiently. Generative Artificial Intelligence (GenAI) directly targets these PM activities with its ability to summarize text, structure data, and prepare reports. However, its use introduces significant uncertainties, including data protection, user competency, and the risk of "hallucinated" outputs. Furthermore, AI-generated content often requires verification and rework, which can offset anticipated efficiency gains. This creates a central challenge for organizations: developing ways to assess GenAI's perceived and true net benefits in practical PM contexts.

Initial evidence suggests GenAI is used for simple, standardized PM tasks rather than complex decision-making [2][3]. While efficiency gains are frequently reported, empirical research remains fragmented, often focusing on isolated tasks and lacking multidimensional measures that capture quality or cognitive relief alongside time savings. Furthermore, it remains unclear how perceived gains are moderated by a project manager's experience level.

To address these gaps, this research develops a multidimensional measurement instrument for perceived productivity gains in PM. We demonstrate its application through a field test with 97 PM professionals, offering preliminary descriptive findings, an initial reliability indication, and groundwork for future validation and causal research.

The paper is structured as follows. Section II reviews the theoretical context of GenAI in PM and productivity measurement. Section III details the methodology and instrument development. Section IV presents the analysis of the survey data from the field test. Then, Section V discusses results and highlights areas for future research. Finally, Section VI summarizes the paper's core contributions.

II. THEORETICAL BACKGROUND

A. Patterns of GenAI Adoption in PM

Recent studies examine the use of GenAI in PM and its productivity implications. What began as exploratory experimentation is evolving into adoption across industries. A global study by the Project Management Institute (PMI) in January 2024 surveyed 500 project professionals across 18 industries in 12 countries, all reporting GenAI use in their projects [4]. The study highlights its growing relevance in PM and identifies four user types:

- **Explorers** use GenAI selectively and experimentally in isolated project tasks.
- **Adapters** apply GenAI regularly to support recurring and moderately complex activities.
- **Innovators** use GenAI extensively across tasks to enhance efficiency and creativity.
- **Trailblazers** embed GenAI strategically and systematically into complex and high-impact tasks.

The categorization emphasizes that GenAI adoption in PM is not uniform but varies in scope, depth, and organizational integration. More intensive adoption is linked to stronger positive assessments of GenAI's impact on key project-level outcomes, such as scope, schedule, cost, and quality management [4]. A similar pattern emerges at the individual level: respondents with high GenAI usage report better outcomes in productivity, collaboration, creativity, problem-solving, and overall effectiveness than those with more limited, exploratory use.

Although the PMI study distinguishes multiple performance dimensions, these are captured broadly. This creates an opportunity to further specify and operationalize productivity aspects for concrete PM activities. Such a systematic investigation requires two key components: a classification framework for GenAI use cases and a model for measuring their productivity impact.

 ISBN: 978-1-68558-400-9

B. Classifying GenAI Use Cases in PM

To classify GenAI use cases in PM, established project management paradigms were first evaluated. Categorizing cases by project lifecycle phases proved unsuitable for cross-industry analysis due to the lack of a standardized phase model across PM bodies (e.g., PMI, PRINCE2), industries, and methodologies (e.g., Agile, Waterfall). Comparability is further limited as activities, such as risk management occur throughout the lifecycle, blurring phase boundaries. An alternative grouping by functional category [5][6], was also considered but found either too vague or too industry-specific. Instead, the PMI's AAA framework was selected, which organizes GenAI use cases in PM into three tiers, based on their increasing complexity and the corresponding need for human intervention [7].

- **Automation** represents the lowest level of complexity, where GenAI independently executes well-defined, rule-based tasks. The human role is typically limited to initiating the process and reviewing the standardized output, such as automated meeting minutes or status reports.
- **Assistance** describes tasks of medium complexity where GenAI acts as a support tool, providing drafts, analyses, or recommendations. The human practitioner then evaluates, contextualizes, and makes the final decision. Examples include generating initial cost estimates or performing risk data analysis.
- **Augmentation**, the highest level of complexity, involves a co-creative partnership in which GenAI enhances human cognitive capabilities. GenAI is used to simulate complex scenarios, identify novel patterns, or generate strategic options, while the human retains full responsibility for the final decision. Augmentation includes strategic tasks like managing project portfolios.

C. Measuring Dimensions for AI Productivity in PM

To measure the impact of GenAI usage, [8] provides a structure. Kunz & Hess argue that both analytical and generative AI can improve productivity across three primary dimensions: **time, cost, and quality**. Applying specific measures to these dimensions allows AI-related productivity gains to be calculated. These dimensions directly align with the foundational "iron triangle" of PM, which conceptualizes project success as the balance of these three constraints [4]. By examining the perceived effects of GenAI through the lens of the iron triangle, we can systematically assess its *direct* contribution to the core goals of PM as understood by practitioners.

However, a comprehensive assessment of GenAI's benefits requires moving beyond the traditional constraints of the iron triangle. The literature indicates that GenAI usage also exerts *indirect* effects on several strategic dimensions by fostering long-term capabilities. First, it enhances **innovation and creativity**, particularly during ideation phases where it can rapidly generate diverse concepts [9]. Second, it optimizes **decision-making** by enabling faster, data-driven analysis, with studies showing that human-AI collaboration yields the highest improvements in decision quality [10][11]. Finally, GenAI strengthens **knowledge management** by improving the creation, structuring, and application of organizational knowledge, thereby reducing information silos and freeing up employees for higher-value learning tasks [12][13].

D. Synthesis and Identification of the Research Gap

A systematic analysis of current GenAI research reveals two critical gaps. First, studies of GenAI's impact are predominantly qualitative. The research scope is fragmented, examining either discrete, simple tasks (e.g., generating meeting minutes), or ill-defined categories (e.g., "planning") [2][3]. This lack of a unified, quantitative framework for measuring productivity provides a partial and skewed view of GenAI's utility.

Beyond *what* is measured, a second critical gap concerns *who* is using the technology. First, digital competency is a key determinant [14]. Counterintuitively, less competent users tend to delegate more complex tasks (e.g., risk management) to AI and assign it a more active, rather than merely supportive, role in the problem-solving process. Second, PM experience also shapes the professional use of GenAI [15]. Seasoned project managers apply GenAI to strategic and complex problems, while novices use it for foundational tasks and skill development. This suggests experienced professionals can better integrate new technologies to exploit their full potential.

Consequently, the central challenge is the lack of a multidimensional instrument that measures perceived gains while also accounting for the moderating effects of user experience. Therefore, empirically investigating GenAI's impact requires first developing this tool.

E. Research Questions

To address the identified research gap, this paper is guided by the following research questions:

RQ1 (Methodological): How can the perceived productivity gains from using GenAI in PM be operationalized into a multidimensional measurement instrument?

RQ2 (Exploratory): Based on this new instrument, what are the initial descriptive findings regarding perceived productivity gains across different PM tasks, and how do these perceptions appear to differ based on user proficiency?

III. METHODOLOGY

This section details the systematic process used to develop, refine, and conduct an initial field test of the measurement instrument.

A. Measurement Development

The measures for this study were developed based on the theoretical foundation from Section II. The operationalization process involved two key principles: first, grounding our measures in the existing literature, and second, ensuring items were specific and behavior-oriented rather than abstract. The survey instrument was developed and administered in German. The English item formulations presented in this paper are translations prepared for reporting purposes only. The following sections detail the operationalization of the different constructs.

1) Self-Assessment of GenAI Proficiency in PM: To assess GenAI proficiency, a single-choice question was developed based on the PMI's four-part user typology presented in Section II.A. The defining characteristics were adapted and formulated as first-person statements, forming a hierarchy of GenAI integration:

- **Explorer: "**I use generative AI sparingly in my projects for selected, straightforward tasks (e.g., brainstorming, text creation, and summarization)."
- **Adapter: "**I use generative AI regularly in my projects for recurring, moderately complex tasks (e.g., creating project plans, risk assessments)."
- **Innovator: "**I use generative AI extensively in my projects across a wide range of tasks and optimize its application (e.g., project budgeting, resource planning)."
- **Trailblazer: "**I use generative AI strategically and systematically in my projects for complex and critical tasks (e.g., forecasting project outcomes, automated decision-making)."

Respondents were asked to select the statement that best described their usage pattern.

2) Operationalizing GenAI Usage: To enable a comparative analysis of GenAI usage across different industries and methodologies, a standardized framework was required. This study adopted the AAA-Framework (Section II.B) for its clear, hierarchical structure and its focus on the nature of use rather than specific contexts. To operationalize this framework, a representative PM task was selected for each category. These tasks serve as proxies to empirically measure and compare GenAI application across different levels of complexity.

For the automation category, we selected *creating meeting minutes*. In this process, GenAI autonomously transcribes discussions and generates structured summaries with key decisions and action items, requiring only a final human review. This choice is validated by its explicit citation as an automation example by PMI [7], and its successful application in PM [14].

We chose *stakeholder communication* for the assistance category. GenAI is used to draft communications like emails or status reports based on project data. The process is characterized by a medium level of complexity, as it requires significant human intervention for contextual adaptation and refinement. Its selection is also supported by its documented success as a GenAI application in PM [14] [16].

For the augmentation category, we selected *project portfolio decision*. In this strategic task, GenAI analyzes project proposals to generate data-driven recommendations. The process requires deep human-AI collaboration, as experts, typically senior management, must provide contextual knowledge and retain final judgment. This synergy on a high-stakes decision makes it a prime example of augmentation.

3) Operationalizing Productivity Gains: We conceptualize perceived productivity gains from GenAI as a multidimensional latent construct. It is latent because it is not directly observed but inferred from indicators like time savings and quality improvements as discussed in Section II.B. It is multidimensional because it encompasses several distinct yet related value contributions, not just a single effect. Consequently, a single percentage-based estimate (e.g., "To what extent has GenAI increased your productivity?") is inadequate, as it wrongly assumes a unidimensional effect and forces respondents to aggregate heterogeneous benefits, thus reducing measurement validity.

Therefore, our study assesses perceived gains indirectly using multiple task-specific indicators for each theoretical dimension, measured on a five-point Likert scale ranging from strong disagreement to strong agreement. This approach ensures contextual validity while consistently measuring the underlying construct. The operationalization is illustrated by the automation task. Following the introductory prompt "Thinking back to the creation of meeting minutes: How helpful was the use of generative AI in this context for you?"., the task-specific items displayed in Table I were presented. The same approach was taken for the assistance and the augmentation tasks.

TABLE I. AUTOMATION: CREATION OF MEETING MINUTES

Task: Meeting Minutes	Dimension
I produce more precise and less error-prone meeting minutes that clearly summarize the discussed content and assign tasks to the appropriate persons.	Quality
I am able to reduce the effort required for manual documentation and follow-up, thereby decreasing personnel and resource expenditure.	Cost
I provide meeting minutes significantly faster after the meeting and require less time for their preparation.	Time
I structure meeting minutes more consistently and make them easier to retrieve, thereby improving knowledge sharing.	Knowledge Management
I automatically identify recurring topics or patterns and thereby receive impulses for improvements.	Innovation & Creativity
I document decisions more precisely, making them easier to understand through clear summaries and supporting arguments.	Decision-Making

B. Survey Design

The questionnaire was structured according to established survey design principles to ensure a logical flow and a positive respondent experience [17]. Questions were arranged in order of increasing difficulty, and the layout was organized into distinct thematic blocks. The survey was presented in the following order: (1) an introductory title page, (2) the block of screening questions, (3) a section for the self-assessment of GenAI usage in PM, (4) a block for demographic information, (5) the main section measuring the construct of perceived productivity gains, (6) an open-ended question about further applications beyond those mentioned, and (7) a concluding thank-you page. Note that for the main assessment in step (5), and to ensure relevance, only participants who confirmed prior GenAI use for each task (which was briefly described) proceeded to the evaluation.

Prior to the main launch, three external participants evaluated the survey for clarity and usability. Based on this pretest, several formulations and scale labels were refined for clarity, and the layout was adjusted.

C. Study Design and Data Collection

To collect data for the initial field application of our newly developed instrument, a quantitative online survey was conducted. The German-language survey was distributed between November 25 and December 29, 2025, targeting PM professionals with prior experience using GenAI in their work. The average completion time for the questionnaire was 5 minutes and 42 seconds.

Participants were recruited via the professional networking platform LinkedIn. Participation was voluntary and unpaid. This recruitment method was selected due to the platform's large number of thematically relevant groups with a direct professional connection to PM and the use of GenAI, which allowed for targeted recruitment. A key component of the distribution strategy was the dissemination of the survey through the PMI Chapter Germany e.V., reaching over 5,000 members. This collaboration was particularly appropriate, as the present study references PMI frameworks.

D. Sample Composition

The survey was designed to target a specific professional demographic. Eligibility was strictly controlled by the three screening questions, ensuring the sample consisted exclusively of PM professionals located in Germany, Austria, or Switzerland who had prior experience using GenAI for PM tasks. From a gross sample of 113 respondents, 16 were removed by the screening questions, leaving a final net sample of 97 participants for analysis. The sample is predominantly composed of professionals in the earlier stages of their PM careers, with a significant majority (62.9%) reporting five years of experience or less. The most seasoned cohort—those with over a decade of experience—was the least represented, accounting for just 13.4% of the total sample. Table II summarizes the sample characteristics.

TABLE II. SAMPLE CHARACTERISTICS: YEARS OF EXPERIENCE IN PM

Years of Experience in Project Management	**Sample**	
	Frequency N	***Percentage (%)***
< 1 year	12	12.4%
1-5 years	49	50.5%
6-10 years	23	23.7%
11-15 years	4	4.1%
> 15 years	9	9.3%
Total	**97**	**100.0%**

An analysis of the participants' current roles shows a clear predominance of Project Team Members, who constitute nearly half of the sample at 48.45% (n=47). The next most common roles were Project Leader/Manager at 14.43% (n=14) and Project Management Office (PMO) members at 12.37% (n=12). All other roles were represented in single-digit percentages: Sub-project Leaders (6.19%, n=6), Functional/Line Managers with project responsibility and External Consultants/Service Providers (both 5.15%, n=5), Scrum Masters/Agile Coaches and Program/Portfolio Manager (both 3.09%, n=3), and Product Owners (2.06%, n=2). The fact that the 'Other' category was not selected (0.00%) indicates the provided role options comprehensively covered the sample population.

The industry distribution, based on categories from a PMI study [2], shows a heavy concentration in the Information Technology sector, representing 23.7% of all responses. Significant cohorts also came from Aerospace (15.5%) and Consulting (14.4%). Other notable sectors included Automotive, Logistics, and Telecommunications, with minimal representation from remaining industries like healthcare, construction, and pharma. Note that participants were permitted to select multiple industries.

IV. RESULTS

This section presents the descriptive statistics and initial empirical findings from the online survey.

A. Self-Assessed Proficiency

The self-assessment of GenAI proficiency reveals that the majority of users are in the intermediate stages. “Adapters” form the largest cohort at 43.3% (n=42), followed by “Innovators” at 30.9% (n=30). Conversely, early-stage “Explorers” (20.6%, n=20) and expert-level “Trailblazers” (5.2%, n=5) represent the smallest groups. This indicates that while deep, complex usage is still rare, a significant majority of respondents have moved beyond initial experimentation and possess solid experience applying GenAI in PM tasks.

Furthermore, a Spearman's rank correlation analysis revealed a significant positive relationship between years of PM experience and self-assessed GenAI proficiency ($\rho = 0.546$, $p < 0.01$, n=97). Following [18], this correlation strength is interpreted as moderate. This finding indicates that as a project manager’s experience increases, so does their self-reported GenAI proficiency. This suggests that extensive professional experience in PM may foster a more rapid development of competency in applying these AI tools in the PM context.

B. GenAI Usage in PM

GenAI adoption varies significantly across the three tasks, which were selected as proxies to represent different levels of complexity. The analysis shows a clear divide between administrative and strategic tasks. The most common application by a wide margin is the creation of meeting minutes, where 84.5% (n=82) of participants have experience. In stark contrast, adoption rates for more complex functions are significantly lower. The use of GenAI for stakeholder communication drops by more than half to 41.2% (n=40), while its application in supporting project portfolio decisions is even less frequent at 33.0% (n=32).

Notably, four participants reported zero usage across the specific PM tasks listed in the survey. However, these individuals were retained in the final sample. This decision was made because, in a concluding open-ended question, they described other valid applications of GenAI in their PM work that were not covered by the predefined tasks.

C. Perceived Productivity Gains

To assess perceived productivity gains, participants rated the benefit of GenAI for each of the three tasks (Meeting Minutes, Stakeholder Communication, Project Portfolio Decisions) on a 5-point Likert scale across six dimensions: Time, Quality, Cost, Decision-Making, Innovation, and Knowledge Management.

1) Preliminary Instrument Reliability: Before analyzing the mean scores, we assessed the internal consistency of the item sets for each task. A composite reliability score (Cronbach's Alpha) was calculated for the six-item "Productivity Gains" construct for each of the three tasks. The scale for "Meeting Minutes" showed acceptable consistency ($\alpha = 0.704$), while the scales for "Stakeholder Communication" ($\alpha = 0.863$) and "Project Portfolio Decisions" ($\alpha = 0.852$) demonstrated good to excellent consistency, interpreted according to Blanz (2015, p. 256). These results provide an initial indication of the internal consistency of the aggregated perceived productivity measures for each task.

2) Descriptive Findings: Table III shows the resulting mean scores and Standard Deviations (SD) for the three tasks. Across all three tasks analyzed, the perceived benefits were consistently dominated by direct, efficiency-related dimensions. Time emerged as the highest-rated perceived productivity gain. Cost and Quality followed closely, with Cost rated slightly higher for meeting minutes and stakeholder communication, while Quality marginally exceeded Cost for project portfolio decisions. The indirect dimensions (Decision-Making, Innovation, Knowledge Management) consistently received lower scores, although their specific order varied depending on the task.

TABLE III. PERCEIVED PRODUCTIVITY GAINS

Dimension	***Meeting Minutes***		***Stakeholder Communicat.***		***Project Portfolio Decision***	
	Mean (N=82)	***SD***	***Mean (N=40)***	***SD***	***Mean (N=32)***	***SD***
Time	4.33	0.546	4.18	0.594	3.91	0.734
Cost	4.20	0.576	3.97	0.577	3.72	0,.581
Quality	4.07	0.466	3.80	0.723	3.75	0.672
Decision-Making	3.80	0.617	3.47	0.877	3.63	0.707
Knowledge Management	3.68	0.564	3.72	0.784	3.41	0.712
Innovation & Creativity	3.67	0.589	3.63	0.774	3.63	0.660
Composite score	**3.96**	**0.357**	**3.80**	**0.521**	**3.67**	**0.514**

Considering the composite score, the application of GenAI for creating meeting minutes received the highest rating (3.96), followed by stakeholder communication (3.80). The task of supporting project portfolio decisions received the lowest rating (3.67), though the score remained significantly above the scale's midpoint of 3.0.

V. DISCUSSION

A. Preliminary Findings and Their Implications

Although the field test is constrained by its relatively small sample size and therefore limited generalizability to the broader population of PM professionals, the analysis revealed noteworthy findings. First, the results suggest that the PM professionals in our sample are pragmatic adopters, using GenAI primarily for low-risk, high-return administrative tasks. This cautious adoption pattern, especially for complex strategic tasks, aligns with literature on initial AI integration [2][3].

Second, consistent with [14], our data hints at heterogeneous GenAI use across experience levels. The tendency for some novice users to apply GenAI to advanced tasks suggests a potential "competency gap" risk, where the tool is used without sufficient domain expertise for verification of AI outputs. Our data offers some support for this, as novice 'Explorers' attempted advanced tasks like portfolio decision-making.

Finally, the perceived productivity gains appear inversely related to task complexity. The highest productivity gains are perceived in standardized administrative tasks like meeting minutes, where the need for human judgment is low. However, even the most complex task was rated positively, suggesting that the surveyed PM professionals find value in using GenAI as a supportive tool across different levels of PM work. Future work should include a more detailed analysis of differences based on experience level, both in PM and in the use of GenAI.

B. Lessons Learned in Instrument Development

From deploying the survey in a first field test, several lessons for instrument development emerged. First, operationalizing GenAI use via proxy tasks involves a trade-off. Using representative tasks allows us to ask for concrete productivity gains, but it constrains generalizability. Respondents may use GenAI heavily in other tasks (as four users noted in the open-ended responses) but not in the specific task presented. A future iteration could include a broader task battery per AAA category, or let respondents select the task(s) they use most within each category. Moreover, as the ratings were restricted to actual users of each task, subgroup sizes may be uneven as in our data. This affects reliability estimates, comparability, and statistical power for subgroup analyses.

Second, the translation of perceived productivity into measurable indicators requires further clarification. Using multiple dimensions improved conceptual coverage but also necessitates clearer boundaries to avoid overlap. Interviews could test how respondents interpret and distinguish these dimensions, particularly for indirect effects. In addition, rework (e.g., due to GenAI hallucinations or compliance constraints) should be explicitly accounted for, as it may offset initially perceived productivity gains. Future refinements could include counter-items capturing

Copyright (c) IARIA, 2026. ISBN: 978-1-68558-400-9

verification effort, rework frequency, and confidence in output correctness, especially for high-stakes tasks.

Third, while self-classified GenAI proficiency is pragmatic, a future iteration could consider adding a small set of objective or behavioral indicators (e.g., frequency of use, number of tools used, prompt complexity) to triangulate proficiency. Together, these lessons suggest directions for the next instrument iteration. Moreover, validation steps should be planned that go beyond the initial internal consistency checks performed in this study.

VI. Conclusion and Future Work

A central challenge in AI productivity measurement is moving beyond generic self-assessments to a concrete, multidimensional measurement. To address this for PM, we developed a survey instrument for assessing perceived productivity gains, grounded in research on GenAI usage and AI productivity measurement. Crucially, our instrument operationalizes productivity by distinguishing between direct gains (time, cost, quality) and indirect gains (e.g., improved decision quality), allowing for an assessment that captures both immediate task efficiency and long-term capability building. To ground these measurements in practice, they were applied to three proxy tasks representing distinct modes of human-AI interaction: automation, assistance, and augmentation. The instrument was empirically applied in a field test with 97 German professionals with practical GenAI experience in PM. The field test indicates that the measurement approach is practically usable and provides a basis for further refinement and validation.

References

[1] L. Hughes *et al.*, "Impact of artificial intelligence on project management (PM): Multi-expert perspectives on advancing knowledge and driving innovation toward PM2030," *J. Innov. Knowl.*, vol. 10, no. 5, p. 100772, Sep. 2025, doi: 10.1016/j.jik.2025.100772.

[2] PMI, "Pushing the Limits: Transforming Project Management with Generative AI Innovation | PMI," Project Management Institute, 2024. [Online]. Available from https://www.pmi.org/learning/thought-leadership/transforming-project-management-with-generative-ai, retrieved: 06, 2026

[3] S. Colditz, "Der KI-Co-Pilot: Warum das PPM ein hybrides Entscheidungsmodell braucht," Deutsch Gesellschaft für Projektmanagement GPM, 2025. [Online]. Available from https://www.gpm-ipma.de/ueber-die-gpm/blog/der-ki-co-pilot-warum-das-ppm-ein-hybrides-entscheidungsmodell-braucht, retrieved: 06, 2026

[4] PMI, "First Movers' Advantage: The Immediate Benefits of Adopting Generative AI For Project Management: Enhancing Project Outcomes Through Intelligent Technologies," Project Management Institute, 2024. [Online]. Available from https://www.pmi.org/learning/thought-leadership/benefits-of-ai-for-project-management, retrieved: 06, 2026

[5] R. Nyqvist, A. Peltokorpi, and O. Seppänen, "Integration of generative artificial intelligence across construction management," *IOP Conf. Ser. Earth Environ. Sci.*, vol. 1389, no. 1, p. 012011, 2024, doi: 10.1088/1755-1315/1389/1/012011.

[6] R. Müller, G. Locatelli, V. Holzmann, M. Nilsson, and T. Sagay, "Artificial Intelligence and Project Management: Empirical Overview, State of the Art, and Guidelines for Future Research," *Proj. Manag. J.*, vol. 55, no. 1, pp. 9–15, 2024, doi: 10.1177/87569728231225198.

[7] PMI, "Shaping the Future of Project Management With AI: Charting Your AI Upskilling Journey with AI," Project Management Institute, 2023. [Online]. Available from https://www.pmi.org/learning/thought-leadership/shaping-the-future-of-project-management-with-ai, retrieved: 06, 2026

[8] S. Kunz and C. Hess, "Artificial Intelligence – Myth or Measurable? A systematic framework to determine AI-induced productivity gains," presented at the DIGITAL 2025, Advances on Societal Digital Transformation, Jul. 2025, pp. 1–6. [Online]. Available from https://www.thinkmind.org/library/DIGITAL/DIGITAL_2025/digital_2025_1_10_18001.html, retrieved: 06, 2026

[9] F. Calvino, J. Reijerink, and L. Samek, "The effects of generative AI on productivity, innovation and entrepreneurship," Paris, OECD Artificial Intelligence Papers 39, Jun. 2025. doi: 10.1787/b21df222-en.

[10] X. Hao, E. Demir, and D. Eyers, "Exploring collaborative decision-making: A quasi-experimental study of human and Generative AI interaction," *Technol. Soc.*, vol. 78, p. 102662, 2024, doi: 10.1016/j.techsoc.2024.102662.

[11] PMI Sweden, "AI and Project Management: A Global Chapter-Led Survey," 2024. [Online]. Available from https://www.pmi.org/-/media/pmi/documents/public/pdf/artificial-intelligence/community-led-ai-and-project-management-report.pdf, retrieved: 06, 2026

[12] M. Alavi, D. Leidner, and R. Mousavi, "Knowledge Management Perspective of Generative Artificial Intelligence (GenAI)," *J. Assoc. Inf. Syst.*, vol. 25, no. 1, 2024, doi: 10.17705/1jais.00859.

[13] M. Glickman and Y. Zhang, "AI and Generative AI for Research Discovery and Summarization," *Harv. Data Sci. Rev.*, vol. 6, no. 2, Mar. 2024, doi: 10.1162/99608f92.7f9220ff.

[14] H. F. Wild and A. Flore, "Nutzenerwartung und Anwendungsfälle von KI im Projektmanagement.," *Proj.-Manag. Aktuell*, vol. 35, no. 5, pp. 42–49, 2024, doi: 10.24053/PM-2024-0089.

[15] B. Hettrich, N. Krings, and A. Kock, "Bridging the Expertise Gap: The Role of Generative AI in Supporting Project Planning Tasks for Novices and Professionals," *Creat. Innov. Manag.*, vol. 34, no. 4, pp. 789–805, 2025, doi: 10.1111/caim.70002.

[16] C. Hess and S. Kunz, "The Project Management Assistant that Never Sleeps - Using Generative Artificial Intelligence to Support Project Management Activities," *Int. J. Adv. Softw.*, vol. 17, no. 1 & 2, pp. 44–58, 2024.

[17] D. A. Dillman, J. D. Smyth, and L. M. Christian, *Internet, phone, mail, and mixed-mode surveys: the tailored design method*, 4th ed. Hoboken: Wiley, 2014.

[18] P. Schober, C. Boer, and L. A. Schwarte, "Correlation Coefficients: Appropriate Use and Interpretation," *Anesth. Analg.*, vol. 126, no. 5, p. 1763, 2018, doi: 10.1213/ANE.0000000000002864.

Enhancing Service Personalization with AI and Robot Agility

Edward C. S. Ku
Department of Travel Management
National Kaohsiung University of Hospitality and Tourism
No.1, Songhe Rd., Xiaogang Dist., Kaohsiung City, 81271, Taiwan (R.O.C),
e-mail: edwardku@mail.nkuht.edu.tw

***Abstract*— Restaurants face increasing challenges in maintaining service flexibility and quality while responding to rapidly changing customer demands in technology-enabled environments. This study investigates how an artificial intelligence climate, customer–artificial intelligence-assisted exchange, and business process innovativeness enhance operational adjustment agility and customer service competencies through robot agility. The study contributes to research on artificial intelligence, digital transformation, and intelligent service systems by explaining how service robots improve organizational adaptability and service performance. Drawing on real options theory, parasocial interaction theory, and hybrid intelligence theory, service robots are viewed as adaptive digital assets that facilitate responsiveness and customer engagement. New empirical evidence was obtained from a two-phase managerial survey of restaurants in Shanghai and Taipei with more than three years of robot deployment. Data from 357 valid responses were analyzed using partial least squares structural equation modeling. The results indicate that an artificial intelligence climate, customer–artificial intelligence-assisted exchange, and business process innovativeness significantly enhance robot agility, which subsequently improves customer service competencies. The findings demonstrate that integrating service robots with artificial intelligence-enabled processes can increase organizational flexibility and maintain service quality in the restaurant industry.**

***Keywords*— service robots; hybrid intelligence; robot agility; customer service competencies.**

I. INTRODUCTION

AI-powered service robots are increasingly adopted in restaurants as digital assets that streamline operations and enhance customer experience [16][30]. These systems automate routine tasks such as ordering and food delivery, improving service speed, precision, consistency, and operational agility. Building on advances in restaurant automation and customer experience management, this study explains how organizational and technological capabilities jointly strengthen responsiveness, customer trust, and loyalty in AI-intensive environments. It addresses an important gap by examining how agility and perceived service quality shape customer experience within hybrid human–AI service settings.

Service robots improve efficiency, perceived value, and customer satisfaction through adaptive interactions [14][26]. Mechanisms such as CAIX and business process innovativeness accelerate service innovation, strengthen coordination in food and beverage delivery, and support personalized customer experiences, particularly in tourism-oriented restaurant contexts where digital differentiation is essential. Prior studies suggest that CAIX promotes emotionally engaging interactions, while perceived anthropomorphism strengthens trust, satisfaction, and loyalty by influencing emotional responses. These mechanisms collectively explain how AI-driven innovation contributes to long-term relational value and digital service excellence.

II. LITERATURE REVIEW

This section reviews the literature on AI-enabled service robots in restaurant environments, focusing on hybrid intelligence, AI climate, CAIX, business process innovativeness, and anthropomorphism. Prior studies suggest that these organizational and technological capabilities enhance robot agility, adaptive service delivery, and customer service competencies in hospitality settings. Drawing on hybrid intelligence and service innovation perspectives, this section develops the conceptual framework and hypotheses linking AI climate, CAIX, business process innovativeness, robot agility, anthropomorphism, and customer service competencies.

A. Hybrid Intelligence and Human–AI Synergy in Service Systems

Hybrid intelligence refers to the strategic integration of human and artificial intelligence, combining their complementary strengths to address complex service challenges more effectively than either could independently [5]. This paradigm utilizes AI for scalability, speed, consistency, and task automation, while preserving human capabilities in contextual reasoning, ethical judgment, and adaptive decision-making [29]. In service environments, hybrid intelligence supports value co-production through coordinated interactions between employees and AI-enabled service robots.

AI enhances operational efficiency, responsiveness, and personalization at scale, whereas employees contribute emotional intelligence and context-sensitive service to maintain relational quality and adaptability in customer experiences [30]. This synergy is particularly important in dynamic, high-contact service settings. Empirical evidence further suggests that AI-conducive organizational climates strengthen agility, stimulate innovation, and improve resource orchestration, thereby supporting the effective implementation of hybrid intelligence in service innovation.

B. AI Climate and Strategic Integration of Service Robotics

In hospitality, AI-powered service robots support operational efficiency and sustainability by improving energy use and reducing reliance on manual labor [28]. A strong AI climate further enhances innovation readiness by embedding AI into strategic initiatives such as circular economy practices and resource-efficient sharing platforms. At the organizational level, human–AI dynamics shape workforce outcomes.

Khaliq et al. [15] found that perceptions of AI or robotic autonomy can increase turnover intention, highlighting the need to manage human–robot coexistence. Building on this, [30] demonstrated that when trust-enabled robots are paired with data-driven personalization, they strengthen engagement and satisfaction and provide a pathway for strategic differentiation in AI-supported hospitality.

C. Robot Agility

Robot agility in restaurants refers to the ability of AI-enabled service robots to adapt to shifting operational conditions, guest preferences, and real-time service demands [26]. Robots should offer innovative options for risk-tolerant customers and intuitive interfaces for those facing time pressure or high information load. Service providers must identify pain points such as confusion, limited time, or low digital literacy and deliver support aligned with customer capability to enhance service quality. A strong AI climate supported by digital infrastructure, organizational policies, and cultural readiness further strengthens this adaptability.

In AI-enabled hospitality settings, continuous data capture and real-time analytics guide robot behavior and enable context-sensitive adjustments [17]. This environment encourages investment in software updates and algorithmic enhancements to sustain adaptability. An AI-conducive climate promotes innovation through experimentation with automation and robotics. AI-integrated training helps employees coordinate effectively with robots, while machine learning allows robots to learn from interactions and improve performance. These conditions form the foundation for Hypothesis 1:

H1: Artificial Intelligence Climate climate of restaurants positively influences robot agility in restaurant environments.

D. Customer–AI-Assisted Exchange (CAIX)

CAIX refers to the interactive processes between human-like service robots and customers in which communication, feedback, and adaptation drive AI-enabled service innovation [22]. Pei et al. [23] noted that these interactions are reshaping service roles as robots increasingly perform frontline tasks once handled by human staff. Jadhav et al. [13] emphasized that CAIX functions as a real-time learning mechanism, allowing robots to refine responses through repeated encounters and tailor service to individual preferences.

These systems process inquiries instantly, adjust dynamically, and deliver personalized recommendations, such as recalling dietary needs or preferred service styles, which enhances responsiveness and satisfaction. As interaction volume and diversity increase, the depth of available data expands, strengthening robots' ability to make context-sensitive decisions. CAIX therefore acts as a key enabler of robotic agility in hospitality by linking user engagement with ongoing AI refinement. Based on this perspective, we propose the following hypothesis.

H2: CAIX positively influences robot agility in restaurant environments.

E. Business Process Innovativeness

Business process innovativeness in service robot deployment refers to the strategic integration of AI-powered robots into organizational workflows to strengthen coordination between automated systems and human operations [9]. By embedding robots into routine processes, firms can automate tasks, improve workflow adaptability, and enhance operational efficiency through intelligent systems and process redesign.

This capability improves customer experience, resource utilization, and digital innovation across hospitality contexts [9][24]. It also supports personalized service delivery and flexible process adjustments in response to changing customer demands. As robots increasingly perform functions such as order handling and guest interaction, machine learning capabilities enable faster and more adaptive service responses, thereby strengthening robotic agility and strategic responsiveness:

H3: Business process innovativeness and service robot innovation positively influence robot agility in restaurant environments.

F. Customer Service Competencies

Customer service competencies in AI-enabled restaurants refer to an organization's ability to meet guest needs effectively through digitally mediated interactions [30]. Prior studies emphasize that incorporating human-like traits, such as conversational ability and emotional responsiveness, into service robots enhances customer engagement and interaction quality [32]. In addition, GenAI-powered robots increasingly perform complex tasks while delivering highly personalized customer experiences.

AI-driven environments also expand service capacity by integrating automation with adaptive technologies [3]. Robots equipped with real-time feedback capabilities can instantly adjust service parameters, improving responsiveness and coordination with employees. In high-contact settings such as fine dining, robots interpret contextual cues to refine service delivery and strengthen guest experience, while business process innovativeness ensures procedural consistency and enhances customer trust and satisfaction.

H4: Business process innovativeness positively influences customer service competencies in AI-enabled restaurant environments.

G. Anthropomorphism

Anthropomorphism refers to the extent to which a digital or robotic agent is perceived as displaying human-like features or behaviors [20]. In AI-enabled service environments, elements such as facial expressions, conversational speech, and interactive dialogue enable robots to communicate more naturally and engage customers more effectively [7]. These human-like cues evoke emotional responses that positively shape perceptions of robotic agility, adaptability, and responsiveness.

Anthropomorphic design also functions as a heuristic for competence and intelligence. When robots exhibit gestures, facial expressions, or modulated voice tones, users are more likely to attribute cognitive capability, adaptive decision-making, and context-sensitive service to the system. Such perceptions increase tolerance toward minor service failures while strengthening trust, satisfaction, and perceived value. Consequently, anthropomorphic design enhances both the perceived intelligence and relational effectiveness of AI-powered service agents.

H5: Perceived anthropomorphism positively moderates the relationship between robot agility and customer service competencies.

Overall, the proposed research model explains how AI climate, CAIX, and business process innovativeness enhance robot agility in AI-enabled restaurant environments. Robot agility subsequently improves customer service competencies, while perceived anthropomorphism strengthens this relationship, as illustrated in Figure 1. The model emphasizes the joint role of organizational capabilities and AI-enabled interactions in supporting adaptive service delivery and customer experience.

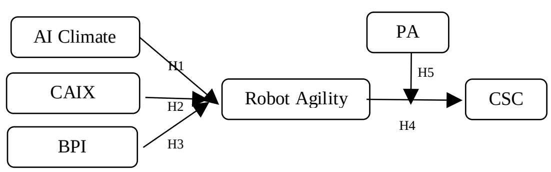

Figure 1. Research model.

Note: CAIX = Customer–AI-Assisted Exchange; BPI = Business Process Innovativeness; PA = Perceived Anthropomorphism; and CSC = Customer Service Competencies.

III. RESEARCH DESIGN

This section describes the research methodology used to examine the proposed relationships among AI climate, CAIX, business process innovativeness, robot agility, perceived anthropomorphism, and customer service competencies. It outlines the questionnaire development process, sampling procedure, and data collection approach employed to ensure the reliability and validity of the empirical investigation.

A. *Questionnaire Development*

Following an extensive literature review, the survey items were customized for the research context, yielding 29 preliminary questions. Wording was refined based on initial assessments. Participants answered using a 7-point Likert scale (where 1 means strongly disagree and 7 means strongly agree), as detailed in Table I, with validated measures chosen specifically for this study. To achieve more precise translations, the instrument was adjusted, translated into both Simplified and Traditional Chinese, and examined by experts. Two hospitality management professors and two information management scholars reviewed the questionnaire for clarity, accuracy, and alignment with the research objectives. A pretest with 30 restaurant employees further assessed item clarity and relevance, confirming the suitability of the last version.

TABLE I. ITEMS IN SURVEY

AI Climate [1]	
AIC1	We use service robots to enhance customer experience.
AIC2	We use service robots to provide expert services.
AIC3	We personalize services through robots.
AIC4	We use service robots to support service strategies
Customer–AI Assisted Exchange [16]	
CAIX1	Customers behave respectfully toward service robots.
CAIX2	Customers value robot-delivered services.
CAIX3	Complaints about robot services are infrequent.
CAIX4	Customers are satisfied with robot services.
CAIX5	Customers enjoy interacting with service robots
Business Process Innovativeness [9, 24]	
BPI1	Robot services align with business processes.
BPI2	We redesign workflows for new opportunities.
BPI3	We adapt robot services to changing needs.
Robot Agility [17]	
RA1	We share robot service to reduce costs.
RA2	Service robots enable new applications.
RA3	We share robots to support innovation.
RA4	Service robots improve maintenance speed
Perceived Anthropomorphism [20]	
PA1	Service robots use human-like voices.
PA2	Service robots provide human-like interactions.
PA3	Service robots have human-like appearances.
PA4	Service robots display human-like behavior.
Customer Service Competencies [30].	
CSC1	Customers are satisfied with service robots.
CSC2	Customers view robot adoption positively.
CSC3	Customers enjoy interacting with service robots.
CSC4	Robot interactions are meaningful and enjoyable.
CSC5	Service robots provide valuable service options.
CSC6	Customers look forward to future robot interactions

B. *Sampling Process*

Restaurants in Shanghai and Taipei were chosen based on their continuous use of service robots for more than three years. Data was collected in January 2025 from 100 robot-assisted restaurants selected through stratified sampling, with 50 from each city. Surveys were distributed by email, and IP verification prevented duplicate entries. This process yielded 357 valid responses, resulting in a 59.5 percent response rate.

IV. ANALYSIS

The data showed that nearly 61.5% of the respondents were female, and over 50% were aged 31 or older. Additionally, 33% of the participants had more than 6 years of work experience.

A. *Measurement Model*

Partial least squares structural equation modeling (PLS-SEM) was employed to evaluate the proposed research model [12]. Measurement reliability was assessed using Cronbach's alpha, rho_A, composite reliability (CR), and outer loadings. All Cronbach's alpha values ranged from 0.837 to 0.939, while CR values ranged from 0.892 to 0.952, exceeding the recommended threshold of 0.70. Convergent validity was further examined using average variance extracted (AVE), with all constructs reporting values between 0.673 and 0.821, surpassing the recommended 0.50 level [21]. These results confirm satisfactory internal consistency reliability and convergent validity of the measurement model.

Discriminant validity was subsequently assessed using cross-loadings and the Fornell–Larcker criterion [12]. All cross-loading values exceeded the recommended threshold of 0.70, providing additional support for convergent validity [4]. Furthermore, the square root of AVE for each construct was greater than its correlations with other constructs, confirming adequate discriminant validity. Overall, the findings demonstrate that the measurement model satisfies the required standards of reliability, convergent validity, and discriminant validity.

B. *Model Testing*

The coefficient of determination (R^2) reflects the variance in endogenous variables explained by the model. In this study, R^2 values were 55.2 percent for robot agility, and 44.3 percent for customer service competencies. Effect size (f^2) values were all below the 0.33 benchmark [11], consistent with an explanatory model. Multicollinearity was assessed using the variance inflation factor (VIF), which should remain below 3.3 [10]. VIF values ranged from 0.056 to 0.602, indicating no collinearity issues.

The results indicate that AI climate ($t = 6.47$, $p < 0.01$), customer–AI assisted exchange (CAIX) ($t = 5.92$, $p < 0.01$), and business process innovativeness ($t = 5.60$, $p < 0.01$) significantly enhance robot agility. These findings are consistent with prior studies highlighting the importance of AI-enabled organizational environments and innovation capabilities in strengthening adaptive service performance [15][16]. Furthermore, robot agility significantly improves customer service competencies ($t = 14.314$, $p < 0.001$), supporting earlier research emphasizing the role of agile service robots in enhancing customer interaction quality and service effectiveness [3][8]. In addition, perceived anthropomorphism significantly moderates the relationship between robot agility and customer service competencies ($t = 2.66$, $p < 0.05$).

V. DISCUSSION

The study's results provide valuable insights into the contributions of examining customer service competencies driven by service robots' innovation.

A. *Theoretical Contribution*

This study extends service robot and AI-enabled hospitality research by moving beyond robot characteristics and customer reactions to examine robot agility as a key organizational capability. Drawing on Real Options Theory, Parasocial Interaction Theory, and Hybrid Intelligence Theory, it develops an integrated framework explaining how AI climate, customer–AI-assisted exchange (CAIX), and business process innovativeness jointly enhance customer service competencies through robot agility. Unlike prior studies that investigate these factors separately, the findings demonstrate their complementary effects in creating responsive and adaptive service systems. The results further show that robot agility mediates the relationship between organizational capabilities and service outcomes, while anthropomorphism strengthens the effectiveness of agile robotic services. By conceptualizing service robots as adaptive socio-technical resources rather than merely automated tools, this study provides a clearer explanation of how AI-enabled hospitality firms achieve superior service performance in dynamic environments.

B. *Implications for Practices*

This study offers strategic guidance for managers integrating AI-enabled service robots into digital transformation initiatives. Establishing a strong AI climate equips robots to respond to fluctuating customer needs, improving operational agility and enabling flexible deployment across varying service conditions. Such digital readiness ensures that robotic systems can adjust to operational changes and support consistent service delivery.

AI should also advance personalized customer experiences through real-time data use. CAIX plays a central role in agile service delivery, as real-time exchanges and adaptive feedback allow robots to tailor responses and enhance efficiency. Managers can strengthen these outcomes by prioritizing customer-centric design, improving human–robot coordination, and monitoring performance continuously. Business process innovativeness further integrates AI systems, robots, and employees. Intelligent automation supports better decision-making, workflow visibility, and interdepartmental coordination, while investments in task optimization and cross-functional data improve scalability and reliability.

Anthropomorphic design contributes to the relational quality of robot interactions. Features such as expressive cues or modulated voice patterns foster trust and emotional engagement, although managers should balance expressiveness with functional efficiency based on customer expectations. Overall, service robots should be viewed as dynamic actors embedded within a socio-technical system. Strengthening AI climate, CAIX, business process innovativeness, agility, and design features collectively enhance service value, operational flexibility, and long-term competitive advantage.

VI. CONCLUSIONS

This study shows how AI-enabled innovation strengthens operational agility and customer service in robot-assisted restaurants through five key enablers: AI climate, CAIX, business process innovativeness, robot agility, and perceived anthropomorphism. A strong AI climate embeds digital capabilities into workflows, improving flexibility and personalization, while CAIX enhances real-time customer–robot interaction for greater accuracy and relevance. Business process innovativeness further supports seamless coordination among AI systems, robots, and staff, ensuring consistent service execution. Robot agility enables rapid task adjustment and adaptive workflow management, and anthropomorphic cues improve emotional engagement and trust. To achieve these benefits, managers should invest in real-time monitoring, AI-integrated training, and data-driven personalization so that technological infrastructure and organizational routines work together to deliver adaptive, high-quality service and reinforce competitive advantage.

This study improves understanding of how service robots support customer service competencies in AI-enabled restaurant settings. However, certain limitations remain. First, while robot agility and AI capabilities are emphasized, implementation challenges are not fully addressed. Future research should explore barriers such as costs, integration complexity, and employee resistance to better reflect the socio-technical aspects of digital transformation.

REFERENCES

[1] Akter, S., Wamba, S. F., Mariani, M., & Hani, U. "How to build an AI climate-driven service analytics capability for innovation and performance in industrial markets?" Industrial Marketing Management, 97, 2021, pp. 258-273. https://doi.org/10.1016/j.indmarman.2021.07.014

[2] Ali, F., Sesliokuyucu, O. S., Khan, K. A., Alotaibi, S., & Wu, C. Z. "The impact of robotic gastronomic experiences on customer value, delight and loyalty in service-robot restaurants." International Journal of Human-Computer Interaction, 41(13), 2025, pp. 8053-8065. https://doi.org/10.1080/10447318.2024.2404503

[3] Balaji, M. S., Jiang, Y. Y., & Zhang, X. Y. "Robots in service: how robot capabilities and personalities drive customer value co-creation and satisfaction." International Journal of Contemporary Hospitality Management, 37(3), 2025, pp. 1016-1035. https://doi.org/10.1108/ijchm-01-2024-0112.

[4] Becker, J. M., Cheah, J. H., Gholamzade, R., Ringle, C. M., & Sarstedt, M. "PLS-SEM's most wanted guidance." International Journal of Contemporary Hospitality Management, 35(1), 2023, pp. 321-346. https://doi.org/10.1108/ijchm-04-2022-0474

[5] Blaurock, M., Büttgen, M., & Schepers, J. "Designing collaborative intelligence systems for employee-AI service co-production." Journal of Service Research. 2024. https://doi.org/10.1177/10946705241238751

[6] Cavusoglu, M., Collins, G. R., Demicco, F., & Cobanoglu, C. "Robot acceptance and service quality in food delivery: An expanded TAM-based study." International Journal of Human-Computer Interaction. 2025, https://doi.org/10.1080/10447318.2024.2445112

[7] Chi, O. H., Chi, C. G., & Gursoy, D. "Seeing personhood in machines: Conceptualizing anthropomorphism of social robots." Journal of Service Research, 28(1), 2025, pp. 78-92. https://doi.org/10.1177/10946705241297196

[8] Cui, J. S., & Zhong, J. A. "The effect of robot anthropomorphism on revisit intentions after service failure: a moderated serial mediation model." Asia Pacific Journal of Marketing and Logistics, 35(11), 2023, pp. 2621-2644. https://doi.org/10.1108/apjml-10-2022-0862

[9] Eckert, T., Hüsig, S., & Doblinger, C. "The influence of market entry strategies on the success of digital service innovations in the mobility sector." Journal of Service Research. 2024. https://doi.org/10.1177/10946705241245463

[10] Goodhue, D. L., Lewis, W., & Thompson, R. "A multicollinearity and measurement error statistical blind spot: correcting for excessive false positives in regression and PLS." MIS Quarterly, 41(3), 2017, pp. 667-684.

[11] Guenther, P., Guenther, M., Ringle, C. M., Zaefarian, G., & Cartwright, S. "Improving PLS-SEM use for business marketing research." Industrial Marketing Management, 111, 2023, pp.127-142. https://doi.org/10.1016/j.indmarman.2023.03.010

[12] Hair, J. F., Babin, B. J., Ringle, C. M., Sarstedt, M., & Becker, J. M. "Covariance-based structural equation modeling (CB-SEM): a SmartPLS 4 software tutorial." Journal of Marketing Analytics. 2025, https://doi.org/10.1057/s41270-025-00414-6

[13] Jadhav, D., Fatima, J. K., & Quazi, A. "Antecedents of team alignment for team performance: length of relationship as a moderator." Journal of Business & Industrial Marketing, 38(12), 2023, pp. 2731-2744. https://doi.org/10.1108/jbim-10-2022-0464

[14] Jain, H., Dhupper, R., Shrivastava, A., Kumar, D., & Kumari, M. "AI-enabled strategies for climate change adaptation: protecting communities, infrastructure, and businesses from the impacts of climate change." Computational Urban Science, 3(1), 2023. https://doi.org/10.1007/s43762-023-00100-2

[15] Khaliq, A., Waqas, A., Nisar, Q. A., Haider, S., & Asghar, Z. "Application of AI and robotics in hospitality sector: A resource gain and resource loss perspective.' Technology in Society, 68, 2022. https://doi.org/10.1016/j.techsoc.2021.101807

[16] Ku, E. C. S. "Digital marketing innovation and industrial marketing: evidence from restaurants' service robots." Asia Pacific Journal of Marketing and Logistics, 36(11), 2024, pp. 3099-3117. https://doi.org/10.1108/apjml-02-2024-0185

[17] Kumar, S., Poudyal, A., Choudhury, S., & Jha, A. "Analysing customer preferences for AI-enabled service robot's physical appearance in the tourism and hospitality industry: insights and exploration." Tourism Planning & Development. 2025. https://doi.org/10.1080/21568316.2025.2461008

[18] Larivière, B., Verleye, K., De Keyser, A., Koerten, K., & Schmidt, A. L. "The service robot customer experience (SR-CX): A matter of ai intelligences and customer service goals." Journal of Service Research, 28(1), 2025. 35-56. https://doi.org/10.1177/10946705241296051

[19] Lucas, D. S. "The effect of regime change on entrepreneurship: A real options approach with evidence from US gubernatorial elections." Journal of Business Venturing, 39(4), 2024. https://doi.org/10.1016/j.jbusvent.2024.106394

[20] Luque-Raya, I., Del Barrio-García, S., & Garcia, O. C. "The social dimension of consumer-AI interaction: alienation, anthropomorphism, and identity in AI-Mediated marketing." Behaviour & Information Technology. 2025. https://doi.org/10.1080/0144929x.2025.2545313

[21] Lyu, T., Guo, Y. L., & Chen, H. “Understanding the privacy protection disengagement behaviour of contactless digital service users: the roles of privacy fatigue and privacy literacy.” Behaviour & Information Technology, 43(10), 2024, pp. 2007-2023. https://doi.org/10.1080/0144929x.2023.2237603

[22] Ma, S., Ge, L., Jia, H., & Wang, Y. G. “Understanding and mitigating the robot disadvantage in luxury services: The role of desire for superiority.” Information Systems Research. 2025. https://doi.org/10.1287/isre.2023.0229

[23] Pei, J. L., Wang, H. L., Peng, Q. P., & Liu, S. S. “Saving face: Leveraging artificial intelligence-based negative feedback to enhance employee job performance.” Human Resource Management. 2024. https://doi.org/10.1002/hrm.22226

[24] Pizam, A., Ozturk, A. B., Hacikara, A., Zhang, T. T., Balderas-Cejudo, A., Buhalis, D.,…State, O. “The role of perceived risk and information security on customers' acceptance of service robots in the hotel industry.: International Journal of Hospitality Management, 117, 2024. https://doi.org/10.1016/j.ijhm.2023.103641

[25] Qin, M., Li, S. Q., Zhu, W., & Qiu, S. S. (2025). “Trust in service robot: the role of appearance anthropomorphism.” Current Issues in Tourism, 28(1), 2025, pp. 36-54. https://doi.org/10.1080/13683500.2023.2295923

[26] Reyes-Uquillas, D., & Hsiao, T. S. “Safe and intuitive manual guidance of a robot manipulator using adaptive admittance control towards robot agility.” Robotics and Computer-Integrated Manufacturing, 70, 2021. https://doi.org/10.1016/j.rcim.2021.102127

[27] Santiago, J., Borges-Tiago, M. T., & Tiago, F. “Embracing RAISA in restaurants: Exploring customer attitudes toward robot adoption.” Technological Forecasting and Social Change, 199, 2024. https://doi.org/10.1016/j.techfore.2023.123047

[28] Simón, C., Revilla, E., & Sáenz, M. J. “Integrating AI in organizations for value creation through Human-AI teaming: A dynamic-capabilities approach.” Journal of Business Research, 182, (2024). https://doi.org/10.1016/j.jbusres.2024.114783

[29] Soprano, M., Roitero, K., La Barbera, D., Ceolin, D., Spina, D., Demartini, G., & Mizzaro, S. “Cognitive biases in fact-checking and their countermeasures: A review.” Information Processing & Management, 61(3), 2024. https://doi.org/10.1016/j.ipm.2024.103672

[30] Wu, S. H.. “Deep affection: service robots increase brand love of restaurants.” British Food Journal, 127(1), 2025, pp.129-147. https://doi.org/10.1108/bfj-06-2024-0631

[31] Wu, S. H., & Ku, E. C. S. ‘Aligning restaurants and artificial intelligence computing of food delivery service with product development.” Journal of Hospitality and Tourism Technology, 15(3), 2024, pp.379-396. https://doi.org/10.1108/jhtt-10-2023-0322

[32] Yang, J. Q., Amrollahi, A., & Marrone, M. “Harnessing the potential of artificial intelligence: Affordances, constraints, and strategic implications for professional services.” Journal of Strategic Information Systems, 33(4), 2024. https://doi.org/10.1016/j.jsis.2024.101864 .

 ISBN: 978-1-68558-400-9

Enhancing Public Sector Agility Through Digital Transformation: A Technology-Organization-Environment Perspective

Aysha Youssef
Doctor of Business Administration (DBA), Teesside University
Middlesbrough, United Kingdom
email: s3108647@live.tees.ac.uk

Abstract— Digital transformation has become a strategic priority for governments seeking to strengthen service delivery, institutional responsiveness and sustainable innovation. However, public sector organizations often struggle to translate digital investments into sustained agility because technology is shaped by leadership, culture, governance and environmental conditions. This paper examines how digital transformation may influence organizational agility in UAE public sector organizations through the Technology-Organization-Environment (TOE) perspective. As the wider doctoral study is ongoing, this camera-ready version reports interim, anonymized evidence only and does not present final thesis findings, a final coding structure or a finalized framework. Drawing on 17 anonymized semi-structured interviews and a structured literature review, the preliminary analysis suggests that agility is not produced by technology alone, but by governed alignment between technological capabilities, organizational readiness and environmental legitimacy. The paper offers a provisional TOE-based analytical model to guide further analysis of public sector agility through digital technologies.

Keywords—public sector agility; TOE perspective; digital transformation; government organizations.

I. INTRODUCTION

Public sector organizations operate in environments shaped by rapid technological change, complex policy demands, rising citizen expectations and pressure to provide more responsive services. Digital government is no longer limited to digitizing existing administrative processes; it increasingly involves the redesign of service delivery, data-informed decision-making and new forms of interaction between government and society [4][5]. In the UAE, digital government is also connected to national ambitions for service excellence, innovation, sustainable development and future readiness [7][11].

Despite this progress, digital transformation does not automatically generate organizational agility. Technology is enacted through structures, routines, interpretations and human agency rather than operating as an independent solution [2]. Digital transformation research similarly emphasizes that technology adoption must be supported by organizational change, leadership commitment, strategic alignment and learning to create meaningful outcomes [4][5].

The research issue addressed in this paper is the gap between digital transformation as technological investment and agility as an institutional capability. Public organizations may introduce dashboards, automation, robotic process automation, artificial intelligence and digital portals while retaining hierarchical approval paths, fragmented data, risk-averse routines and limited cross-functional authority. In such cases, digital tools may accelerate information flow but not necessarily improve decision speed, coordination, adaptation or innovation.

Organizational agility refers to the capability to sense, interpret and respond to changing conditions while maintaining continuity and accountability [3]. In public sector settings, agility is especially challenging because governments must balance responsiveness with legality, transparency, equity, security and institutional stability [6]. This means that agility in government cannot be understood as unrestricted speed. Rather, it must involve disciplined adaptation: rapid responsiveness within clear governance boundaries.

This paper adopts the Technology-Organization-Environment (TOE) framework because it explains transformation as the result of technological, organizational and environmental conditions rather than technology alone [1]. The technological dimension captures digital systems, data, automation, artificial intelligence and interoperability. The organizational dimension captures leadership, culture, skills, decision rights and governance routines. The environmental dimension captures regulation, cybersecurity expectations, national strategies, citizen needs and institutional pressures. Recent extensions of TOE also show that the framework can be adapted to digital transformation contexts by incorporating contextual capabilities and orientations beyond the original adoption focus [8].

The provisional contribution of this paper is threefold. First, it repositions TOE from an adoption-oriented framework toward an agility-oriented lens for public sector digital transformation. Second, it uses selected anonymized interview evidence illustratively to show how technological, organizational and environmental conditions interact in government contexts. Third, it proposes a provisional TOE-based model of governed alignment that will be refined after completion of the wider doctoral analysis.

The remainder of this paper is structured as follows. In Section 2, the literature on public sector digital transformation, organizational agility and TOE is reviewed. In Section 3, the methodology, sample, interview protocol and analysis procedures are described. In Section 4, preliminary findings are presented through TOE-coded themes and interview evidence. In Section 5, the theoretical and practical implications are discussed. In Section 6, the conclusion and future work are presented.

 ISBN: 978-1-68558-400-9

II. LITERATURE BACKGROUND

This section reviews the literature informing the study. It focuses on digital transformation in the public sector, organizational agility as a dynamic capability and the rationale for using TOE to examine agility through digital technologies.

A. Digital Transformation in the Public Sector

Digital transformation in the public sector involves changes in organizational processes, service models, decision practices and relationships with citizens and stakeholders [4]. It is broader than the implementation of information systems because it affects the way public institutions organize, collaborate and create public value. Vial argues that digital transformation creates change through the combined effects of digital technologies, organizational responses and strategic consequences [5].

Government organizations face distinctive transformation conditions. They operate within formal accountability systems, policy mandates, audit requirements and data protection obligations. Consequently, public sector digital transformation must reconcile innovation with compliance, transparency and continuity. This makes the public sector a valuable context for examining whether digital tools create agility or only automate existing routines.

B. Organizational Agility and Adaptive Governance

Organizational agility is closely related to dynamic capability theory, particularly the capacity to sense change, seize opportunities and reconfigure resources [3]. In government, this capability supports adaptive governance, crisis response, service redesign, policy adjustment and future readiness. However, public sector agility differs from private sector agility because the goal is not only competitive advantage, but also public value, legitimacy and service continuity.

Adaptive governance literature stresses the need to combine responsiveness with stability and accountability [6]. Participants in this study echoed this view by describing agility as disciplined adaptation rather than unrestricted speed. This framing is important because public organizations must act quickly while protecting sensitive data, observing legislation and maintaining citizen trust.

C. Technology-Organization-Environment Perspective

The TOE framework provides a holistic basis for examining digital transformation because it considers technological, organizational and environmental factors together [1]. This is particularly relevant to public sector agility, where digital systems interact with leadership behavior, culture, skills, regulation, cybersecurity, national priorities and citizen expectations.

The extended TOE literature also supports the adaptation of TOE to newer digital transformation contexts. Nguyen, Le and Vu extend the framework by adding contextual orientations and capabilities that improve explanatory power in digital transformation settings [8]. This paper builds on that logic by using TOE not only to explain adoption, but also to analyze post-adoption agility outcomes. In other words, the question is not simply whether a technology is adopted, but whether it strengthens sensing, decision-making, coordination and adaptive service delivery.

III. RESEARCH METHODOLOGY

This section summarizes the methodological approach, empirical sources, participant selection, interview protocol and analytical steps used to generate the preliminary evidence reported in this paper.

A. Research Design and Data Sources

The study adopts an exploratory qualitative design as part of an ongoing Doctor of Business Administration research project on agility in UAE public sector organizations. The qualitative approach is appropriate because the study seeks to understand how senior professionals interpret digital transformation, organizational readiness and agility in practice rather than test predefined statistical relationships.

Two sources inform this paper. First, a structured literature review was conducted across public administration, organizational change, agility and digital transformation literature to identify definitions, models, gaps and theoretical foundations. Second, semi-structured interviews were conducted with senior professionals in UAE public sector contexts. The paper reports preliminary findings from 17 anonymized participant accounts. Because the research is qualitative, response distribution is reported as thematic recurrence rather than statistical generalization.

TABLE I. RESEARCH DESIGN AND SAMPLE.

Research approach	**Exploratory qualitative study informed by systematic literature review**
Empirical source	17 anonymized semi-structured interviews
Participant profile	Senior or specialist roles in strategy, technology, finance, foresight, legislation, institutional excellence, project management and service delivery
Sampling logic	Purposive selection of participants with direct exposure to digital transformation or agility-related work
Analysis method	Thematic analysis mapped inductively and deductively to TOE dimensions
Ethical protection	Participants and entities anonymized; identifiable demographic and organizational details removed

B. Interview Protocol and Analytical Procedure

The interview protocol was designed to elicit evidence on agility practices rather than abstract opinion only. Participants were asked to define agility, describe digital and artificial intelligence initiatives, explain changes in work and decision-making, identify implementation barriers, discuss leadership and culture and reflect on future public sector agility. Examples included questions such as: How do you

 ISBN: 978-1-68558-400-9

define organizational agility in a government context? Which digital tools or systems have supported faster response or decision-making? What challenges have emerged in technology adoption? How do leadership, culture, regulation and national direction influence agility?

Interviews were recorded and transcribed for analysis. Thematic analysis followed iterative familiarization, open coding, theme grouping and comparison against the TOE framework [9]. Trustworthiness was strengthened through anonymization, transcript review where available, iterative coding, comparison with literature and attention to credibility, dependability and confirmability principles [10][12].

TABLE II. PRELIMINARY THEMATIC DISTRIBUTION ACROSS 17 INTERVIEWS.

Theme	**Indicative preliminary recurrence and illustrative evidence**
AI, analytics and automation	Indicative recurrence: participants cited dashboards, RPA, risk engines, portals, Copilot and internal AI platforms.
Leadership and change readiness	Indicative recurrence: sponsorship, empowerment, cross-functional teams, learning culture and resistance management.
Governance, cybersecurity and data	Indicative recurrence: data quality, confidentiality, controlled AI, legal boundaries and system integration.
Public value and service impact	Indicative recurrence: reduced visits, elderly service users, faster services and societal value of AI.

IV. PRELIMIBARY EMPIRICAL INSIGHTS

At this stage, the preliminary interview evidence suggests that public sector agility emerges through alignment among technological capability, organizational readiness and environmental legitimacy. These insights should not be read as final thesis findings; rather, they are interim patterns used to strengthen the conference paper and demonstrate the direction of analysis. The evidence does not support a technology-deterministic view. Instead, digital tools appear to enhance agility when embedded in governance, leadership, culture and data practices.

A. Technology: From Digitization to Decision Support

Participants described digital systems as enablers of faster access to information, improved reporting and better coordination. Examples included artificial intelligence dashboards for forecasting, robotic process automation for repetitive financial tasks, predictive risk engines, legislative observatories, open data systems, digital portals and business process mapping. One participant explained that information that previously required lengthy searching can now be retrieved in seconds, while another described automation as a way to reduce manual errors and speed operational processes.

However, the technology dimension also revealed limits. Participants emphasized that artificial intelligence should support, not replace, decision-makers. They also identified data quality, cybersecurity, infrastructure readiness, interoperability and model reliability as key conditions. This suggests that technology contributes to agility when it improves sensing and decision support while remaining trustworthy and integrated into operational workflows.

B. Organization: Leadership, Culture and Cross-Functional Work

Organizational conditions were the strongest explanatory factor in the data. Leadership support, empowerment, agile structures, cross-functional teams and a learning culture repeatedly appeared as enablers. Participant 01 stated that agility requires employees to move between strategic and technical levels when project needs demand it. Participant 04 described reduced hierarchy and the use of cross-functional teams as important for flexibility. Participant 12 summarized public sector agility as disciplined adaptation, meaning rapid response while maintaining governance.

Resistance to change was also recurrent. Several participants described employees or departments that initially resisted new systems because benefits were unclear, time was limited or fear of replacement existed. Participants indicated that proof of concepts, awareness sessions, visible leadership sponsorship and evidence of tangible benefit helped reduce resistance. Therefore, agility is not simply a technical capability; it is a social and managerial capability that converts technology into coordinated action.

C. Environment: National Direction, Regulation and Public Trust

Environmental factors acted simultaneously as drivers and constraints. UAE national strategies, digital government agendas, artificial intelligence leadership roles and service-excellence expectations created pressure to transform. Participants also referred to COVID-19, heavy rain events and service expectations as examples showing the importance of readiness and continuity. These pressures pushed organizations toward mobile services, remote work, automation and data-driven decision-making.

At the same time, regulation, cybersecurity and data confidentiality defined the boundaries of acceptable agility. Participants described controlled internal clouds, on-premises artificial intelligence environments, approval committees and concern about confidential data leakage. Participant 10 emphasized that digital services and artificial intelligence must create tangible societal value rather than being adopted for technology's sake. This illustrates why public sector agility must be bounded by legitimacy, ethics and trust.

D. Cross-Dimensional Preliminary Insight: Governed Alignment

The emerging cross-dimensional insight is that agility appears strongest when the three TOE dimensions are aligned. Digital tools become agility mechanisms when leaders define priorities, data are reliable, employees are capable, governance is clear and external legitimacy is protected. This alignment may explain why similar technologies can produce different outcomes across entities.

TABLE III. TOE-CODED PRELIMINARY INSIGHTS WITH ANONYMIZED PRIMARY EVIDENCE.

TOE dimension	**Preliminary insight and illustrative anonymized evidence**
Technology	AI, RPA and analytics may improve sensing and speed when reliable data and workflow integration are in place.
Organization	Leadership, culture and decision rights appear to shape responsiveness; participants cited sponsorship, cross-functional teams and resistance management.
Environment	National strategies, cybersecurity, legislation and public value expectations act as both drivers and boundaries for disciplined adaptation.

V. DISCUSSIONS

The preliminary analysis provides illustrative evidence for the research issue while deliberately avoiding presentation of the final thesis results. At this stage, the interview data suggest that public sector agility is not a single outcome of digital adoption. Rather, it appears to be a governed capability that develops when digital tools are connected to decision rights, culture, security, national priorities and public value.

A. Theoretical Contribution

The provisional theoretical contribution is the extension of TOE from adoption explanation to agility explanation. Traditional TOE studies tend to ask which factors influence technology adoption [1]. This paper asks how adopted technologies may contribute to organizational agility. The shift is important because an organization may adopt artificial intelligence, dashboards or automation without becoming more agile if decision rights, data governance and cultural readiness remain weak.

The provisional model in Figure 1 conceptualizes governed alignment as the mechanism through which technology, organization and environment may support public sector agility. The model highlights that artificial intelligence, automation and analytics do not directly create agility unless they are linked with data quality, leadership, skills, governance, legal compliance and public-value objectives.

TOE CONTEXTS
Technology + Organization + Environment
v
GOVERNED ALIGNMENT
disciplined adaptation | secure data | decision rights
v
AGILITY MECHANISMS
speed | resilience | public value
Provisional analytical model for ongoing doctoral research

Figure 1. Provisional TOE-based governed alignment model for public sector agility.

The model provisionally reframes agility in government as disciplined adaptation. This concept emerged strongly in the interview evidence and clarifies the difference between public and private sector agility. Public sector agility means the ability to move quickly when needed while still protecting rights, data, equity, national security and accountability.

B. Practical Contribution

For practitioners, the preliminary insights suggest that public organizations should assess digital transformation through agility outcomes rather than technology deployment alone. Relevant indicators include decision-cycle speed, service continuity, public-value creation, quality of data use, employee readiness and governance reliability.

The preliminary insights also suggest four practical actions. First, artificial intelligence and automation should be linked to clear business cases and public-value outcomes. Second, entities should create safe experimentation mechanisms such as sandboxes, internal platforms and controlled proof-of-concept environments. Third, digital skills, change management and employee participation should be strengthened. Fourth, governance should be embedded into artificial intelligence workflows so that agility is accelerated without increasing unacceptable risk.

VI. CONCLUSION AND FUTURE WORK

This paper examined, as part of an ongoing doctoral study, how digital transformation may influence organizational agility in UAE public sector organizations through the TOE perspective. It addressed the problem that digital transformation often remains framed as technological implementation rather than as an organizational capability requiring alignment across technology, organization and environment.

The preliminary evidence from 17 anonymized interviews suggests that agility is not generated by technology alone. Artificial intelligence, automation, dashboards and digital platforms may support agility when connected to leadership, culture, decision rights, data quality, cybersecurity and public-value objectives. The paper therefore identifies governed alignment as a provisional analytical mechanism linking technology capability, organizational readiness and environmental legitimacy.

 ISBN: 978-1-68558-400-9

The study remains ongoing and has limitations. The insights presented here are interim, anonymized and illustrative; they are not the final DBA thesis findings. To protect the integrity of the ongoing thesis, the paper does not disclose the full coding hierarchy, final participant-level analysis, final framework or final conclusions. These will be reported in the thesis and, where appropriate, in subsequent publications. Future work will complete the coding analysis, refine the model and examine how digital technologies shape agility across different government functions.

REFERENCES

[1] L. G. Tornatzky and M. Fleischer, The Processes of Technological Innovation. Lexington, MA: Lexington Books, 1990.

[2] W. J. Orlikowski, “The duality of technology: Rethinking the concept of technology in organizations,” Organization Science, vol. 3, no. 3, pp. 398-427, 1992.

[3] D. J. Teece, G. Pisano, and A. Shuen, “Dynamic capabilities and strategic management,” Strategic Management Journal, vol. 18, no. 7, pp. 509-533, 1997.

[4] I. Mergel, N. Edelmann, and N. Haug, “Defining digital transformation: Results from expert interviews,” Government Information Quarterly, vol. 36, no. 4, Article 101385, 2019.

[5] J. Vial, “Understanding digital transformation: A review and a research agenda,” The Journal of Strategic Information Systems, vol. 28, no. 2, pp. 118-144, 2019.

[6] M. Janssen and H. van der Voort, “Adaptive governance: Towards a stable, accountable and responsive government,” Government Information Quarterly, vol. 33, no. 1, pp. 1-5, 2016.

[7] OECD, Digital Government Review of the United Arab Emirates: Towards a More Innovative and Data-Driven Public Sector. Paris: OECD Publishing, 2021.

[8] T. H. Nguyen, X. C. Le, and T. H. L. Vu, “An extended technology-organization-environment (TOE) framework for online retailing utilization in digital transformation: Empirical evidence from Vietnam,” Journal of Open Innovation: Technology, Market, and Complexity, vol. 8, no. 4, Article 200, 2022.

[9] V. Braun and V. Clarke, “Using thematic analysis in psychology,” Qualitative Research in Psychology, vol. 3, no. 2, pp. 77-101, 2006.

[10] Y. S. Lincoln and E. G. Guba, Naturalistic Inquiry. Beverly Hills, CA: Sage, 1985.

[11] UAE Government, “The UAE Digital Government Strategy 2025,” The Official Portal of the UAE Government, 2025. Available: https://u.ae/ [accessed: 2026.05.16].

[12] J. W. Creswell and C. N. Poth, Qualitative Inquiry and Research Design: Choosing Among Five Approaches, 4th ed. Thousand Oaks, CA: Sage, 2018.

The Role of Digital Government in the Digital Transformation of Society: Case Study of North Macedonia

[1] Sandra Cvikić
Regional Center Vukovar
Ivo Pilar Institute of Social Sciences
Vukovar, Croatia
sandra.cvikic@pilar.hr

[2] Nadica Nikoloska
Institute for Intellectual Property - Skopje
Public Scientific Institution, Skopje, North Macedonia
n.nikoloska@is.edu.mk

[3] Goce Naumovski
Cyril and Methodius University in Skopje, Faculty of Law
"Iustinianus Primus", Republic of North Macedonia
g.naumovski@pf.ukim.edu.mk

[4] Filip Vučen
Institute for Intellectual Property - Skopje
Public Scientific Institution
Skopje, North Macedonia
filip.vucen@gmail.com

***Abstract* - The Republic of North Macedonia serves as an excellent case study of the digital transformation of a society aspiring to EU membership, where digital government plays a prominent role. Using the Dynamic Policy Maturity Benchmark Model in qualitative research focused on the continuous evolution of digital policies and strategic documents for digital transformation in the country, this paper demonstrates how the EU integration process has significantly influenced digital governance in recent years. Government trade-offs in the development of solutions and negotiation of digital policies that have advanced the digitalisation of Macedonia represent a stable strategy designed to address the challenges posed by the rapid digital transformation of society. The proposed model for assessing the current situation and the potential for improved digitalisation of Macedonian society has proved to be a useful tool for future policy development and decision-making.**

Keywords-North Macedonia; digital government; digital transformation; EU integration; Dynamic Policy Maturity Benchmark Model.

I. Introduction

The European Commission (hereafter EC) initiated accession negotiations with Macedonia in October 2009, requiring subsequent governments to comply with recommended conditionalities to progress, pending a positive assessment that led to renewed accession negotiations in June 2019. Successive efforts on key issues such as corruption, judicial reform, intelligence and security services reform, and public administration reform resulted in renewed support from the ministers of European affairs in March 2020, as the European Union (hereafter EU) confirmed its political agreement to open accession negotiations with Albania and Macedonia, which was formally endorsed by the European Council on 26 March 2020. Setting aside the complex negotiated political conditionalities and criteria of the EU accession process, this paper focuses on the Republic of North Macedonia's (hereafter Macedonia) digital Europeanisation process, particularly the role of digital government in the digital transformation of society. Section 2 outlines the methodological and policy framework used in the analysis and interpretation of the research data. Section 3 provides the context of digitalisation and the background to digital transformation. Section 4 discusses the results of the qualitative analysis – the assessment of Macedonian digital transformation. Section 5 interprets the research findings of the assessment, and Section 6 presents the concluding remarks and insights into the role of digital government in the contemporary digital transformation of Macedonian society, with recommendations for future work.

II. Scope of Research

The digital transformation of Macedonian society, as a complex social phenomenon, has significantly affected the population, politics, and economy, particularly through the introduction of Artificial Intelligence (hereafter AI) into almost every aspect of daily life. The aim of this paper is to present evidence on the digital transformation of Macedonian society and the role of digital government in this process. The question that guides this inquiry is whether strategic development based on the digitalisation process in Macedonia can advance policy evolution through benchmark assessment. This is based on a review of publications – including policy reports, policy briefs, research papers, and studies – produced by domestic and international organisations, experts, and scholars. The methodological framework developed to analyse selected texts includes a case study of Macedonian digital transformation since 2021, together with a process evaluation examining how national policies on digitalization were implemented. This approach determines whether the policy is being implemented as planned and identifies any issues during the implementation process. Therefore, the research method employs the following analytical steps: a) data gathering and selection, b) qualitative sociological data analysis, and c) process evaluation of digital policy implementation - Dynamic Policy Maturity Benchmark Model (hereafter DPMBM) [1].

 ISBN: 978-1-68558-400-9

Considering the size of the territory and the population (1,836,713 inhabitants according to the latest census from 2021), the DPMBM is proposed as suitable for assessing implemented digital policies designed to ensure policy evolution that responds to technological changes and societal needs in Macedonia [2]. This model enables the evaluation of the influence and impact of newly introduced digital technologies and services within the digital ecosystem in Macedonian society. The benchmark ensures seamless alignment between strategy and the impacts of digitalisation, and each time new technologies are introduced an assessment is triggered to determine whether they support citizens' people-centric values. Therefore, this paper chooses to use the ITU's Benchmark for Fifth Generation Digital Collaborative Regulation (G5 Benchmarks), which often serves as a national benchmark for inclusive digital transformation across all sectors of a country's economy and society [3]. The benchmark design indicators are grouped around four pillars: 1) national collaborative governance, 2) policy design principles, 3) digital development toolbox, and 4) digital economic policy agenda. As a structured approach to policy evolution, this paper uses the above-mentioned model to introduce an analytical tool – dynamic benchmarks transformed into Key Performance Indicators (hereafter KPIs) that measure and adjust policy effectiveness. The KPIs for digitalisation include the following: data protection, privacy, confidentiality of communications, digital governance, connectivity, data literacy, digital skill development, level of digital innovation, inclusive investment, digital equity, data ethics, cybersecurity, open data, data accessibility/democratisation, level of data culture, digital-by-default rules, use of digital twins, use of digital simulation, availability in continuous operations, and interoperability. These indicators were used to assess Macedonian digital transformation and were further evaluated using the following assessment criteria: aligned (+), not aligned (-), partially aligned (PA), and no information (NI), referring to the alignment of digitalisation policies with national strategic documents and international documents. The assessment criteria aligned (+) and partially aligned (PA), in the context of digital systems, imply that planned policies in Macedonia exist and are in place within developing digital infrastructure, and are thus fully or partially aligned with proposed measures, priorities and goals of strategic development documents [4]. However, as proposed by Klievink et al., digital policy optimization, assessed according to the evaluation criteria fulfilled (+), not fulfilled (-), partially fulfilled (PF), and no information (NI), implies that policies are implemented in integrated digital infrastructure and thus either fully or partially contribute to the digital transformation of Macedonian society [5]. In this respect, digital transformation is assessed as a continuous process of policy change and development towards policy maturity. As indicated by tables in the following sections, the outlined KPIs are assessed according to the above-mentioned criteria, based on the data presented in strategic documents and assessments (domestic and international).

Continuous assessment of this kind enables policies to be updated so they remain relevant and effective in responding to the challenges of rapid digitalisation in Macedonia. As all strategic planning in Macedonia must ensure sustainable, inclusive, socially acceptable, and ecologically sound development policies in accordance with EU accession political criteria, legislation, and professional standards, the concept of a DPMBM enables the assessment of digital transformation in a people-centred country [6]. Therefore, the proposed benchmark model evaluates the potential and impact of the introduced digital technologies and services in the Macedonian digital ecosystem as a sustainable country. It is designed to define Macedonia's distinctive digital policy maturity levels. However, this paper focuses on the digital transformation assessment of Macedonian society according to policy criteria KIPs indicators proposed by the DPMBM at two levels:

1. institutions and mechanisms
2. digitalisation.

The remaining levels (society, people-centric, sustainability, and economy) were not considered due to limited data resources for in-depth analysis.

Assessment findings were interpreted within the framework of Thomas Sowell's understanding of policy solutions [7]. His theoretical approach to core decision-making in policy creation, development, and implementation rests on the assumption that there are no solutions, only trade-offs, in the economy underlying any human undertaking. In this paper, Sowell's understanding is considered indispensable for the study of Macedonia's digital transformation, where policy choices are not cost-free, and decisions taken in that context resolve one problem while simultaneously sacrificing something else. Decision-makers must therefore carefully balance policy outcomes, as there is no perfect solution and the inherent costs of implementation may eliminate one problem but often create or worsen another. Within the framework of digital Europeanisation, Macedonian governments must therefore accept responsibility for the costs of the trade-offs they choose to make on their path of negotiated EU integration. The paper focuses on the set of trade-offs Macedonian governments have made in their European negotiation process that shape the digitalisation of the entire society.

III. Context and Background

A. Context

As of 2019, the digitalisation of Macedonian public services, together with the introduction of digital wallets in accordance with Regulation (EU) No 910/2014 of the European Parliament and of the Council on electronic identification and trust services for electronic transactions in the online market and its repeal, has been a challenging task for both government authorities and citizens [8]. The process has been characterised by insufficient knowledge, trust, and digital skills among citizens on one hand, and on the other, by inert state institutions and a lack of legal provisions and institutional trust. Thus, as an integral part of Macedonia's European integration process, digitalisation and digital wallets, which fall under Chapter 10 of the acquis Communautaire: Information society and media, have posed

 ISBN: 978-1-68558-400-9

challenges for the authorities in adopting the necessary changes through the long-term Information and Communication Technology Strategy. As will be evident from the following sections, Macedonia has made substantial progress since the 2021 EU Progress Report, which stated that the country was moderately prepared in this area of European integration.

B. Background

The Law on Electronic Documents, Electronic Identification, and Confidential Services was adopted in 2019 (fully aligned with the eIDAS Regulation of 2014), creating the preconditions for the Macedonian government to develop a stable legal framework for digitalisation in subsequent years. This has also enabled the creation of digital identity as a driver of the digitalisation of public services and the development of a functional common digital market, both of which have contributed to the market for confidential services and increased trust in electronic services in Macedonia [9]. The established legal framework has regulated the work of public administration bodies involved in data and document transfers and the implementation of electronic services. Subsequently, the 2019 Law on Electronic Documents, Electronic Identification, and Confidential Services has enabled digital judicial practices, electronic employment [10], and governance, made operable by the newly formed National Interoperability Platform and managed by the Ministry of Digital Transformation. In this way, the automated and interoperable system – the ‘Macedonian Information Highway’ – helps the government bridge the gap between policy and system implementation [11]. However, according to expert assessment, creating and cultivating digital culture in Macedonia was and remains a challenging task [12]. The extensive legislation, comprising 131 laws to be prepared by the established Inter-Ministerial Working Group since 2021, required substantial time to align with the rest of the legislative framework and the new digitalisation plans. Whatever moderate progress was made by the Ministry of Information Society and Administration, the following three years will demonstrate how improved interoperability of the digitalisation system, including public sector institutions and critical infrastructure, is achieved according to strategic planning. Although the proposed National Information and Communication Technology Strategy 2021–2025 (including the Action Plan) was never adopted by the Assembly and was subsequently amended and revised by the newly formed Ministry of Digitalisation in 2025, reforms to digitalise society were implemented despite challenging political circumstances with Macedonian neighbours during EU negotiations until June 2022. Recommendations made by experts in 2022 have been incorporated into development policies, creating a more favourable environment for the progressive digital transformation of Macedonian society [13]. To overcome the evident tension between progressive legislation and deep-seated institutional inertia, Macedonian digital governance had to confront bureaucratic resistance, fragmented inter-agency data sharing, and limited public digital skills and literacy [14]. Embedding legislation into digital governance has made administrative reform possible and opened it up to strategic collaboration, participation, transparency, and access to information [15]. As indicated by Gil-Garcia et al., information technology has become a catalyst for open government, producing diverse effects in a digitalised society. In the Macedonian case, it has fostered a sense of transparency in the provision of high-quality government information and data, as well as enhanced participation, including citizens in government decision-making processes while collaborating with individuals and/or organisations to provide services. The developed adaptability of digitalised governance in Macedonia has transformed government capacity, creating more resilient and agile institutions [16]. However, the extent to which the Macedonian government has become agile in its software development, acquisition, project management and evaluation is presented in the following sections on the digital transformation of Macedonian society, according to the Dynamic Policy Maturity Benchmark Model.

IV. Assessment of Macedonian Digital Transformation

Two strategic documents have shaped the development of digitalisation in Macedonian society since 2023. The first, adopted in 2022, is the Development Framework of the Republic of North Macedonia 2030 (hereafter DFRNM 2030) [17]. The second, adopted in 2024, is the 2024–2044 National Development Strategy (hereafter NDS) [18]. Subsections A and B provide insight into the assessed potential for digital transformation in Macedonia, while subsection C concerns the assessment of strategically implemented digitalisation policies based on selected international policy progress assessments. Subsection D then provides insight into digital transformation according to domestic policy progress assessments.

A. The Development Framework for the Republic of North Macedonia 2030 (DFRNM 2030)

DFRNM 2030, together with other priorities, identifies the digitalisation of Macedonian society as a complex and interconnected priority essential for real progress by 2030. It regards digitalisation as a strategic development goal necessary for achieving other development priorities, so it can be considered both a separate priority and a means to achieve other priorities. Reappropriation of the concept of digitalisation in real-world contexts concerns four interconnected aspects of technologically advanced societies: skills, governance, business, and public infrastructure, which form the basis for development strategies. Macedonian society is no exception. The DFRNM 2030 assessment of the current state of digitalisation notes that although the e-service system was introduced to Macedonian public services, it was still not fully functional in 2022. However, according to State Statistics from 2020, 91% of companies with more than ten employees use computer technology and broadband Internet. More than half of these companies have developed their own websites, and 6.2% use electronic sales platforms (e-commerce). The assessment indicates that 83% of Macedonian households have Internet access at home, and most use broadband Internet. Still, the priorities of digitalisation in sustainable development require Macedonian authorities to invest in high-quality education, gender equality, industry, innovation and infrastructure, as well as in peace, justice and strong institutions.

Therefore, DFRNM 2030 proposes four key interventions for digitalising society to support the development needs of a country aspiring to EU membership. The first area involves accelerating the country's digital transformation by investing in digital infrastructure and technology, providing digital literacy training for the wider population, including the labour force, and introducing electronic identity measures and public-private partnership models. The second area focuses on promoting digital economic development in Macedonia through investment in businesses that facilitate access to information about available digital technologies and the business opportunities they offer, promoting domestic and international trade in goods and services via formal electronic services, implementing the Digital Agenda for the Western Balkans, and improving the digital skills of the Macedonian workforce. The third area concerns developing the Information Technology sector (hereafter IT) in Macedonia by investing in Information and Communication Technology (hereafter ICT) expert education, adopting a strategic approach to prevent the brain drain of IT personnel, developing the IT sector as a prosperous service activity, and making targeted investments in the IT sector. Finally, the fourth area concerns the full digitalisation of Macedonian public services through investment in the e-services system, while expanding public institutions' capacity to deliver e-services and enhancing digital literacy among citizens and businesses to use e-services. Therefore, to harness the benefits of new digital technologies in this era of rapid digital transformation, Macedonian strategic development must consider the dynamics of policy maturity.

TABLE I. DFRNM 2030 POTENTIAL FOR DIGITALIZATION 2022 -2024

No.	**Dynamic Policy Maturity Benchmark Assessment**	
	Assessment criteria	***Aligned (+) Not aligned (-) Partially aligned (PA) No information (NI)***
1	**data protection**	NI
2	**privacy rights**	NI
3	**digital governance**	+
4	**confidentiality of communication**	NI
5	**digital inclusion/connectivity**	+
6	**data literacy**	+
7	**digital skill development**	+
8	**level of digital innovation**	PA
9	**inclusive investment**	+
10	**digital equity for all**	PA
11	**data ethics**	PA
12	**cybersecurity**	NI
13	**availability in continuous operations**	+
14	**open data**	NI
15	**data democratization**	NI
16	**level of data culture**	+
17	**digital by default rules**	+
18	**use of digital twins/simulation**	NI
19	**interoperability**	NI

Macedonian strategic development in the digital transformation of society, according to the DPMBA assessment (Table I), demonstrates fragmented capacity and potential for digitalisation during the 2022–2024 period. The potential for digitalisation is developed mainly in digital governance, digital inclusion and connectivity, data literacy, digital skills, inclusive investment, availability in continuous operations, level of data culture, digital by default rules, and to a lesser extent in digital innovation, digital equity for all, and data ethics. The potential for digitalisation of Macedonian society therefore requires additional government incentives and diligent policy making and development.

B. The 2024-2044 National Development Strategy (NDS)

The NDS was adopted by the Assembly of the Republic of North Macedonia in 2024, in line with the country's aspirations to join the EU by 2030 and the sustainable development goals set by the 2030 Agenda [19]. North Macedonia has therefore based its development on three fundamental national development goals across six key strategic areas of interest. The development of human capital and the digitalisation of society are central elements and fundamental prerequisites for the country to achieve successful progress in each strategic area. Among the national capabilities and capacities expected to improve as key factors in developing the strategic areas of the NDS – microeconomic stability, strong institutions, territorial organisation and functional decentralisation, inclusion and social cohesion, climate change and environmental issues, and political consensus – the one that stands out as a horizontal measure is data management and use, digital technology, and the digitalisation of Macedonian society. Therefore, as in DFRNM 2030, digitalisation becomes a priority, aiming to achieve a higher degree of Macedonian modernisation through information technologies in both the private and public sectors. Public authorities and the business sector must provide quality, timely, and disaggregated data essential for creating national evidence-based development policies. The NDS recommends that the state invest in a digital strategy focusing on four key pillars:

1. developing a skilled workforce to attract top professionals from the digital sector,
2. building secure and environmentally friendly digital infrastructure,
3. facilitating the digital transformation of businesses, and
4. providing public digital services focused on users.

This will require substantial changes in administrative and management structures so that implemented reforms enable interoperability and systems for data collection, management, protection, and use in decision-making processes. To initiate overall social transformation, the NDS proposes several strategic areas of intervention. It emphasises that data and technologies are crucial drivers of progress, with the potential to reduce regional development disparities and enhance cohesion in diverse Macedonian geographical areas and communities. This will enable fact-based, targeted government interventions and investments to support underdeveloped Macedonian areas, in line with overall regional development. The NDS states that digital platforms play a key role in this respect, as they provide a wide range of information and facilitate the sharing of best practices, enabling stronger collaboration and mobilisation of local and regional resources for the benefit of all. It is expected that established digital platforms will equip all local communities with the basic tools and knowledge to access, on equal terms, the world of digital technologies and infrastructure in Macedonia. Furthermore, the development of smart cities integrated with Internet of Things (hereafter IoT) devices is expected to coordinate and improve the functioning of all infrastructure and services provided by local authorities, making them more efficient over time. Digital finance and blockchain technology are expected to change the fiscal decentralisation process, as digitalisation in every sense has become essential to the overall transformation of Macedonian society. To accelerate its digital transformation, Macedonia is committed to promoting the digital economy and education, enhancing citizens' digital literacy, and fully digitalising public services, which enable free access to information provided by public authorities. In this strategic area, the horizontal measure of society's digitalisation focuses on the rule of law and good governance. It is assumed that digitalisation will significantly improve efficiency, effectiveness, transparency, accountability, and responsibility, which are necessary for the healthy development of Macedonian democracy. The use of digital tools in this respect is indispensable. E-government initiatives also represent modern interaction between citizens and state authorities, making public services more accessible through streamlined processes. Public sector data is more accessible to all citizens through open data platforms, and clear, presentable information enhances government accountability, enabling authorities to ensure legal certainty and impartial justice. In this respect, online legal services and digital courts enable and support access to justice and improve the efficiency and confidentiality of legal proceedings. As a key component of Macedonia's digital transformation, the country must invest further in developing digital skills and strengthening cybersecurity. To address these challenges, the government has adopted a defense and security strategy (the Strategy for Defence and Long-Term Capability Plan for 2019–2028), aiming to create policies that systematise the workforce and ICT jobs, as well as broadband and cybersecurity measures. The digitalisation process, in this context, reaffirms the NDS's commitment to leveraging new technologies for real-time data analytics within an established digital framework, enabling Macedonian society to confront a range of security challenges. In this way, the digitalisation of Macedonian society becomes a roadmap, creating an environment that fosters the resilience and sustainability of all government and public systems of operation.

TABLE II. NDS POTENTIAL FOR DIGITALIZATION 2024 -2025

No.	Dynamic Policy Maturity Benchmark Assessment	
	Assessment criteria	*Aligned (+) Not aligned (-) Partially aligned (PA) No information (NI)*
1	**data protection**	+
2	**privacy rights**	PA
3	**digital governance**	+
4	**confidentiality of communication**	+
5	**digital inclusion/connectivity**	+
6	**data literacy**	+
7	**digital skill development**	+
8	**level of digital innovation**	+
9	**inclusive investment**	+
10	**digital equity for all**	+
11	**data ethics**	NI
12	**cybersecurity**	+
13	**availability in continuous operations**	+
14	**open data**	NI
15	**data democratization**	PA
16	**level of data culture**	+
17	**digital by default rules**	+
18	**use of digital twins/simulation**	NI
19	**interoperability**	+

According to the NDS, the digital transformation of Macedonian society, based on the DPMBM assessment, shows significant improvement (Table II). The capacity and potential for digitalisation in the 2024–2025 period have increased in all areas except those for which no data are available (data ethics, open data, and the use of digital twins/simulation). This positive trend in policy maturity is also observed in relevant international assessments of digital transformation progress in Macedonia and the Western Balkans generally.

C. Digital Transformation in Accordance with International Policy Progress Assessments

The Macedonian integration process since 2004, when the country applied for EU membership, represents a political decision endorsed by all governments as the only viable development option. However, the journey has been marked by numerous economic challenges and socio-political tensions resulting from project-oriented and technocratic development principles within the EU. The transitional nature of the ongoing reforms implemented so far demonstrates the extent to which the dynamics and impact of these policies have been shaped by political discontinuities and transformative liberal democratic ideas. This section assesses the actual progress of digitalisation in Macedonian society according to selected reports and policy papers from international organisations present in the country and the Western Balkans.

a) Assessment according to Digital Compass for 2030

The extent to which the character of Macedonian digital transformation has evolved to accommodate transitional reforms and policies aligned with the EU integration process is first assessed in comparison with the Digital Compass for 2030 – Digital Decade policy programme, which provides clear guidelines, targets, and objectives [20]. As a benchmark for assessing Macedonia's digital progress, this policy pathway outlines four crucial areas of transformation: skills, secure and sustainable digital infrastructure, digital transformation of business, and digitalisation of public services. According to key indicators for assessing digital transformation - Digital Economy and Society Index (hereafter DESI), it is possible to monitor Macedonia's progress compared to the EU. Of all DESI areas, the greatest progress has been made in digital public services, followed by human capital development, connectivity, and integration of digital technology. However, digital development still lags behind the EU average in all DESI areas. Despite numerous policy initiatives to enhance overall digital skills, human capital development in this area remains significantly below the EU average. This is due to lower proficiency in internet user skills, although some positive developments have been observed in the percentage of ICT graduates and female ICT specialists. Despite significant progress in digital services and service delivery, Macedonia still faces challenges in establishing domestic portals dedicated to digital services, which would provide easier access to essential public services for citizens and businesses. Efforts have been made to improve and strengthen the legal framework for digital services and the provision of e-signatures. Furthermore, Macedonia has enhanced the availability of and access to open data. Although progress has been made, advancement and policy focus remain insufficient to reach the EU average (16.8) in digital public services, even though, for example, Macedonia is slightly above (75.4%) the EU average of 64.5% for pre-filled online forms. In conclusion, both this and the pandemic lockdowns, which accelerated the integration of online sales, still lacked progress and left considerable room for improvement in the overall technological development and digital transformation of Macedonian society.

TABLE III. ASSESSMENT OF THE ALIGNMENT WITH DIGITAL COMPASS FOR 2030

No.	**Dynamic Policy Maturity Benchmark Assessment**	
	Assessment criteria	***Aligned (+) Not aligned (-) Partially aligned (PA) No information (NI)***
1	**data protection**	PA
2	**privacy rights**	PA
3	**digital governance**	PA
4	**confidentiality of communication**	PA
5	**digital inclusion/connectivity**	PA
6	**data literacy**	PA
7	**digital skill development**	PA
8	**level of digital innovation**	PA
9	**inclusive investment**	PA
10	**digital equity for all**	PA
11	**data ethics**	PA
12	**cybersecurity**	PA
13	**availability in continuous operations**	PA
14	**open data**	PA
15	**data democratization**	PA
16	**level of data culture**	PA
17	**digital by default rules**	PA
18	**use of digital twins/simulation**	NI
19	**interoperability**	PA

As the DPMBM assessment in Table III shows, advancing the digitalisation of public services in Macedonia largely depends on strong leadership and political will capable of effectively managing digital development and societal transformation. Empowering appointed state officials and authorities at all government levels to proactively manage state policies within budgetary provisions and financial constraints requires modernising the administrative system through digital optimisation of processes and services, as well as re-engineering in line with the 'once-only principle'. A more inclusive and open public administration system therefore necessitates promoting its services with strong user involvement to ensure the successful implementation of digital plans. This snapshot of Macedonian digitalisation efforts, driven by the government's EU integration process, indicates slow but

 ISBN: 978-1-68558-400-9

steady progress in the digital development and transformation of society (Table III).

b) Assessment according to Strategic Digital Policy Development

The European Commission's 2025 Report commends Macedonia's "smooth functioning of the internet market for electronic communications, electronic commerce and audio-visual services" (EC, 2025: 73) [21]. It states that "the rules protect consumers and support the universal availability of modern services" (EC, 2025: 73). Compared to 2023, significant progress has been made in Macedonian digital transformation. Long-awaited laws regulating digital development in Macedonia were adopted in June 2025. In connection with this legal and normative framework, Macedonia has adopted a long-term Information and Communication Technology Strategy and an Action Plan. However, the report notes that there is still room for improvement, particularly in enhancing the capacities of the Ministry of Digital Transformation "to more effectively coordinate the country's digital transformation policies" (EC, 2025: 73). Further improvements are required in electronic communications to align the country with the EU Gigabit Infrastructure Act, based on the still valid 2019–2029 broadband operational plan. The Broadband Competence Office states that, according to 2024 data, "the fixed broadband coverage" in Macedonia amounts to 87.29% of fast broadband coverage. However, public broadcasting enterprises in Macedonia need to strengthen their capacities to develop more efficient optical infrastructure and broadband access in "areas with insufficient commercial investment from operators" (EC, 2025: 73). The report recommends that the Agency for Electronic Communications strengthen its independence, which is deemed insufficiently developed in the process of Macedonian digital development. Nonetheless, the Macedonian National eService Portal has enhanced its services for businesses through an improved user access interface and expanded e-services, providing new services requested by citizens, such as birth, marriage, and citizenship certificates, which are now available online. In January 2025, the national cybersecurity strategy was adopted, which is expected to mainstream government efforts to strengthen state infrastructure and enhance cybersecurity expertise. Consequently, the established national Computer Incident Response Centre has further improved and developed its operational capacities and staff. While a new telecom operator was successfully integrated into the now greatly expanded Macedonian 5G mobile network, the report nevertheless insists on the continuous implementation of the EU 5G Cybersecurity Toolbox measures in this respect. Thus, the developed legal and normative framework still lacks effective implementation in practice, as the existing tools for e-identification and e-signature certificates are not user-friendly and are not free of charge for Macedonian citizens. As shown in Table IV, alignment with the Digital Compass for 2030 in Macedonia has contributed to the strategic development of digital policies, driving the country's digital transformation in the near future. This cultural shift marks a milestone, representing the negotiated groundwork for sustained innovation and future development of public service delivery, maximising the benefits of digital transformation for all.

TABLE IV. ASSESSMENT OF STRATEGIC DIGITAL POLICY DEVELOPMENT IN 2025

No.	Dynamic Policy Maturity Benchmark Assessment	
	Assessment criteria	***Fulfilled (+) Not fulfilled (-) Partially fulfilled (PF) No information (NI)***
1	**data protection**	PF
2	**privacy rights**	PF
3	**digital governance**	+
4	**confidentiality of communication**	PF
5	**digital inclusion/connectivity**	PF
6	**data literacy**	+
7	**digital skill development**	+
8	**level of digital innovation**	PF
9	**inclusive investment**	+
10	**digital equity for all**	+
11	**data ethics**	+
12	**cybersecurity**	+
13	**availability in continuous operations**	+
14	**open data**	+
15	**data democratization**	+
16	**level of data culture**	+
17	**digital by default rules**	+
18	**use of digital twins/simulation**	PF
19	**interoperability**	+

c) The United Nations Development Program assessments of digital policy optimisation in Macedonia

The United Nations Development Program (hereafter UNDP) digital readiness assessment of Macedonia from 2023 states that the country is making significant progress in its digital transformation journey [22]. Designed to help Macedonia strategically and holistically address inclusive digital transformation, it also provides insight into plans to build a national digital ecosystem based on the principles of universality, responsibility, equality, and openness. Evidence shows that change is visible in the newly introduced infrastructure, government, regulation, business sector, and society in general. Macedonia now ranks 30th globally for mobile broadband speeds, with 3G and 4G coverage successfully rolled out and 5G now available in 16 cities. In line with the EU framework, digital governance is steadily progressing, with the Macedonian startup ecosystem primarily centred in Skopje. Improved digital infrastructure is present almost everywhere, although most technological innovations occur in the capital, Skopje, which drives the advancement of the entire country. Notably, half of Macedonian businesses use online tools and e-commerce, and approximately 80% of the population is connected to the internet. Progress is therefore evident in improved regulatory transparency, e-commerce, and cybersecurity.

Another report from the UNDP and Indago in 2023 shows that Macedonia has made steady progress in citizens' use of digital tools, reducing the detected digital gap in the country [23]. Macedonian citizens have invested significantly in purchasing technologies such as smartphones, indicating that digital devices are now an integral part of everyday life. However, use of the national e-Uslugi platform indicates a relatively low level of awareness and, consequently, limited use of such digital public services. Although there are reported difficulties with using the webpage, overall satisfaction with the provided digital public services is generally high. However, many citizens still strongly prefer direct human contact and in-person visits to offices. In conclusion, the UNDP report emphasises that the anticipated technological impacts of digital transformation in Macedonia are most evident, according to various data sources, in positive changes in the use of banking services, work or study, and access to government or public services.

TABLE V. ASSESSMENT OF DIGITAL POLICY OPTIMISATION IN MACEDONIA, 2023

No.	**Dynamic Policy Maturity Benchmark Assessment**	
	Assessment criteria	***Fulfilled (+) Not fulfilled (-) Partially fulfilled (PF) No information (NI)***
1	**data protection**	PF
2	**privacy rights**	PF
3	**digital governance**	+
4	**confidentiality of communication**	PF
5	**digital inclusion/connectivity**	PF
6	**data literacy**	PF
7	**digital skill development**	PF
8	**level of digital innovation**	PF
9	**inclusive investment**	PF
10	**digital equity for all**	PF
11	**data ethics**	PF
12	**cybersecurity**	PF
13	**availability in continuous operations**	PF
14	**open data**	PF
15	**data democratization**	PF
16	**level of data culture**	PF
17	**digital by default rules**	PF
18	**use of digital twins/simulation**	NI
19	**interoperability**	PF

If the Digital Compass for 2030 alignment has had a clear impact on Macedonia's strategic development of digital policies according to European Commission reports, then the policy optimisations made in 2023, based on the DPMBM assessment, have contributed to the country's digital transformation, despite its limited capacity for change (Table IV).

D) Digital Transformation According to Domestic Policy Progress Assessments

This section evaluates policy progress for 2024–2025, based on the Ministry of Digital Transformation's reports, as an integral part of the Strategic Plan 2025–2027. According to the Strategic Plan 2025–2027, Macedonia's digital development is crucial for the country's future prosperity [24]. The Ministry focuses on digitalisation and technological advancement, aiming to improve digital skills among public sector employees and citizens, and to enhance cybersecurity. The overall goal is to enable Macedonia's administration to manage, establish, and unify the state system with local and regional authorities and relevant institutions that provide services to citizens. Given its special role in supporting the digital transformation of state authorities in Macedonia, the Ministry has defined the following areas of strategic interest as priorities: 1) digitalisation of public administration and delivery of electronic services, 2) digital connectivity and information and communication technology infrastructure, 3) cooperation and establishment of joint standards for cybersecurity, and 4) strong digital skills and digital literacy among administration staff. Therefore, the Ministry develops policies to advance the information society, digitalisation, and cybersecurity using the most advanced technology to optimise and automate public administration processes. The extent to which the Ministry has developed policies to create an information society is evident from the National Broadband Competency Office report for 2025 [25]. Taking into account the results of the National Broadband Competency Office report for 2025 and based on data from the Ministry of Digital Transformation's reports for 2024–2025 (aligned with Strategic Plan 2025–2027), the following two tables (Table VI and Table VII) assess Macedonia's digitalisation process, in line with the DPMBM [26].

TABLE VI. ASSESSMENT OF THE DIGITALIZATION PROCESS – 2024-2025 MINISTRY OF DIGITALISATION REPORTS

No.	Dynamic Policy Maturity Benchmark Assessment	
	Digitalization process	***Assessment criteria: Existing, Current, Evolving, Transformative, Sustainable***
1	**Institutions and mechanisms**	Existing, Evolving, Transformative, Sustainable
2	**Legal representation of stakeholders**	Existing, Evolving, Transformative, Sustainable
3	**Policy governance**	Existing, Evolving, Transformative, Sustainable
4	**IT Governance in Government (e-Government)**	Existing, Evolving, Transformative, Sustainable

TABLE VII. ASSESSMENT OF DIGITALISATION PROCESS, 2024 – 2025 (ACCORDING TO STRATEGIC PLAN 2025-2027)

No.	Dynamic Policy Maturity Benchmark Assessment	
	Digitalization process	***Assessment criteria: Existing, Current, Evolving, Transformative, Sustainable***
1	**Digitalisation**	Existing, Evolving, Transformative, Sustainable
2	**Data protection**	Existing, Evolving, Transformative, Sustainable
3	**Privacy**	Existing, Evolving, Transformative, Sustainable
4	**Confidentiality of communication**	Existing
5	**Digital governance**	Existing, Evolving, Transformative, Sustainable
6	**Connectivity**	Existing
7	**Data literacy**	Existing, Evolving, Transformative, Sustainable
8	**Digital skill development**	Existing, Evolving, Transformative, Sustainable
9	**Level of digital innovation**	Existing, Evolving, Transformative, Sustainable
10	**Inclusive investment**	Existing
11	**Digital equity**	Existing, Evolving, Transformative, Sustainable
12	**Data ethics**	Existing, Evolving, Transformative, Sustainable
13	**Cybersecurity**	Existing, Evolving, Transformative, Sustainable
14	**Open data**	Existing, Evolving, Transformative, Sustainable
15	**Data accessibility / democratization**	Existing, Evolving, Transformative, Sustainable
16	**Level of data culture**	Existing, Evolving, Transformative, Sustainable
17	**Digital default rules**	Existing, Evolving, Transformative, Sustainable
18	**Use of digital twins**	Existing
19	**Use of digital symulation**	Existing, Evolving, Transformative, Sustainable
20	**Availability in continuous operations**	Existing
21	**Interoperability**	Existing, Evolving, Transformative, Sustainable

Assessment of the Macedonian digitalisation process shows clear rapid progress in the country's digital transformation, reflecting the Ministry's diligent efforts to improve challenging conditions imposed by domestic issues and EU accession difficulties.

V. DIGITAL GOVERNANCE IN MACEDONIA

Negotiating and navigating digital governance in Macedonia has been, and remains, a challenging task not only for government institutions but also for regional and local authorities. It would be highly presumptuous to propose a straightforward solution to the challenges identified in the assessments presented in this paper. However, a more sensible approach is adopted, supporting a realistic understanding of policy development and policymaking mechanisms. Developing a strong and functional digital environment in Macedonia – is discussed within the framework of the trade-offs that the government, its authorities, and institutions can realistically negotiate not only among themselves, but more importantly with EU institutions as an integral part of the accession process [27]. It is important to note that, although uncommon, such an approach is indispensable if the young Macedonian democracy is to position itself as an equal partner in negotiating the best possible outcomes of the forthcoming EU accession process, including the ways in which the country will digitally develop and transform its society. As noted by Naumovski and Popovic, the development of information technologies creates new legal challenges that necessitate proactive governance [28]. Empowering appointed state officials and authorities at all levels of government to proactively manage state policies within budgetary provisions and financial constraints requires modernising the administrative system through digital optimisation of processes and services, as well as re-engineering in line with the 'once-only principle'. A more inclusive and open public administration system therefore requires promoting its services with strong user involvement to ensure the successful implementation of digital plans. The 'trade-offs' that Macedonian authorities must negotiate and renegotiate at all levels of government – both among themselves to achieve consensus among all political parties and citizens, and with EU authorities – to protect national interests and the positive outcomes of digital development policies and overall social transformation, require persistent dedication and diligence.

The policy assessment in this paper shows that development plans have established a set of clearly defined, realistic strategic options that relevant Macedonian stakeholders can successfully negotiate or renegotiate to

 ISBN: 978-1-68558-400-9

ensure positive outcomes in the evolving digital landscape. A comprehensive digital roadmap with a strategic outlook, integrating digital technologies into government services and supported by updated and improved laws and regulations, now provides an inclusive (re)negotiation framework managed by knowledgeable and skilled professionals. A solid foundation of sustained and enhanced digitalisation activities, measures, and policies, continuously monitored and evaluated by impartial agencies, therefore ensures fact-based negotiation positions that support Macedonian national interests. Moreover, the most important component in the digitalisation of Macedonian public services is the development of digital skills and education among civil servants, which has so far been carefully balanced to meet the challenging demands of the employment market and the professionals produced by the education system [29]. Thus, the digitalisation of Macedonian public services concerns secure, high-performing, and sustainable digital infrastructure. Although it is not possible to provide a comprehensive overview of Macedonian digital infrastructure, the analysed documents and research on the subject are sufficient to present the current state of the developed digital foundation in the country. Advancing Macedonian digital infrastructure depends greatly on the state's ability to provide and sustain the necessary resources to support digital development in an ecological manner, as required by the EU. As emphasised by Naumovski and Kamboska, the Internet and electronic communication redefine traditional legal frameworks, particularly regarding liability, privacy, and data protection, which are essential considerations in infrastructure development [30]. Emerging threats and vulnerabilities identified in current digital system development in Macedonia have been sufficiently addressed through coordinated policies and stable regulatory frameworks. As previously mentioned, trade-offs in this context are made during internal negotiations among Macedonian stakeholders responsible for developing the high-speed broadband (fiber optics) system, which enables faster access to digital services, and between state authorities and businesses that use up-to-date digital technologies and mobile electronic devices to ensure efficient and secure digital ID networks, data transmission and transactions, digital connections, data hubs, and centres.

Finally, the digitalisation of Macedonian public services concerns also the digitalisation of businesses, specifically the process that enables the successful integration of digital technology into various aspects of a company’s operations, activities, and strategies. The introduction of new and innovative digital technologies into business processes improves decision-making and creates new opportunities for growth and profit. Evidence so far shows that the digital development of businesses in Macedonia slowly creates an agile and innovative culture that is free to experiment with new technologies and knowledge.

The overall digitalisation of Macedonian society, although greatly improved in recent years, has proved to depend largely on how digital governance is implemented within the EU integration alignment process. Therefore, it is of utmost importance for the respective government to position Macedonian digital development and transformation in the global context, and to ensure that negotiated trade-offs are based on critical thinking when digital technological innovations promise progressive social change with no alternative.

VI. Conclusion and Future Work

Since 2021, the digital transformation of Macedonian society has created opportunities for the government to address outstanding structural, economic, political, and social challenges. This paper demonstrates how Macedonia can substantially benefit from digital transformation to advance economic development within a fully functional and efficient government administration and rule of law. Macedonia has fully embraced digital transformation, and the newly formed Ministry of Digitalisation has ensured a regulatory framework that is future-proof and sustainable, with policies that benefit both businesses and citizens. The strategic development framework for its programmes and subprogrammes has enabled meaningful connectivity as a foundation for digital development and transformation in Macedonia. People-centred and government-centric digital transformation is built on and evolves from a solid IT infrastructure, which supports the progress of sector-centric digital transformation. It is evident that Macedonia’s advancements in digital public services have surpassed the Western Balkans average, while the proactive adoption of cutting-edge digital technologies, such as big data, cloud computing, and AI by enterprises, greatly contributes to the country’s economic progress [31]. Half of the country’s businesses use online tools and e-commerce, while 80% of the population is connected to the internet [32]. There has been visible improvement in legislation and the regulatory framework, marking substantial progress in transparency, e-commerce, and cybersecurity. Similarly, efforts in digital governance and public procurement have been enabled by alignment with EU standards, supported by significant progress in digital infrastructure, including broadband coverage and internet access.

In this paper, the reappropriated DPMBM, as an innovative research tool, has enabled data analysis whose results can inform a whole-of-government approach to policy making and a whole-of-society approach to the digital transformation of Macedonian society. Aiming to serve as a reference for the development of future improved digital governance and Macedonian digital transformation, it also provides ample evidence to support further research on this global phenomenon that affects every person and country worldwide. According to the expert report on the Western Balkans Digital Economy and Society Index 2025, it is clear that additional data is needed to compare the digitalisation process in countries aspiring to EU membership [33]. The identified data gaps in digital transformation indicators, as demonstrated by the Macedonian case, require further research, while the future digitalisation process demands a strategic approach with a broader agenda and policies for developing data-driven technological infrastructure.

ACKNOWLEDGMENT

This research paper on digital transformation of Macedonian society has been prepared within the framework of the project "Factors of Intellectual Property Management and Digital Transformation of Macedonia (2025-2026)", financed by the Ministry of Education and Science. The project is implemented by the Public Scientific Institution, Institute for Intellectual Property – Skopje.

REFERENCES

[1] United 4 Smart Sustainable Cities, "Policy benchmarks for digital transformation of people-centred cities", 2024, Accessed: 2026-03-10. [Online]. Available at: file:///C:/Users/Korisnik/Downloads/2401530_u4ssc-policy-benchmarks-for-digital-transformation-of-people-centre.pdf .

[2] Republic of North Macedonia State Statistical Office, "Statistical Yearbook of the Republic of North Macedonia 2023", 2023, Accessed: 2026-03-15. [Online]. Available at: https://www.stat.gov.mk/PrikaziPublikacija_en.aspx?id=34&rbr=897 .

[3] ITU, "Banchmark of fifth-generation collaborative digital regulation", 2021, Accessed: 2026-03-15. [Online]. Available at: https://www.itu.int/dms_pub/itu-d/opb/pref/D-PREF-BB-2021-01-PDF-E.pdf .

[4] B. Klievink, N. Bharosa and Y. H. Tan, The collaborative realization of public values and business goals: Governance and infrastructure of public-private information platorms. Government Information Quarterly, 33(1), 2016: 67-79.

[5] B. Klievink, N. Bharosa and Y. H. Tan, The collaborative realization of public values and business goals: Governance and infrastructure of public-private information platorms. Government Information Quarterly, 33(1), 2016: 67-79.

[6] OFFICIAL GAZETTE OF THE REPUBLIC OF NORTH MACEDONIA, "Методологија за стратешко планирање и подготвување на Годишна програма за работа на Владата на Република Македонија (Methodology for Strategic Planning and Creation of Yearly Operational Programmes for the Republic of Macedonia Government), 124/2008, 2018, 2019; GOVERNMENT OF THE REPUBLIC OF NORTH MACEDONIA, BRITISH EMBASSY SKOPJE, UK INTERNATIONAL DEVELOPMENT, CENTER FOR CHANGE MANAGEMENT, Процената на влијанието на регулативата (ПВР) во Северна Македонија (Assessment of the Legislation Impact in North Macedonia), 2023.

[7] T. Sowell, A Conflict of Visions: Ideological Origins of Political Stuggles, William Morrow&Co., 1987.

[8] POLICY AND PRACTICE IN PROVIDING AN ELECTRONIC IDENTIFICATION SERVICE, Version: 2.0, EVROTRUST, 2024, Accessed: 2026-06-29. [Online]. Available at: https://evrotrust.mk/en/wp-content/uploads/2024/05/ET-Macedonia-Policies-and-Practices-for-eID-Service-Provision-v.-2.0-en.pdf .

[9] Z. T. Kamilovska, NATIONAL REPORT FOR NORTH MACEDONIA ON ELECTRONIC EVIDENCE AND VIDEOCONFERENCING, Project DIGI-GUARD 2023, 2023, Accessed: 2026-06-29. [Online]. Available at: https://www.pf.um.si/site/assets/files/7177/north_macedonia_evidence_final.pdf .; T. Kalamatiev and N. Murdzev, "Electronic documents and electronic signatures as a means of concluding an employment contract", Časopis za radno i socijalno pravo, XVIII, 2/2024: 23-37.

[10] I. Margel, Y. Gong and J. Bertot, Agile government: Systemic literature review and future research, Government Information Quarterly, 35(2), 2018: 291-298.

[11] A. S. Jovkovska, Đ. A. Bogdanovska and I. B. Karakašev, "Digitalna Dekada Kvalitet na e-uslugi Study on the quality of e-services for citizens in North Macedonia", 2024, Accessed: 2026-04-10. [Online]. Available at: https://knowledge-center.org/images/DigitalDecade/Studies/Citizen_-_CRM088_-_EN.pdf .

[12] D. Nikolovski and A. Stojanovski, "ÖGfE Policy Brief 22 Digital Europeanisation in North Macedonia: slowly moving forward", 2022, Accessed: 2026-02-12. [Online]. Available at: https://www.wb2eu.eu/wp-content/uploads/2023/01/PB-222022.pdf .

[13] UNDP, "Digital Readiness Assessment North Macedonia", 2024, Accessed: 2026-02-16. [Online]. Available at: https://www.undp.org/sites/g/files/zskgke326/files/2024-01/dra_mk_06.pdf .

[14] J. R. Gil-Garcia, M. Gasco-Hernandez and T. A. Pardo, "Beyond Transparency, Participation, and Collaboration? A Reflection on the Dimensions of Open Government", Public Performance & Management Review, 2020.

[15] I. Mergel, Y. Gong and J. Berot, "Agile government: Systematic literature review and future research", Government Information Quarterly, 35, 2018: 291-298).

[16] I. Mergel, Y. Gong and J. Berot, "Agile government: Systematic literature review and future research", Government Information Quarterly, 35, 2018: 291-298).

[17] DEVELOPMMENT FRAMEWORK OF THE REPUBLIC OF NORTH MACEDONIA MKD 2030, 2022, Accessed: 2026-02-12. [Online]. Available at: https://mkd2030.mk/wp-content/uploads/2022/06/Ramka-za-razvoj-MKD2030-EN.pdf

[18] THE NATIONAL DEVELOPMENT STRATEGY 2024-2044, Accessed: 2026-02-14. [Online]. Available at: https://www.nrs.mk/en-GB/dokumenti.nspx .

[19] UNDP, "Digital Readiness Assessment North Macedonia", 2024, Accessed: 2026-02-16. [Online]. Available at: https://www.undp.org/sites/g/files/zskgke326/files/2024-01/dra_mk_06.pdf ; Agenda 2030, 2015, Accessed: 2026-02-16. [Online]. Available at: https://docs.un.org/en/A/RES/70/1.

[20] 2030 Digital Compass: Your Digital Decade, Accessed: 2026-02-12. [Online]. Available at: https://futurium.ec.europa.eu/en/digital-compass ; B. Bajić, S. Damjanović and G. Pastorovic, "Policy Brief Digital Transformation in the Western Balkans", 2024, Accessed: 2026-02-12. [Online]. Available at: https://prod5.assets-cdn.io/event/10047/assets/8284323237-339dde7c43.pdf .

[21] EUROPEAN COMMISSION, "Commission Staff Working Document North Macedonia 2025 Report, SWD(2025) 753 final", 2025, Accessed: 2026-02-12. [Online]. Available at: https://enlargement.ec.europa.eu/document/download/267b368e-6b55-4a42-bb72-6395593de4da_en?filename=north-macedonia-report-2025.pdf .

[22] UNDP, "Digital Readiness Assessment North Macedonia", 2024, Accessed: 2026-02-16. [Online]. Available at: https://www.undp.org/sites/g/files/zskgke326/files/2024-01/dra_mk_06.pdf ; Agenda 2030, 2015, Accessed: 2026-02-16. [Online]. Available at: https://docs.un.org/en/A/RES/70/1.

[23] UNDP and INDIGO, "Research Report on the Use of Digital Tools by Citizens and Assessing the Digital Gap", 2023, Accessed: 2026-02-20. [Online]. Available at: https://www.undp.org/sites/g/files/zskgke326/files/2024-02/2023_DigitalGapAssesment_v.2.0_ENG_Final.pdf .

[24] MINISTRY OF DIGITAL TRANSFORMATION, “Strategic Plan 2025 – 2027 Ministry for Digital Transformation”, 2025, Accessed: 2026-02-25. [Online]. Available at: https://portal.mdt.gov.mk/post-body-files/godisni-planovi-file-slsv.pdf .

[25] REPUBLIC OF NORTH MACEDONIA BROADBAND COMPETENCE OFFICE. Reports, 2025, Accessed: 2026-03-20. [Online]. Available at: https://bco.mdt.gov.mk/?page_id=228&lang=en .

[26] MINISTRY OF DIGITAL TRANSFORMATION, Reports, 2024, 2025, Accessed: 2026-02-07. [Online]. Available at: https://mdt.gov.mk/en-GB/dokumenti/izvestai .

[27] K. Kolozova, “Monitoring Influence & Disinformation Campaignes in North Macedonia”, 2025, Accessed: 2026-06-02. [Online]. Available at: https://dwp-balkan.org/monitoring-influence-disinformation-campaigns-in-north-macedonia/ .

[28] G. Naumovski and D. Popovic, Information Technology Law, Skopje: Epoka University Library, 2013.

[29] ITU, “Digital Skills Assessment Republic of North Macedonia”, 2021, Accessed: 2026-02-27. [Online]. Available at: https://www.itu.int/en/ITU-D/Regional-Presence/Europe/Documents/Publications/2022/North%20Macedonia%20Digital%20Skills%20Assessment/ITU_NorthMacedonia_DigitalSkillsAssessment_20211223_FINAL_20220606.pdf .

[30] G. Naumovski and M. Kamboska, Intellectual Property Law in Macedonia, The Netherlands: Kluwer Law International B. V., 2022.

[31] B. Bajić, S. Damjanović and G. Pastorović, “POLICY ANSWERS Policy Brief. Digital Transformation in the Western Balkans”, 2024, Accessed: 2026-02-27. [Online]. Available at: https://mir.org.mk/wp-content/uploads/2024/10/policy_brief_digital_transformation_web.pdf .

[32] UNDP, “Digital Readiness Assessment North Macedonia”, 2024, Accessed: 2026-02-16. [Online]. Available at: https://www.undp.org/sites/g/files/zskgke326/files/2024-01/dra_mk_06.pdf .

[33] Western Balkans Digital Economy and Society Index 2025 Report, Accessed: 2026-06-16. [Online]. Available at: 2026, file:///C:/Users/Korisnik/Downloads/WB%20DESI%202025%20Report.pdf .

Digital Authoritarianism

Brian Gannon
Fachgruppe IT und Technik
IU Internationale Hochschule,
Erfurt, Germany
e-mail: brian.gannon@iu.org

***Abstract*— Modern democratic societies have in recent centuries adopted a social contract whereby deference is shown by citizens, to state or public authority (and to the laws and institutions through which that authority is expressed) in return for a voice that determines the nature and extent of that authority. This social contract is now threatened by the emergence of a small number of private organizations which have assumed some of the functions of state authority. This shift from public to private authority has been enabled by the extent to which these private organizations – very large online platforms – use powerful tools to exploit vast amounts of data amassed over the past few decades. It has profound implications for individuals, particularly in respect of their fundamental human rights, and it undermines the basis of the contract that binds citizen and state. This gradual expansion of private digital power at the expense of public authority, described here as digital authoritarianism, is the subject of this research.**

Keywords- digital sovereignty; authoritarianism; online surveillance.

I. Introduction

Digital platforms have developed highly lucrative and successful business models which are predicated on the fact that their raw material – personal data – is provided willingly and for free by individual users in exchange for access to online services provided by the digital platforms. They have taken advantage of rapid advances in technology, widespread connectivity, and associated network effects to establish dominant positions in the online marketplace. This condition is greatly enhanced by the conjunction of ubiquitous intelligent devices (the 'Internet of Things' – 'IoT') and powerful Artificial Intelligence (AI) technology.

From this vantage point, digital platforms have induced in users a profound deference to digital authority, engendered primarily by a demand for convenience. While the economic and normative effects of digital authoritarianism are eloquently described by Cohen, Pasquale, Zuboff and others, this paper focuses on describing the dependence/deference duality and its intersection with convenience [10][36][49]. The entrenchment of digital authority is diminishing human agency and thus has significant implications for society, its governing institutions and the rule of law.

The research is exploratory. It assesses the operation of digital platforms in enabling control and surveillance, recognizing that platforms have other powerful capabilities and effects, alluded to but not investigated. It starts by setting out contextual underpinning for the analysis in Section 2; describing the types of enabling infrastructure that will transform its current capability in Section 3; and elaborating on the entrenchment of the digital authoritarian construct as an inevitable outcome of these developments in Section 4. Section 5 summarizes the role of regulation in constraining the digital authoritarian environment, and Section 6 presents a conclusion and future work.

II. Context: Antecedents, Definitions, Foundations

A. *Summary of literature and definitions*

The most relevant antecedents for this discussion in the context of wider academic research are the notions of digital sovereignty, and of control and choice in cyberspace. The analysis starts by defining key terms.

'Digital platform' is used here as an umbrella term to describe private companies that provide large-scale digital infrastructure and services including search, social media, cloud computing, e-commerce, and operating systems to billions of users worldwide. It covers various aspects of these firms, including both technical capability and their business models. The digital platform is highly visible in a way that computing rarely was in the past. It is an entrenched social construct as much as a technical phenomenon, infecting language, custom, behavior, norms and even personal relationships. It not only facilitates human expression and interaction but also molds it, shaping opinion, generating ideas, entertaining, unearthing and disseminating knowledge, enabling commerce, and aiding destruction. It is nonetheless a new phenomenon, and its effects are incompletely understood.

'Authoritarianism' is usually understood in political terms and often used synonymously with the term 'totalitarian'. Diebert defines authoritarianism as '…state constraints on legitimate democratic political participation, rule by emotion and fear, repression of civil society, and the concentration of executive power in the hands of an unaccountable elite.' [14]. By extension, in this research 'digital authoritarianism' defines a state of society where important aspects of a citizen's life are determined by data gathered by ubiquitous sensors and analyzed to provide real-time regulation, recommendation and guidance. In such circumstances, the data (or the party holding the data) is the authority and the citizen is authorized to act according to the data. In the context of the online world an essential element of authoritarianism is the concentration of power in the hands of a few unaccountable individuals.

'AI' is described as adaptive computer activity that 'arises when computer systems perform more complex tasks which previously required human intelligence …<involving> machines which are capable of analyzing situations and learning for themselves and then generating answers which may not even be foreseen or controlled by their programmers.' [38]. The distinguishing factor, emphasized in this definition, is foresight: traditional computing is eminently predictable - a computer does exactly what it is programmed to do, no more and no less, while AI systems may not. The deliberate 'training' of AI systems enhances their ability to predict and perform tasks. This is machine learning (ML), which seeks to mimic biological thought and learning processes [40].

The IoT is 'an infrastructure in which billions of sensors embedded in common, everyday devices … record, process, store and transfer data and… interact with other devices or systems' [2]. IoT devices are often capable of actuation - taking action to adjust physical processes such as modifying temperatures. Hildebrandt conflates IoT with AI to define 'data-driven' agents - an 'ambient intelligence' that forms a 'cyber-physical infrastructure' which puts physical infrastructure such as homes, vehicles, fields, bio-forms and weapons online, 'to enable persistent monitoring and surreptitious adaptation.' [23]. In this research, the term 'intelligent agent' is used is a similar manner as a shorthand to cover the co-existence of IoT and AI capabilities in such agents.

B. Digital Sovereignty

The meaning of the term 'digital sovereignty' has changed over time, although the central concern – who controls and regulates the online environment – remains the same. Initially focused on regulation of cyberspace, subsequent inquiry addressed misuse of digital technology by state actors. For example, Diebert was predominantly concerned with risks of state authoritarianism through use of digital technology [14]. As the reach and scope of digital platforms grew, researchers explored the transfer of authority from state to private actors, described by Pasquale as 'functional sovereignty' [35]. In 2020, the chairman of the US House Judiciary Subcommittee on Antitrust, Commercial and Administrative Law questioned the heads of Amazon, Apple, Google and Facebook, noting that his committee's investigation '…goes to the heart of whether we, as a people, govern ourselves, or whether we let ourselves be governed by private monopolies.' [5].

Floridi defines sovereignty as a '…a form of legitimate, controlling power', noting that until recently, it had always been 'analogue', that is, bounded by the physical world (territories, cities, nations…), and controlling physical limited resources (land, oil, gold…) [19]. In the digital age, the infosphere is not a territory and '…digital assets are largely private and subject to market forces …' and '…whoever will win the fight for digital sovereignty will determine the lives of all people on both sides of the digital divide.' [19].

Contemporary digital sovereignty is understood as an attempt by the state to reassert control over digital assets and information, so that the power of digital platforms to dictate terms is diminished. It implicitly acknowledges the incursions made by digital platforms into state domains and seeks to reverse these. The EU has led the way by laying the foundations of what Bradford describes as a 'digital constitution', meaning '…a set of expansive digital regulations that form a normative and principled foundation for a digital economy and society as construed by European laws.' [4].

Regardless of the form of state governance adopted in the past, authority has invariably been public and, usually, visible: a citizen (or subject) might not always have been familiar with prevailing rules, actors, and institutions but generally knew who was in charge. In democratic societies, the citizen has scope to exercise her own judgement and to conduct her life within agreed legal boundaries. Fundamental rights are guaranteed according to the jurisdiction and these, along with social, economic and other rights (and associated obligations and constraints) afford a greater or lesser degree of *choice*. Similarly, the citizen as consumer may cede authority to private agents to effect control.

C. Control, choice, consent, convenience

The most frequent basis for engagement between user and digital platform is underpinned by the contractual notion of consent, effected through 'opt-in' or 'opt-out' choices. Thus, the individual consents to terms and conditions usually involving the surrender of personal data and receives software-based benefits such as social media apps for this consideration. Sometimes, consent is nugatory - the choice is 'take it or leave it'.

This 'informed consent' mechanism is flawed, if not absolutely '…broken beyond any regulatory repair.' [37]. This is because first, there is a cognitive deficit - the user is rarely fully informed of the nature and scope of her obligations, even if they are documented in the platform's terms and conditions. Second, there is a structural deficit - the user can have little awareness of the wider use of her data when she consents to its collection. Third, there is an asymmetry between the trading partners that is heavily weighted in favor of the platform. A confounding factor is that many IoT devices eliminate the interface between user and data, making it difficult to collect consent.

The tension between control and choice or consent in the online environment is thus relevant. In the early days of the internet there was relative equality of arms between user and provider, but now that digital platforms are weaponized with AI, IoT and other powerful controlling technology, user choice '…will inevitably be influenced by default rules, framing effects, and starting points'. [42]. Yeung identifies the inherent power of such choice architectures in influencing user behavior, that facilitate 'undermining an individual's capacity to exercise independent discretion and judgment'. [47]. Macro-economic and societal studies include Cohen's comprehensive picture of the imbalance between corporate and individual actors online, and how this has unfolded and become entrenched; Pasquale's extensive analysis of 'black box' algorithmic processes that deny choice; and, Zuboff's forcible identification of online economic and social inequality as 'surveillance capitalism'. [10][36][49].

A further overarching reason is convenience. Koops, discussing the role of consent in the context of data protection notes that, '(A)nother challenge of relying on consent is that convenience and people's limited capacity to make rational decisions prevent people from seriously spending time and intellectual effort on reading the privacy statements of every website, app, or service they use.' [28]. The demand for convenience in the digital environment is as persistent as it is offline. Consumers seek instant gratification, and digital platforms oblige with easy to use, intuitive interfaces offering simple 'click here' options, backed by impenetrable, formalistic contracts that are inconvenient to read. Digital platforms take advantage of consumer bias towards convenience and of their propensity to stick with default options, and use this propensity to influence and manipulate user choice [39][41]. The exacerbating effects on control and choice resulting from the widespread deployment of AI/ML technology by digital platforms are analyzed next.

III. Expansion of Digital Platform Capability

A. Changes in digital platform capability

For the most part, citizens appear unconcerned by the coincidence of powerful new control and surveillance technologies in the hands of a few outsize corporations. The services provided by digital platforms are often free, easy to use and have engendered social good. It is plausible to make a case that the benefit of such services, even if provided by a powerful oligopoly, outweigh potential risks.

The counterargument to complacency starts by highlighting the pace and scale of industrialization of surveillance, and the potential for almost unlimited control enabled by three directional trends. These are the *collection vector*: ubiquitous intelligent agents will generate vast real-time data sets; the *analysis vector*: powerful AI/ML technology will turbocharge efficient search, interrogation and manipulation of this data, generating actionable information; and the *distribution vector*: fast global networks will enable platforms to distribute the resulting information more efficiently than ever before. Together, this leads to greater dependence on, and deference to, digital platforms, and to a corresponding diminishment in human agency.

B. Collection, Analysis and Distribution

Notwithstanding the vast scale of data already compiled by digital platforms, and their ability to manipulate it, much data is incomplete and of poor quality. For example, in the UK's National Health Service (NHS), many patient records are still paper-based [32]. The deployment of intelligent agents allows collection of data that is real-time, accurate, and independent of human agency. Conservative estimates suggest that more than 30 billion individual connected devices are already in place, and that annual growth rates of greater than 10% are predicted leading to deployment of hundreds of billions of devices by 2030 [13][26].

As with collection, analysis of data in the past has often been slow, cumbersome, and inaccurate. Now, availability of low-cost, high-performance processing and increased sophistication of analysis techniques mean that digital platforms are increasingly capable of accurate inference.

Together, the widespread deployment of connected smart devices combined with the deployment of advanced AI and ML capability trained on vast rich datasets extends the power and influence of digital platforms. This is because in addition to controlling the means of production of real-time, actionable, accurate information, they also control distribution.

At an infrastructural level, distribution capability is enhanced by 5G networks and almost total ubiquity of smart mobile handsets. This amplifies the power and effect of digital platforms across all spheres of social, commercial and political activity, including distribution of social mores and fashions via recommender and influencer functions, distribution and control of speech by controlling access to news platforms, and distribution of influence over democratic processes by targeting voters.

C. Emerging Effects of Intelligent Agents

Humans are willing to cede agency in this new technology landscape because it is increasingly convenient and easy. There is a view that it will in any case soon be impossible for users to exercise any meaningful agency. Yeung refers to Harari's concept of 'dataism', a perception that the job of processing data into information should be left to intelligent algorithms because the human brain no longer has the capacity to process such vast amounts of data [48]. This operates in tandem with Morozov's solutionism – the belief that technology can fix any problem. Intelligent agents 'offer speed, lower transaction costs and efficiency in decision-making [21].

These viewpoints are evidence of an increasing reliance and deference to digital platforms. This amounts to a delegation of responsibility to unaccountable actors, and implies a more controlled and compliant society. This default to the 'convenience' option, and deference to the digital authorities who control the levers of information production, leads directly to digital authoritarianism.

IV. Digital Authoritarianism

A. Digital authoritarianism

The conduct, ubiquity and reach of digital platforms has been described as 'totalitarian' by sources as ideologically opposed as Cuban communists and American capitalists [22][25]. The idea of extreme surveillance and control is not new: writing in 1787, Jeremy Bentham described the Panopticon and presented his Inspection Principle, paraphrased as 'the more strictly we are watched, the better we behave'[3]. Foucault examined a similar theme in his writing on prisons [20]. More frequently, it is Orwell's 'Big Brother' which provides the modern metaphor for Big Tech's all-seeing eye [34]. In such descriptions, the emphasis is on total, comprehensive and always-on surveillance and control.

For digital platforms, totalitarianism is essentially a capitalistic process. Zuboff lays bare the economic rationale and ruthlessness underpinning the digital platform business model [49]. Her analysis implies that users view these platforms as *authorities*. It is certainly true that they are authoritative on social practice, style, rumor, commerce and much more. As digital platform knowledge is enhanced by information provided and analyzed by intelligent agents, the platform's authority is correspondingly enhanced to the point where it leads to a state of 'digital authoritarianism'.

Like Arendt's totalitarianism, digital authoritarianism is: '… international in organization, all-comprehensive in its …scope, and global in its … aspiration…' [1]. Unlike totalitarianism, the regime is not state imposed, nor does it incorporate tools of terror or extreme coercion. It is not politically inspired but commercial in intent: authority is vested not in the state or in elected officials but in private commercial agents. The nature of the social contract is one where the regime is sustained primarily by data volunteered by users in return for benefits and convenience (better health, safety, education, entertainment…) - a regime of bread and circuses rather than Big Brother.

B. Normative implications of digital authoritarianism

Arendt uses words such as 'indifferent', 'apathetic', 'childish stupidity', and 'isolation' to describe the readiness of the masses to accept totalitarian rule, embodying a form of deference that is not class or ideology-based, but is dependent on 'the atomized and individualized mass'. [1]. While such attributes certainly assist in the establishment of a state of digital authoritarianism, the essential pre-condition is a demand for *convenience*.

This is not on its own a problem: services provided by digital platforms are easy to use, helpful and sometimes, wonderful. They operate as a public good, and like utilities have become indispensable. Search engines tell us what we know; online retailers tell us what we want; social media networks tell us who we are [15].

Yet, in providing convenience, digital services exacerbate the tyranny of choice [42]. The anxiety induced by choice is relieved by authoritative decision-makers in the form of digital assistants (Alexa, Siri…), recommenders (GoCompare, Trip Adviser…), bots (autofill, chatbots…), and more. These are increasingly informed by distributed intelligent agents collecting data points and measurements that enhance the accuracy and relevance of the choice presented. AI-based inference lends further authority to digital platforms because of its predictive capability and 'wisdom'.

This authority is at the heart of automation bias - the propensity for humans to defer to automated decision-making systems despite the presence of contradictory information. Errors of automation bias tend to occur when decision-making is dependent on computers or other automated aids, and the human is in an observatory role but able to make decisions [12]. Such deference to digital authority is increasingly problematic as humans integrate AI/ML decision makers into everyday life. Its impact is perhaps obvious in critical decision-making fields such as weapons control systems, but the effect of embedding such bias in widespread IoT – for example in recommender systems – could be debilitating.

The normative consequences of digital authoritarianism for the human rights of the individual are profound. The first proceeds from the location of authority in private commercial agents rather than in the state. The international human-rights architecture is constructed vertically: its instruments bind states and conceive of the individual as requiring protection principally against public power. The rights most obviously engaged by pervasive observation — the right to privacy under Article 7 of the Charter of Fundamental Rights of the European Union (CFREU), Article 8 of the European Convention on Human Rights (ECHR), and Article 12 of the Universal Declaration of Human Rights (UDHR) — were drafted against the spectre of state intrusion [8][11][45].

The articulation of Article 8 CFREU ("Protection of personal data") was the first time a constitutional-level instrument expressly separated data protection from the older right to privacy and gave it standalone status. The moving force was the need to give constitutional permanence to a right that the Article 7 privacy guarantee could not adequately deliver in an era of automated data handling. What it seeks to protect is the individual's informational self-determination — the fair, transparent and supervised processing of their personal data — as a right distinct from, though closely related to, private life.

A second implication arises from the consensual character of the regime, sustained as it is by data volunteered in exchange for benefit. The normative danger is that the idiom of voluntary exchange recasts a rights-impairment as an ordinary market transaction, thereby dissolving the claim that any right has been infringed at all. This is a pivotal move, for it shifts the question from whether rights are violated to whether they may be validly bartered — a question that engages the partial inalienability of fundamental rights.

Finally, there is the impact on freedom of speech. This is particularly acute, because the construct of digital authoritarianism cedes editorial power over public discourse to the platform. Where the dominant forums for political, civic and cultural expression are privately owned, the practical scope of speech is determined solely by the terms of service drafted unilaterally and enforced at scale by automated systems. The platform is thus a de facto regulator of expression, free to suppress lawful speech that no state would legitimately restrict and, conversely, to amplify or demote material on commercial rather than rights-based criteria. The normative upshot is that freedom of expression, conceived as a shield against the state, offers scant protection against a commercial sovereign whose control over the conditions of speech is more total, and less visible, than that which the right was designed to resist.

C. Operational implications of digital authoritarianism

Ceding authority to digital platforms raises practical issues of security and privacy. Since the IoT comprises billions of connected devices, the attack surface (the possible range of targets) is vastly expanded. Vulnerability is compounded by the fact that many IoT devices are not implemented with security in mind, perhaps because they perform non-sensitive tasks, or because they are designed for implementation within secure network perimeters and not intended for direct connection to web servers.

The broader issue is that apathy and deference to digital authority has inured many to data privacy infringement, and engendered data promiscuity. Data generated by intelligent agents may thus overwhelm flimsy safeguards for individual privacy that derive from the most fundamental common law and constitutional principles. In the light of these risks, the question of how digital authoritarianism and its enabling platform technologies are regulated and controlled becomes profoundly important. This challenge is addressed next.

V. REGULATING DIGITAL PLATFORMS

A. Challenges with regulation of Digital Platforms

Efforts to curtail the power of digital platforms can seem Sisyphean. Cohen describes a surveillance-innovation complex in which surveillance (and control) is 'light, politically nimble and relatively impervious to regulatory constraint.' [10] This is underscored by frequent infringements and grandstanding self-regulation on the part of digital platforms, and by weak enforcement by regulators. The progressive deployment of intelligent agents, AI and pervasive distribution intensifies the potential for harm.

At the same time, regulators suffer from structural disadvantages compared to digital platforms. Marsden talks of the Portarlington effect, a shorthand for the extreme inequality of arms between small national regulators and powerful digital corporations [24]. Regulators suffer from informational and capability disadvantages compared to digital firms. They are poorly placed to conduct the constant monitoring required when proactive remedies are imposed, and they are exposed to the risk of regulatory capture. This frequently results in delayed investigation of infringement and lax or ineffective enforcement.

Nonetheless, legislators across several jurisdictions are attempting to curb the power of digital platforms. In Europe, under the banner of the Digital Single Market initiative, lawmakers have launched a significant programme of legislation that places significant obligations on large digital platforms, notably the Digital Services Act (DSA) and the Digital Markets Act (DMA) [17][18]. In the UK, the Online Safety Act (OSA) introduces liability for harm caused by specific digital firms [44]. Even in the US, where there has long been a broad interpretation of First Amendment rights, there have been moves to provide online privacy for citizens [7]. Digital platforms also conduct self-regulation by contractually setting norms or 'house rules' on their platforms. In this way they govern policy and usage and can take action against infringement, such as exclusion of a user.

These regulatory initiatives are taking place in the context of today's challenges, even though the next wave of digital firepower is already here. The debate has already shifted, and the stakes considerably raised, by the widespread deployment of smart ubiquitous devices and machine learning technology.

B. Possible models of regulation

The complexity of the digital environment is engendered in part by the multiplicity of stakeholders involved. Digital platforms typically operate as multi-sided markets, and thus have many constituent actors (content providers, customers, advertisers, retailers, regulators and so on). Governance of intelligent agents used by these platforms will require a form of algorithmic regulation, an activity that necessarily requires cooperation of these various state and private actors, perhaps by adopting a model of decentered regulation.

Other proposals include regulating intelligent agents and their underlying algorithms along the lines of the pharmaceutical industry, using a central agency (the 'FDA for algorithms') to vet for safety and efficacy [43]. Marsden argues in favor of OffData, a new regulatory agency that '…draws on competition, technology regulation, fundamental human rights, privacy, and free expression to ensure a fair and neutral deal for digital prosumers and citizens [30]. In any scenario, given the pace, volatility and complexity of the digital environment, it is difficult not to envision the need for complex *dynamic* regulation as proposed by Murray [31].

There are calls also for international instruments building on shared principles of administrative law, which could resolve conflicts of law in the digital economy, although such initiatives might be difficult to achieve [29]. Regional attempts at regulation stand a better chance. For example, the EU's AI Act applies to 'AI systems used for indiscriminate

Copyright (c) IARIA, 2026. ISBN: 978-1-68558-400-9

surveillance applied in a generalized manner to all natural persons without differentiation.'[16]. Such regulation does at least prohibit the use of intelligent agents in specified sensitive circumstances and introduces rules for AI systems designated as high risk.

VI. Conclusion and Future Work

Large digital platforms are extending their influence over increasingly indifferent users who are willing to trade autonomy for convenience. This phenomenon, described here as digital authoritarianism, is enhanced by the widespread deployment of intelligent agents and risks undermining elected state authorities. It also places immense power in the hands of a small number of people. Addressing this democratic imbalance will require decentered dynamic regulation, taking account of the interests of all stakeholders rather than a simple binary mandate. Investigating suitable forms of regulation offers a substantive topic for future cross-disciplinary research.

There is scope also to explore further the constitutional implications touched on in this paper, specifically to assess the robustness of regulatory instruments as a shield for human rights. Equally, the inevitability of digital authoritarianism and the uniformity of user deference posited in this research is likely to promote diverse user practices and resistance which warrant further investigation.

References

[1] H. Arendt, 'The Origins of Totalitarianism', 1966, 3rd ed., Harcourt Brace.

[2] Article 29 Data Working Party: Opinion 8/2014 on Recent Developments on the IoT, 2014

[3] J. Bentham, "Panopticon writings", (M. Bozovic, Ed.). Verso, (Original work published 1787)

[4] A. Bradford, "Europe's Digital Constitution", Virginia Journal of International Law, 2023, Vol.64, Iss. 1, pp. 1-68

[5] D. N. Cicilline, Opening statement at hearing on "Online Platforms and Market Power, Part 6, U.S. House of Representatives Antitrust, Commercial, and Administrative Law Subcommittee. (2020, July 29) https://cicilline.house.gov/press-release/cicilline-opening-statement-at-big-tech-antitrust-hearing (retrieved: June, 2026)

[6] R. Calo, "Robotics and the Lessons of Cyberlaw", W. Barfield and U. Pagallo (eds) "Research Handbook on the Law of Artificial Intelligence", Edward Elgar, Cheltenham, 2018

[7] California Consumer Privacy Act of 2018, Cal. Civ. Code § 1798.100 et seq. (2018)

[8] Charter of Fundamental Rights of the European Union, 2012 O.J. (C 326) 391

[9] A. Chatfield and C. Reddick, "A framework for Internet of Things-enabled smart government", Government Information Quarterly, 2019, Vol. 36, pp. 346–357

[10] J. Cohen, "Between Truth and Power", Oxford University Press, NY, 2019

[11] Convention for the Protection of Human Rights and Fundamental Freedoms (ECHR), 213 UNTS 221 (1950)

[12] D. Danks and A. London, "Algorithmic Bias in Autonomous Systems", Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence (IJCAI-17), 2017

[13] DataReportal, We Are Social, and Meltwater, 2025, Number of connected devices worldwide as of October 2025, Statista. https://www.statista.com/statistics/1559435/connected-devices-worldwide/ (retrieved: March, 2026)

[14] R. Diebert, "Authoritarianism goes Global: Cyberspace under Siege", Journal of Democracy, 2015, Vol. 26, No. 3, pp. 64-78

[15] G. Dyson, "Turing's Cathedral", Penguin Books, London, 2012, p.308

[16] European Parliament & Council of the EU, Regulation (EU) 2024/1689 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act), 2024, OJEU

[17] European Parliament & Council of the EU, Regulation (EU) 2022/1925 on contestable and fair markets in the digital sector (Digital Markets Act), 2022, OJEU

[18] European Parliament & Council of the EU, Regulation (EU) 2022/2065 on a single market for digital services (Digital Services Act), 2022, OJEU

[19] L. Floridi, "The Fight for Digital Sovereignty: What It Is, and Why It Matters, Especially for the EU", Philosophy & Technology, 2020, Vol. 33, pp. 369–378

[20] M. Foucault, "Discipline and Punish: The Birth of the Prison", 1977, (trans. A. Sheridan), Pantheon Books

[21] M. Gal, "Algorithmic Challenges to Autonomous Choice", Michigan Telecommunications and Technology Law Review, 2017, http://dx.doi.org/10.2139/ssrn.2971456 (retrieved: March, 2026)

[22] Granma:http://en.granma.cu/mundo/2019-06-07/digital-totalitarianism; (retrieved: March, 2026)

[23] M. Hildebrandt, "Law as Information", Modern Law Review, 2016, Vol. 79, Issue 1, pp 1-30

[24] House of Lords Communications Committee, "Regulating in Digital World' – oral & written evidence", p.906- 934 theadvertisingindustry (parliament.uk) (retrieved: February, 2026)

[25] N. Howe, "The Rise of Totalitarian Technology", Forbes, 2019, https://www.forbes.com/sites/neilhowe/2019/03/06/the-rise-of-totalitarian-technology/ (retrieved: March, 2026)

[26] IoT Analytics, 2024, Number of connected IoT devices growing 13% to 18.8 billion globally, Statista. https://www.statista.com/statistics/1101442/iot-number-of-connected-devices-worldwide/ (retrieved: March, 2026)

[27] N. Kobie, "The Complicated Truth about China's Social Credit System", Wired, 2019 https://www.wired.co.uk/article/china-social-credit-system-explained (retrieved: March, 2026)

[28] B. J. Koops, "The trouble with EU DP law", International Data Privacy Law, 2014, Vol. 4, No. 4, pp. 250-261

[29] E. Marique and Y. Marique, "Sanctions on Digital Platforms - Beyond the Public/Private Divide", Cambridge International Law Journal, 2019, vol. 8., Issue. 2, pp. 258 - 281

[30] C. Marsden, "Prosumer law and network platform regulation", Georgetown Law Technology Review, Vol. 2, Issue 2 2018 pp. 376-398

[31] A. Murray, "Mapping the Cyber-Regulatory Environment", Routledge, London, 2006

[32] NHS Data Blog, "NHS still reliant on paper notes and drug charts despite electronic upgrades - BMJ Group", (retrieved: March, 2026)

[33] K. O'Hara, K and N. Shadbolt, "The right to be forgotten", European Data Protection Law Review, 2015, Vol. 1, Issue 3, pp. 178-189

[34] G. Orwell, "1984", Secker & Warburg Ltd., London, 1949

[35] F. Pasquale, "From Territorial to Functional Sovereignty", LPE Blog, 2017, From Territorial to Functional Sovereignty: The Case of Amazon - LPE Project (retrieved: March, 2026)

[36] F. Pasquale, “The Black Box Society”, Harvard University Press, 2016

[37] I. Rubinstein, “Big Data: The End of Privacy or a New Beginning?”, New York University Public Law and Legal Theory Working Papers, Paper 357, 2012

[38] P. Sales, “Algorithms, artificial intelligence and the law”, Sir Henry Brooke Lecture for BAILII, 2019, British and Irish Legal Information Institute (BAILII)

[39] H. Schweitzer, J. Haucap, W. Kerber and R. Welker, “Modernising the law on abuse of market power: Report for the Federal Ministry for Economic Affairs and Energy (Germany)”, Federal Ministry for Economic Affairs and Energy (BMWi), 2018

[40] Stanford University, “100 year Study on AI”, https://ai100.stanford.edu (retrieved: March, 2026)

[41] Stigler Committee on Digital Platforms, Policy Brief, Stigler Centre, 2019, p.2

[42] C. Sunstein and R. Thaler, “Libertarian Paternalism is not an Oxymoron”, University of Chicago Public Law & Legal Theory Working Paper No. 43, 2003, Libertarian Paternalism Is Not an Oxymoron (retrieved: March, 2026)

[43] A. Tutt, “An FDA for Algorithms”, 69 Administrative Law Review 83, 2017 pp. 83-123

[44] UK Parliament, Online Safety Act 2023, The Stationery Office, https://www.legislation.gov.uk/ukpga/2023/50, (retrieved: March, 2026)

[45] Universal Declaration of Human Rights, UNGA Res 217 A(III) (1948)

[46] R. Van Loo, “The Corporation as Courthouse”, Yale Journal on Regulation, 2016, Vol. 33, Iss. 2, p. 547

[47] K. Yeung, “Algorithmic Regulation: a critical interrogation”, Regulation and Governance, 2017, Vol. 12, pp. 505–523

[48] K. Yeung, “‘Hypernudge’: Big Data as a mode of regulation by design”, Information, Communication & Society, 2017, Vol. 20, No. 1, pp. 118–136

[49] S. Zuboff, “The Age of Surveillance Capitalism”, Profile Books, London, 2019

Agent-Based Simulation and Digital Twin Modeling for Dynamic Storage Strategies in Dark Warehouses

Simon Holzhauer and Philipp Nenninger
Institute of Energy Efficient Mobility
Karlsruhe University of Applied Sciences
Karlsruhe, Germany
e-mail: {simon.holzhauer | philipp.nenninger}@h-ka.de

***Abstract*—The increasing agility and dynamics of global e-commerce, coupled with steadily rising operational costs and labor shortages, necessitate the transition toward fully automated storage facilities known as dark warehouses. Within these environments, optimizing the Storage Location Assignment Problem (SLAP) is critical to maintaining high-throughput logistics. However, evaluating complex, dynamic storage strategies in active facilities is computationally demanding, requiring robust predictive models rather than static equations. To address this challenge, this paper presents an agent-based discrete-event simulation environment developed as the computational core of an industrial digital twin. Implemented in Python, the model accurately maps the physical kinematics and asymmetric interactions between Automated Guided Vehicles (AGVs) and Automated Storage and Retrieval Systems (AS/RS) within a pallet high-bay warehouse. The simulation systematically evaluates four storage strategies: Random Fixed, Shared Random, Access-frequency-based Fixed Location (FSN), and a Dynamic Predictive Reslotting strategy under varying stochastic load scenarios and sudden seasonal demand shifts. Quantitative analysis reveals a non-linear relationship between hardware resource utilization and systemic performance. The study demonstrates that while static models severely degrade during demand volatility, the dynamic, predictive reslotting algorithm is superior. By leveraging idle time to proactively reorganize inventory, the dynamic strategy minimizes AGV traffic congestion, significantly reduces average retrieval times and enhances overall throughput predictability.**

Keywords-Digital Twins; Agent-Based Simulation; Dark Warehouse; Storage Location Assignment Problem; Intralogistics; Automated Guided Vehicles.

I. INTRODUCTION

The contemporary supply chain is under unprecedented pressure, characterized by the demand for maximum agility and the mitigation of steadily rising costs for storage space, energy, and human capital. This operational pressure is further amplified by a growing global shortage of skilled labor, forcing industrial enterprises to design resilient, highly efficient, and fail-safe automated processes. In this context, *dark warehouses*, which are fully automated storage facilities operating without direct human intervention or environmental conditioning, have emerged as a critical evolutionary step in intralogistics [1][2]. Understanding the economic drivers behind this paradigm shift requires a closer look at the traditional cost structures governing modern warehousing. Evaluating these baseline expenditures helps contextualize why rigid, legacy control frameworks are increasingly being replaced by cyber-physical optimization strategies.

Warehousing and inventory management represent a substantial financial burden within global supply chains, conventionally accounting for roughly 20% to 30% of total logistics costs [3]. Within these facilities, the physical retrieval and order-picking processes are highly labor- and capital-intensive, historically consuming an estimated 55% to 65% of the total warehouse operating budget [4][5]. The optimization of storage and retrieval processes represents a pivotal lever for reducing overall supply chain expenditure. By deploying automated hardware alongside intelligent, data-driven storage strategies, dark warehouses mitigate workforce shortages while simultaneously enhancing throughput velocity, operational flexibility, and order fulfillment precision [6][7].

Evaluating the performance of these complex systems under fully automated conditions requires sophisticated simulative approaches. While industrial software exists for warehouse design, it is frequently proprietary, vendor-locked, and unsuitable for independent, strategic algorithmic evaluation [8][9]. Additionally, scientific literature often isolates specific subproblems of warehousing, failing to capture the systemic interdependencies, race conditions, and spatial congestion inherent in a fully automated facility [4][7].

To address this gap, this investigation focuses on the development of an agent-based simulation environment that acts as the analytical engine for an industrial digital twin [10]–[12]. This framework enables the holistic, quantitative analysis of various storage strategies, spanning static fixed-location models to dynamic, predictive reslotting algorithms. By subjecting these cyber-physical models to varying stochastic demand scenarios, including severe utilization loads and sudden seasonal structure shifts, the research quantifies the limits of hardware scaling versus software orchestration.

The paper is organized as follows: Section II discusses related work surrounding digital twins and simulation methodologies. Section III details the methodology, specifically the storage location assignment problem formulation and the cyber-physical environment architecture. Section IV outlines the mathematical kinematic equations governing the agent movement profiles. Section V presents the algorithmic frameworks for the evaluated storage strategies. Section VI describes the experimental setup and scenario generation. Section VII presents the quantitative results of the simulation. Section VIII provides a comprehensive discussion of these findings. Finally, the conclusion and future work are given in Section X.

 ISBN: 978-1-68558-400-9

II. Related Work in Industrial Twins and Automated Warehousing

The concept of the digital twin has evolved from passive geometric modeling into a predictive, analytical engine essential for complex manufacturing and logistics networks [13]. This evolution is formalized by the five-dimensional digital twin model, comprising the physical entity, virtual entity, services, data, and connections [12]. In dark warehouses, the virtual entity must accurately mirror physical kinematics and spatial constraints; however, the ultimate value lies in the "services" dimension: utilizing the digital replica to run predictive simulations and proactively optimize system parameters before physical execution [12].

The current technological frontier integrates DTs with Agent-Based Simulation (ABS) to create autonomous, self-optimizing entities [14]. Unlike top-down system dynamics or rigid Discrete-Event Simulation (DES), ABS models systems from the bottom up. By defining independent decision-making agents, such as AGVs and AS/RS, ABS captures emergent system dynamics, spatial congestion, and routing conflicts that linear models systematically fail to represent [15][16]. When embedded within a digital twin architecture, ABS acts as the critical bridge between theoretical Warehouse Management System (WMS) optimization and real-time, collision-free execution in automated environments [17][18].

III. Formulating the Storage Location Assignment Problem

The core operational challenge addressed by the simulated digital twin is the Storage Location Assignment Problem (SLAP). SLAP is formally defined as the optimal mapping of a set of products, or Stock Keeping Units (SKUs), to a set of available storage locations to maximize warehousing efficiency [19][20]. Efficiency in this context encompasses the minimization of transport distances, the reduction of retrieval times, and the optimal utilization of spatial capacity.

In its most fundamental form, often referred to as the *Greenfield* scenario where the warehouse is assumed to be entirely empty, SLAP can be mathematically formulated as a classical binary integer programming problem. The theoretical objective function Z aims to minimize the expected retrieval costs based purely on historical access frequencies and spatial distances [21]. This objective function is expressed as:

$$\min Z = \sum_{i \in I} \sum_{j \in J} f_i \cdot d_j \cdot y_{ij} \tag{1}$$

In this formulation, i serves as the index for a product from the set of all products I, while j is the index for a storage location from the set of all locations J. The parameter f_i denotes the access frequency of product i, and d_j represents the physical distance of storage location j to the target input/output (I/O) point. The binary decision variable y_{ij} indicates whether product i is assigned to location j (taking a value of 1) or not (taking a value of 0).

This optimal assignment is subject to strict capacity constraints, ensuring that each product is assigned to exactly one location ($\sum_{j \in J} y_{ij} = 1$) and that each storage location holds no more than one product ($\sum_{i \in I} y_{ij} \leq 1$) [20]. If the minimum of Z is achieved under these conditions, the storage allocation is considered optimal for a static environment.

However, the Greenfield assumption is rarely applicable in continuous dark warehouse operations. A functional warehouse represents a *Brownfield* scenario, containing existing, dynamically fluctuating stock. Any algorithmic attempt to optimize the layout during active operations (reslotting) must mathematically account for the immediate operational penalty of moving an item. Therefore, an extended, practice-oriented objective function must be formulated. This function weighs the long-term retrieval efficiency against the short-term kinematic effort required for relocation:

$$\min Z = \sum_{i \in I} \sum_{j \in J} c_{ij} \cdot x_{ij} + \sum_{i \in I} \sum_{j \in J} u_{ij} \cdot x_{ij} \tag{2}$$

In (2), the binary decision variable x_{ij} dictates whether product i is assigned to or relocated to location j during the active operational phase. Here, c_{ij} represents the expected retrieval cost, analogous to $f_i \cdot d_j$ from the Greenfield model. The critical addition is u_{ij}, which represents the precise cost of physically relocating product i from its current position to the new optimal position j [22].

The inclusion of these relocation costs, combined with the scale of modern distribution centers, transforms the SLAP into a highly complex, NP-hard problem [19]. Because exact mathematical optimization models, such as Mixed-Integer Linear Programming (MILP), suffer from exponential calculation times and are unsuitable for real-time digital twin synchronization, the framework relies on highly efficient heuristic approaches. Specifically, greedy algorithms and metaheuristics are utilized, making locally optimal choices at discrete intervals to approximate a global optimum without violating the real-time computational constraints of the simulation engine [19].

These computational bottlenecks and the inherent spatial complexities of the SLAP can be effectively overcome by embedding this heuristic logic within a dynamic, agent-based simulation. By acting as the virtual entity of the digital twin, this environment evaluates the storage algorithms not in a static mathematical way, but against the emergent traffic and kinematic constraints of a real system.

IV. Architecting the Digital Twin and Cyber-Physical Environment

To ensure that this simulative evaluation maintains empirical validity and industrial applicability, the digital twin must be constructed based on exact physical parameters representative of modern industrial pallet high-bay warehouses. The architecture establishes strict rules regarding spatial topology, capacity, and the interaction interfaces between different classes of autonomous agents.

The reference layout simulates a freestanding high-bay warehouse. The structural geometry consists of 12 distinct racks, each containing 200 compartments arranged in a grid of 20 vertical columns and 10 horizontal levels, yielding a total systemic capacity of 2,400 storage locations. The storage methodology is strictly single-depth. This architectural decision ensures direct, unhindered access to every individual pallet, allowing the simulation to enforce strict first-in-first-out (FIFO) retrieval principles, which are often compromised in double-deep or block storage configurations. The simulation assumes standard European pallets, leading to a parameterized compartment width of 1000 mm, depth of 1500 mm, and a height of 2000 mm to accommodate both the pallet and the stored goods [23].

The spatial dimension of the digital twin is modeled using the Python library NetworkX, which maps the physical layout into a rigid, undirected graph structure as seen in Figure 1[24]. Within this graph, nodes represent physical positions, encompassing intersections, I/O points, and racking interfaces, while weighted edges represent navigable paths based on exact physical distances. This graph serves as the absolute coordinate system for all agent routing.

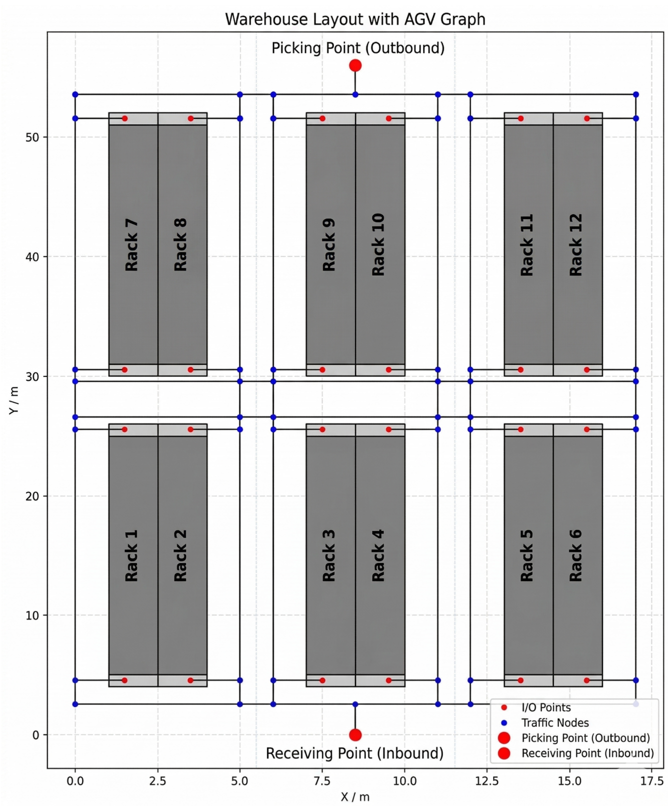

Figure 1. Warehouse Layout with Graph

To simulate the traffic constraints and emergent congestion characteristic of dark warehouses, strict capacity limits are enforced on the graph nodes via the SimPy discrete-event framework. The central goods receipt and goods issue points act as the simulation boundaries and possess high-capacity buffers to accommodate simultaneous staging by multiple agents. The nodes representing the physical handover points between AGVs and SRUs are strictly limited to a capacity of 1, as mechanical constraints dictate that an SRU can only interface with a single AGV at any given moment. Furthermore, intersection and transit routing nodes are limited to a capacity of 2. This prevents deadlocks while enabling fluid bidirectional traffic within the aisles.

Pathfinding within this constrained graph relies on the A^* search algorithm, which guarantees shortest-path optimality across weighted networks [25]. Because the central WMS logic generates thousands of routing requests per hour, the A^* algorithm is optimized via caching mechanisms. Calculated paths between static nodes are retained in memory, drastically reducing computational overhead and allowing the simulation to process extended time horizons rapidly.

V. Kinematic Equations and Agent Movement Profiles

A critical component of a detailed digital twin is the precise mathematical representation of cyber-physical kinematics [26]. The simulation advances beyond static topological routing by calculating the exact temporal duration of every mechanical action, utilizing distinct velocity profiles that account for the acceleration and deceleration phases of AS/RS.

A. AS/RS Kinematic Modeling

An AS/RS operates on a two-dimensional vertical plane within the storage aisles. Because the horizontal drive motor along the X-axis and the hoisting mechanism along the Z-axis operate concurrently and independently, the total travel time T_{travel} is dictated by the slower of the two axes. This relationship is mathematically represented by applying the Chebyshev distance metric to the temporal domain [27]:

$$T_{total} = \max(T_x, T_z) + T_{cycle} \tag{3}$$

In (3), T_{cycle} represents the constant duration for the mechanical fork cycle, encompassing the time required to physically extract or deposit a pallet. The simulation defines this interaction time as 5.0 seconds for pickup and 5.0 seconds for drop-off, yielding a 10-second T_{cycle}.

The calculation for the travel time of an individual axis (e.g., T_x) must differentiate between long and short travel distances. For long distances, the machine reaches its maximum velocity, forming a trapezoidal velocity profile (encompassing acceleration, constant velocity, and deceleration phases). For short distances, the deceleration phase must commence before maximum velocity is achieved, forming a triangular profile. The critical distance d_{crit} required to reach the maximum velocity v_{max} and immediately decelerate to zero, assuming a constant acceleration and deceleration rate a, is calculated as:

$$d_{crit} = \frac{v_{max}^2}{a} \tag{4}$$

Consequently, the precise time T_x required to cover a horizontal distance Δx is computed as a piecewise kinematic function [26]:

Copyright (c) IARIA, 2026. ISBN: 978-1-68558-400-9

$$T_x = \begin{cases} \frac{v_{max,x}}{a_x} + \frac{\Delta x}{v_{max,x}} & \text{if } \Delta x \geq \frac{v_{max,x}^2}{a_x} \text{ (Trapezoidal)} \\ 2 \cdot \sqrt{\frac{\Delta x}{a_x}} & \text{if } \Delta x < \frac{v_{max,x}^2}{a_x} \text{ (Triangular)} \end{cases} \quad (5)$$

The calculation for the vertical axis time T_z follows identical logic, substituting the respective vertical velocity and acceleration parameters. In the simulation, the AS/RS parameters are calibrated to match modern industrial performance standards: the horizontal maximum velocity is configured as $v_{max,x} = 3.0$ m/s with an acceleration of $a_x = 1.0$ m/s^2, while the vertical hoisting velocity is set to $v_{max,z} = 1.5$ m/s with an acceleration of $a_z = 1.0$ m/s^2.

B. Automated Guided Vehicle Kinematics

In contrast to the continuous multi-axis movement of AS/RS, AGVs navigate sequentially along an orthogonal grid on the horizontal plane. Their theoretical travel distances conform strictly to Manhattan distance geometry. To maintain detailed physical representation, the simulation models AGVs with strict kinematic constraints: they must decelerate to a complete stop to execute 90-degree rotations at grid intersections.

The total travel time T_{travel} for an AGV is therefore the sum of all linear transit segments and rotational delays:

$$T_{travel} = \sum_{i=1}^{N} T_{seg,i} + \sum_{j=1}^{M} T_{turn,j} \quad (6)$$

In (6), N represents the total number of straight segments, M denotes the number of direction changes (where $M = N - 1$), $T_{seg,i}$ is the kinematic time required to traverse the i-th segment, and $T_{turn,j}$ applies a constant rotational penalty of 2.0 seconds per 90-degree turn.

The linear traversal time $T_{seg,i}$ is computed using the identical piecewise trapezoidal and triangular kinematic logic established for the AS/RS. To reflect standard industrial safety and performance limits, the baseline AGV kinematic parameters are defined with a maximum velocity of $v_{max} = 1.5$ m/s (equivalent to 5.4 km/h) and a constant acceleration and deceleration rate of $a = 1.0$ m/s^2, which strictly complies with the DIN EN ISO 3691-4 specifications [28].

C. Kinematic Validation

To ensure the integrity of the digital twin, the discrete-event simulation engine was validated against manual kinematic calculations. An edge-case scenario was tested: the retrieval of a pallet located in the compartment closest to the output node (rack 10, compartment 20/1). The manual calculation aggregated the triangular velocity profiles of the short SRU movement, the SRU interaction times, the piecewise AGV segment times across an 8.5-meter multi-segment path, and six total 90-degree AGV rotations. The theoretical manual calculation yielded a total execution time of 54.24 seconds. The Python-based digital twin simulated the identical event in 54.22 seconds, confirming a near-perfect mathematical alignment and validating the engine for large-scale scenario testing.

VI. Storage Strategies and Dynamic Reslotting Logic

The validated simulation serves as the computational testbed for evaluating four distinct storage assignment policies. These strategies advance in complexity from static, rule-based allocations to highly dynamic, predictive algorithms, allowing for a comprehensive benchmarking of intralogistics optimization techniques [19].

The first strategy evaluated is *Dedicated Random Storage*. This baseline strategy mirrors the most rudimentary form of warehouse management. Upon initial system setup, each SKU is assigned a permanent, randomly selected storage compartment. This assignment remains fixed regardless of fluctuating demand or access frequency [20]. While computationally trivial to implement, this approach severely restricts spatial efficiency and leads to suboptimal routing distances, serving primarily as a lower-bound benchmark for measuring subsequent optimizations.

The second strategy is *Shared Random Storage* (often colloquially referred to in industry as chaotic storage). In this approach, no dedicated locations exist. Any incoming pallet is routed to a randomly selected, currently available empty slot anywhere in the warehouse network [29]. While this strategy maximizes spatial density by preventing reserved-but-empty slots, it inherently lacks intelligent geographic positioning. Because retrieval typically follows strict First-In-First-Out (FIFO) rules, the system must locate the oldest instance of an SKU wherever it was randomly placed, heavily penalizing the retrieval times of fast-moving items [4].

The third strategy is *FSN-Based Dedicated Storage*. This methodology introduces statistical categorization based on the velocity of inventory turnover, explicitly dividing SKUs into Fast-moving, Slow-moving, and Non-moving (FSN) tiers [21]. By analyzing historical order frequencies rather than economic value, the WMS partitions the rack topologies. Fast-moving items are statically assigned to the most accessible zones nearest to the I/O points, while non-moving items are relegated to the deepest, most distant racks [29]. Though highly efficient in a perfectly stable environment, the rigid nature of this static zoning creates severe vulnerabilities when demand patterns undergo seasonal or sudden micro-structural shifts, leading to systemic degradation over time.

The fourth, and most advanced, framework implemented within the digital twin is the *Dynamic Reslotting Strategy*. This approach merges intelligent initial slotting with proactive, continuous background reorganization (interleaving) [22]. The WMS operates as an adaptive agent, continuously monitoring SKU access frequencies over a rolling 24-hour historical window. The algorithmic evaluation is triggered at discrete 5-minute intervals. To prevent interference with active fulfillment, the physical reorganization only executes during periods of low systemic load, specifically when AGV utilization drops below 50%.

A greedy heuristic identifies "fast-movers" currently stranded in low-quality, distant locations and "slow-movers" occupying high-value, accessible coordinates near the I/O points. The algorithmic decision to execute a background relocation is governed by a localized profit function G:

$$G = (t_{current} - t_{new}) \cdot f_{sku} \cdot \beta - t_{reloc} \tag{7}$$

This localized profit function represents a novel approach to task interleaving, actively evaluating immediate kinematic penalties against long-term routing efficiency. The economic viability weighting factor, β, was empirically calibrated to 0.70. This threshold ensures that background relocations are only executed when the anticipated long-term retrieval savings significantly outweigh the immediate energy consumption and mechanical wear incurred by the move.

In this profit function, the term $(t_{current} - t_{new})$ represents the expected temporal savings (in seconds) per future retrieval operation if the item is moved to the improved location. This delta is multiplied by f_{sku}, the projected access frequency of the item, and β, a weighting factor for economic viability (calibrated to 0.70). From this projected temporal profit, the immediate kinematic cost t_{reloc}, the exact time incurred by the AGVs and AS/RS to physically perform the move, is subtracted [22].

If G yields a positive value, the digital twin autonomously dispatches an AGV to reorganize the physical environment. To prevent spatial deadlocks and race conditions during these complex product swaps, the WMS utilizes dedicated mutex markers (defined as _SWAP_ states) that temporarily lock specific rack coordinates. This guarantees absolute data consistency and collision-free routing while the AGV acts as a mobile buffer [17].

VII. Experimental Setup and Scenario Generation

The quantitative evaluation of these complex assignment strategies requires robust, highly realistic data inputs. Because proprietary industrial order logs are typically confidential, the digital twin utilizes a stochastic scenario generator parameterized by an open-source, real-world eCommerce dataset from a major footwear manufacturer [30]. Analysis of this reference dataset reveals a pronounced turnover distribution, classic to warehousing logistics: merely 24.0% of the SKUs account for 79.8% of the total order frequency (constituting the high-velocity F-articles), while 23.6% of the SKUs represent 15.1% of the orders (S-articles), and the remaining 52.4% represent a mere 5.1% of the orders (the N-articles) [20][30].

Because the reference dataset originally consists of individual footwear SKUs, order lines were computationally aggregated into homogeneous unit loads (full-pallet equivalents). This transformation ensures the demand profiles accurately reflect the physical and volumetric constraints of a pallet high-bay warehouse.

To model natural logistical clustering, operational peaks, and variance, the simulation relies on an exponential distribution for order inter-arrival times. This ensures the system is subjected to realistic burst demands rather than a perfectly uniform, deterministic stream of tasks [31]. A 30-day continuous simulation horizon is utilized. This extended timeframe is critical to ensure statistical validity and to capture both micro-fluctuations (e.g., daily operational cycles) and macro-trends (e.g., seasonal catalog shifts) without being skewed by initial warm-up periods [31].

To rigorously test the resilience and adaptability of the storage algorithms, the digital twin is subjected to four distinct stress scenarios [22]:

- **Scenario 1 (Low Load):** The base arrival rate is tuned so the system operates at approximately 40% AGV utilization. Resources are abundant, queueing delays are minimal, and the environment represents optimal, non-congested operating conditions.
- **Scenario 2 (High Load):** The order arrival rate is drastically increased to push AGV utilization to a sustained 80%. The system operates near its mechanical breaking point, severely limiting robot idle time and inducing heavy spatial congestion at grid intersections.
- **Scenario 3 (Medium Load with Volatility):** Base utilization is maintained at 60%, but the scenario incorporates aggressive day/night demand shifts and a sudden "seasonal shift" at Day 15. At this juncture, the statistical access probabilities of the F-articles and S-articles are inverted to simulate a rapid product catalog transition.
- **Scenario 4 (High Load with Volatility):** This scenario combines the sustained 80% stress load with the day/night cycles and the Day 15 seasonal inversion. It represents the most constrained operating environment for the WMS, designed to test the limits of algorithmic adaptability under severe physical constraints.

VIII. Simulation Results

The execution of the continuous 30-day simulation experiments across the four scenarios generated extensive telemetry datasets, allowing for a deep quantitative analysis of both hardware resource dimensioning and strategic software efficacy.

Prior to comparing the storage strategies, the digital twin was utilized to determine the optimal AGV fleet size. Holding the baseline arrival rate constant, the fleet was scaled from 1 to 30 units. The analysis revealed a distinct non-linear relationship. At fleet sizes below 7 units, the AGVs represented a critical systemic bottleneck, with wait times growing exponentially. As the fleet expanded to 10 units, retrieval times plummeted and stabilized. Increasing the fleet from 10 to 30 AGVs yielded statistically negligible improvements in throughput. Therefore, a fleet of 10 AGVs was selected for all scenario testing, providing a baseline utilization of 60–70% that balances efficiency with realistic capital expenditure (CAPEX) constraints. Conversely, the telemetry data from the 12 AS/RS units revealed chronic underutilization across all scenarios, indicating an asymmetric hardware bottleneck within the reference design.

Table I presents the results for Scenario 1, operating under low load conditions.

 ISBN: 978-1-68558-400-9

TABLE I. SCENARIO 1: LOW LOAD CONDITIONS

Strategy	Storage [s]	Storage Δ[%]	Retrieval [s]	Retrieval Δ[%]	SD [s]
Random	80.48	0.00	89.00	0.00	19.6
Shared	78.86	-2.01	88.30	-0.79	19.8
FSN Fixed	84.99	+5.60	81.70	-8.20	16.4
Dynamic	76.02	-5.54	75.60	-15.06	15.6

Under these ideal conditions, the Dynamic Reslotting strategy significantly outperformed, reducing average retrieval times by 15.06% compared to the baseline (from 89.00 s to 75.6 s). Furthermore, it reduced operational variance, dropping the standard deviation to 15.6 seconds. The FSN strategy performed adequately in retrieval (-8.20%) but suffered during storage (+5.60%), as inbound AGVs frequently bypassed empty slots to reach reserved, deeper zones [29].

Table II presents the results for Scenario 2, operating under sustained high load.

TABLE II. SCENARIO 2: SUSTAINED HIGH LOAD

Strategy	Storage [s]	Storage Δ[%]	Retrieval [s]	Retrieval Δ[%]	SD [s]
Random	86.32	0.00	97.60	0.00	23.7
Shared	83.35	-3.44	94.20	-3.48	22.3
FSN Fixed	93.91	+8.79	94.40	-3.28	21.3
Dynamic	95.15	+10.23	88.70	-9.12	20.2

The high load fundamentally altered the efficiency matrix. The Dynamic strategy maintained its lead in retrieval (-9.12%), but resulted in a corresponding 10.23% increase in storage times (from 86.32 s to 95.15 s). Because the AGVs were constantly occupied with active order fulfillment, the WMS successfully suppressed background reslotting algorithms, temporarily preventing perfect layout optimization to prioritize immediate throughput [22]. Simultaneously, the FSN strategy experienced significant degradation. By statically clustering all fast-movers in the front racks, the FSN strategy created intense spatial hotspots, causing multiple AGVs to queue and gridlock at the same AS/RS transfer nodes [17].

Table III details the system's performance in Scenario 3, incorporating demand volatility and a seasonal shift under medium load.

TABLE III. SCENARIO 3: MEDIUM LOAD WITH VOLATILITY

Strategy	Storage [s]	Storage Δ[%]	Retrieval [s]	Retrieval Δ[%]	SD [s]
Random	80.07	0.00	90.47	0.00	20.4
Shared	78.96	-1.39	89.41	-1.17	20.5
FSN Fixed	85.20	+6.41	83.10	-8.25	18.6
Dynamic	80.20	+0.16	77.62	-14.21	18.1

In Scenario 3, the resilience of the Dynamic slotting strategy became evident. Despite the inversion of product demand profiles at Day 15, the algorithm utilized the night cycle lulls to completely restructure the physical layout. By swapping obsolete stock with newly trending SKUs, it maintained a 14.21% retrieval advantage over the baseline (from 90.47 s to 77.62 s).

Table IV presents the results for Scenario 4, combining high load, high volatility, and a seasonal shift.

TABLE IV. SCENARIO 4: HIGH LOAD WITH VOLATILITY

Strategy	Storage [s]	Storage Δ[%]	Retrieval [s]	Retrieval Δ[%]	SD [s]
Random	83.53	0.00	94.34	0.00	22.4
Shared	81.79	-2.08	93.12	-1.30	22.0
FSN Fixed	89.23	+6.83	89.61	-4.99	21.9
Dynamic	90.04	+7.80	82.65	-12.39	19.9

In the most demanding scenario, the systemic breakdown of static optimization was highly pronounced. A temporal analysis of the FSN strategy revealed that immediately following the Day 15 shift, retrieval times spiked drastically and remained elevated. The fixed assignments became severely suboptimal, with obsolete products blocking premium slots while newly trending products were stranded deep in the warehouse network [21]. In contrast, the Dynamic slotting strategy detected this statistical drift and continuously reallocated inventory during the brief night shifts to recover efficiency, ultimately achieving a sustained 12.39% retrieval advantage even under maximum load (from 94.34 s to 82.65 s). This advantage in scenario 4 is also shown in Figure 2, which compares the retrieval time of the different strategies.

IX. DISCUSSION

The vast telemetry datasets generated by the agent-based simulation yield several higher-order insights with profound implications for the design and deployment of industrial digital twins.

The most critical revelation is the inevitable decay of static optimization models. The failure of the FSN strategy in Scenarios 3 and 4 highlights a primary limitation of static facility design: static allocation models inevitably degrade over time due to stochastic shifts in consumer demand [20]. An industrial digital twin must not merely serve as a passive monitoring dashboard; it must feature continuous, autonomous realignment mechanisms [13]. The superiority of the Dynamic Reslotting strategy proves that optimization is not a singular project to be completed at facility commissioning, but a continuous computational cycle. The digital twin must act as a closed-loop control system, detecting statistical anomalies and autonomously dispatching agents to rebalance the logistical topology before overall throughput is compromised [12].

The success of the dynamic reslotting algorithm relies heavily on the concept of spatiotemporal arbitrage. The algorithm actively trades excess systemic capacity in one time period, such as the lower-demand night shifts, for enhanced kinetic performance during peak daytime operations. During complex

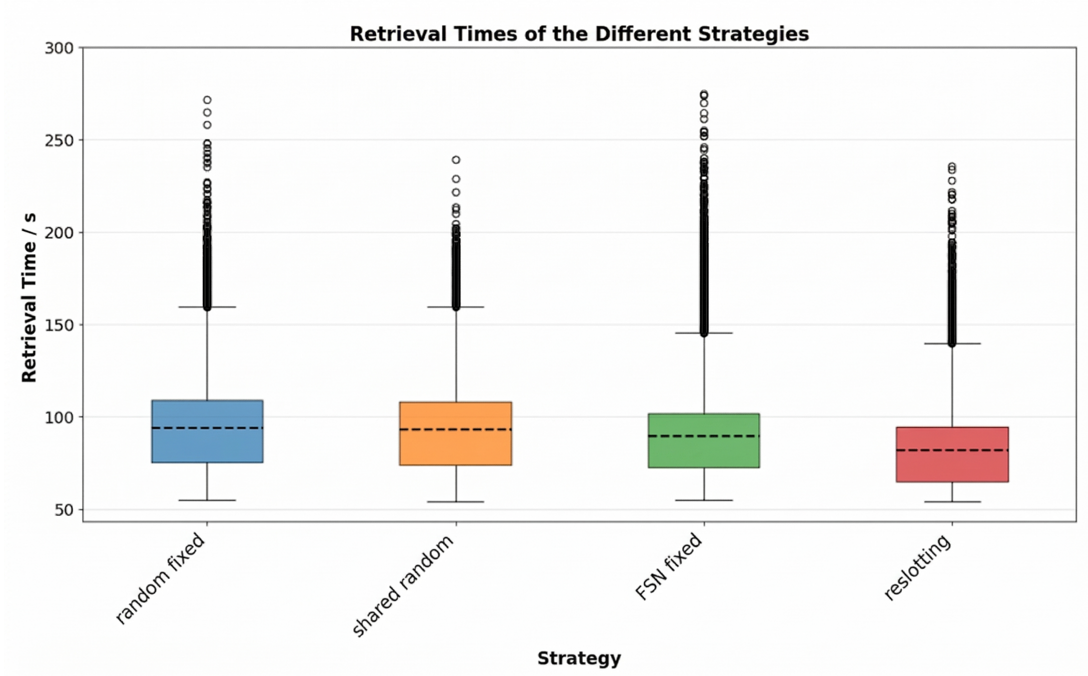

Figure 2. Scenario 4 - Retrieval Times of the Strategies

"swap" operations, the WMS utilizes the AGV's payload capacity as a temporary mobile buffer while the AS/RS maneuvers pallets into new positions [22]. This algorithmic behavior effectively integrates transport and storage capacities. In a sophisticated digital twin, AGVs are no longer viewed simply as delivery vectors moving from point A to point B, but as dynamic, floating nodes within the active memory architecture of the warehouse itself [1].

The observed increase in storage times under the Dynamic strategy is a calculated operational trade-off. By proactively reserving highly accessible slots near the I/O points for fast-moving retrievals, inbound storage tasks are frequently relegated to deeper racks. However, because outbound order fulfillment is typically the critical bottleneck in e-commerce intralogistics, sacrificing storage velocity to guarantee retrieval velocity yields a net-positive systemic throughput.

This leads to a counterintuitive insight regarding the total cumulative distance traveled by the AGV fleet. One might logically assume that the background relocations executed by the dynamic slotting strategy would vastly inflate the total mileage. However, our primary data from Scenario 3 reveals that the total fleet distance for the baseline random strategy was 4,647 km, while the dynamic slotting strategy, including all background relocation drives, totaled only 4,337 km. The optimization of the spatial layout compensates for the added mechanical effort of the reorganization itself. By proactively bringing the most frequently accessed goods closer to the I/O nodes, the marginal travel savings of every subsequent commercial pick accumulate into a net reduction in overall AGV travel distance.

Finally, the simulation exposes the strict limits of physical hardware scaling. The saturation curve of the AGV fleet, demonstrating diminishing returns beyond 10 units, proves that physical scaling cannot unilaterally solve logical bottlenecks. Introducing additional hardware into a highly utilized, constrained spatial graph simply generates exponential traffic conflicts, thereby lowering the marginal utility of each robot [17]. Maximum warehouse throughput is fundamentally bound by software orchestration. The finding that the 12 AS/RS units were chronically underutilized while AGVs were saturated indicates that future dark warehouse facility designs need to evaluate the number of AS/RS units as well.

X. Conclusion and Future Work

The transition to fully automated dark warehouses represents a massive capital investment for the logistics sector. To maximize the return on this infrastructure, enterprises must move beyond rigid control systems and embrace adaptive, AI-driven operations. This paper demonstrated the efficacy of an agent-based simulation acting as the core of an industrial digital twin, quantitatively proving that dynamic, predictive reslotting significantly outperformed static models. The dynamic strategy achieved the shortest retrieval times, reducing them by up to 15.06% and drastically collapsed operational variance, transforming temporal downtime into kinetic efficiency and enhancing overall throughput predictability. Most critically, it demonstrated the unique capacity to autonomously adapt to unforeseen seasonal demand shifts, leveraging off-peak hours to rebalance the physical layout.

Based on the insights gained from this validated architecture, future work must expand the capabilities of the digital twin. To thoroughly validate the proposed greedy heuristic, future investigations will benchmark it against exact Mixed-Integer Linear Programming (MILP) solvers on smaller-scale instances to quantify the optimality gap. Additionally, the evaluation framework will be expanded to financially quantify the added energy consumption, mechanical wear, and labor costs associated with continuous reslotting. Finally, to transition this architecture toward real-world deployment under highly congested multi-agent conditions, future models will evaluate the integration of Deep Reinforcement Learning (DRL) algorithms to replace the heuristic, allowing the warehouse to completely and autonomously self-govern its spatial topology.

ACKNOWLEDGMENTS

This work was developed as part of the move.mORe project. The move.mORe project funded by the Innovation University (IHS) promotes research-based regional transfer of ideas, knowledge and technology in the field of sustainable mobility and is jointly run by the universities of Karlsruhe and Offenburg

REFERENCES

[1] G. Fragapane, R. de Koster, F. Sgarbossa, and J. O. Strandhagen, "Planning and control of autonomous mobile robots for intralogistics: Literature review and research agenda," *European Journal of Operational Research*, vol. 294, no. 2, pp. 405–426, 2021.

[2] S. Winkelhaus and E. H. Grosse, "Logistics 4.0: A systematic review towards a new logistics system," *International Journal of Production Research*, vol. 58, no. 1, pp. 18–43, 2020.

[3] P. Baker and Z. Halim, "An exploration of warehouse automation implementations: Cost, service and flexibility issues," *Supply Chain Management: An International Journal*, vol. 12, no. 2, pp. 129–138, 2007.

[4] R. De Koster, T. Le-Duc, and K. J. Roodbergen, "Design and control of warehouse order picking: A literature review," *European Journal of Operational Research*, vol. 182, no. 2, pp. 481–501, 2007.

[5] J. A. Tompkins, J. A. White, Y. A. Bozer, and J. M. A. Tanchoco, *Facilities planning*, 4th ed. Hoboken, NJ: John Wiley & Sons, 2010.

[6] N. Boysen, R. de Koster, and F. Weidinger, "Warehousing in the e-commerce era: A survey," *European Journal of Operational Research*, vol. 277, no. 2, pp. 396–411, 2019.

[7] K. Azadeh, R. De Koster, and D. Roy, "Robotized and automated warehouse systems: Review and recent developments," *Transportation Science*, vol. 53, no. 4, pp. 917–945, 2019.

[8] S. Weyer, M. Schmitt, M. Ohmer, and D. Gorecky, "Towards industry 4.0-standardization as the crucial challenge for highly modular, multi-vendor production systems," *IFAC-PapersOnLine*, vol. 48, no. 3, pp. 579–584, 2015.

[9] E. Negri, L. Fumagalli, and M. Macchi, "A review of the roles of digital twin in cps-based production systems," *Procedia Manufacturing*, vol. 11, pp. 939–948, 2017.

[10] G. Lugaresi and M. Macchi, "Digital manufacturing modelling: A review of digital twin applications in machining and intralogistics," *IFAC-PapersOnLine*, vol. 51, no. 11, pp. 1143–1148, 2018.

[11] B. Rodič, "Industry 4.0 and the new simulation modelling paradigm," *Organizacija*, vol. 50, no. 3, pp. 193–207, 2017.

[12] F. Tao et al., "Digital twin-driven product design framework," *International Journal of Production Research*, vol. 57, no. 12, pp. 3935–3953, 2019.

[13] W. Kritzinger, M. Karner, G. Trampusch, and M. Sihn, "Digital twin in manufacturing: A categorical literature review and classification," *IFAC-PapersOnLine*, vol. 51, no. 11, pp. 1016–1022, 2018.

[14] S. Wang, J. Wan, D. Li, and C. Zhang, "Towards smart factory for industry 4.0: A self-organized multi-agent system with big data based feedback and coordination," *Computer Networks*, vol. 101, pp. 158–168, 2016.

[15] C. M. Macal and M. J. North, "Tutorial on agent-based modelling and simulation," *Journal of Simulation*, vol. 4, no. 3, pp. 151–162, 2010.

[16] P. O. Siebers, C. M. Macal, J. Garnett, D. Buxton, and M. Pidd, "Discrete-event simulation is dead, long live agent-based simulation!" *Journal of Simulation*, vol. 4, no. 3, pp. 204–210, 2010.

[17] Y. A. Bozer and D. A. Aldarondo, "A simulation-based comparison of routing algorithms for multi-load agvs," *European Journal of Operational Research*, vol. 268, no. 2, pp. 500–516, 2018.

[18] N. Kousi et al., "A digital twin for adapting robot behavior in flexible intralogistics," *Procedia Manufacturing*, vol. 28, pp. 171–176, 2019.

[19] J. J. Reyes, E. L. Solano-Charris, and J. R. Montoya-Torres, "The storage location assignment problem: A literature review," *International Journal of Industrial Engineering Computations*, vol. 10, no. 2, pp. 199–224, 2019.

[20] J. J. Bartholdi and S. T. Hackman, *Warehouse & Distribution Science*, Release 0.98. Supply Chain and Logistics Institute, Georgia Institute of Technology, 2019.

[21] W. H. Hausman, L. B. Schwarz, and S. C. Graves, "Optimal storage assignment in automatic warehousing systems," *Management Science*, vol. 22, no. 6, pp. 629–638, 1976.

[22] J. A. Pazour and H. J. Carlo, "Reslotting strategy for a unit-load automated storage and retrieval system," *IIE Transactions*, vol. 47, no. 8, pp. 823–839, 2015.

[23] *EN 13698-1: Pallet production specification - Part 1: Construction specification for 800 mm x 1200 mm flat wooden pallets*, European Committee for Standardization (CEN), Brussels, Belgium, 2003.

[24] NetworkX, "Networkx documentation," 2025, Accessed: 2026-05-27. [Online]. Available: https://networkx.org/

[25] P. E. Hart, N. J. Nilsson, and B. Raphael, "A formal basis for the heuristic determination of minimum cost paths," *IEEE Transactions on Systems Science and Cybernetics*, vol. 4, no. 2, pp. 100–107, 1968.

[26] H. Hwang, W. Baek, and M.-K. Lee, "Travel-time models considering the operating characteristics of the storage and retrieval machine," *International Journal of Production Research*, vol. 42, no. 14, pp. 2877–2889, 2004.

[27] Y. A. Bozer and J. A. White, "Travel-time models for automated storage/retrieval systems," *IIE Transactions*, vol. 16, no. 4, pp. 329–338, 1984.

[28] ISO, *Industrial trucks — safety requirements and verification — part 4: Driverless industrial trucks and their systems*, Standard DIN EN ISO 3691-4, International Organization for Standardization, Geneva, CH, 2020.

[29] C. G. Petersen and G. Aase, "An evaluation of routing and volume-based storage policies in an order picking operation," *Decision Sciences*, vol. 35, no. 2, pp. 269–289, 2004.

[30] R. Furlan de Assis, *Order picking dataset from a warehouse of a footwear manufacturing company*, version 1, Mendeley Data, 2024. Accessed: 2026-05-27. [Online]. Available: https://data.mendeley.com/datasets/pf2w725pw3/1

[31] A. M. Law, *Simulation Modeling and Analysis*, 5th ed. New York: McGraw-Hill Education, 2015.

AI Agent-Driven Insurance Platform

A Multi-Agent Architecture for Intelligent, Deterministic, and Auditable Insurance Processing

Vidya More
Independent Researcher
USA
e-mail: vidyamore@yahoo.com

James Zhang
Independent Researcher
USA
email: james_s_zhang@yahoo.com

***Abstract*—The insurance industry faces mounting pressure to reduce claims leakage, accelerate cycle times, enhance underwriting and customer experience, and maintain regulatory compliance, all while operating on legacy platforms that were not designed for *modern* automation capabilities. This paper presents a comprehensive architectural framework for an Artificial Intelligence (AI) Agent-Driven Insurance Platform that addresses these challenges through a novel approach: replacing monolithic, hard-coded workflow engines with a collaborative ecosystem of specialized AI agents governed by an intelligent orchestrator. The proposed architecture introduces three transformative capabilities: (1) business-owned policy management through natural language, allowing business analysts to express rule changes in plain English that are *automatically* transformed into typed, versioned Domain-Specific Language (DSL) with comprehensive testing, conflict detection, and governed deployment; (2) agentic orchestration where specialized agents each responsible for discrete functions such as intake, document understanding, classification, coverage verification, fraud detection, adjudication, payment processing, and communications collaborate under unified governance rather than competing within rigid workflow queues; and (3) explainable and auditable outcomes through deterministic execution, where every decision carries explicit rationale, policy citations, version identifiers, and immutable audit trails. The architecture enables insurers to materially reduce cost-per-claim and cycle time while improving compliance, explainability, and customer outcomes, positioning them for competitive advantage in an increasingly digital marketplace.**

Keywords*—*AI agents; insurance automation; multi-agent systems; claims processing; explainable AI

I. INTRODUCTION

The AI Agent-Driven Insurance Platform addresses these challenges by replacing brittle, hard-coded workflows with a collaborative set of specialized AI agents guided by an Orchestrator. Each agent focuses on a specific responsibility, intake, document understanding, classification, coverage, fraud detection, adjudication, payments, and communications, enabling the overall system to behave like a coordinated team rather than a monolith.

In this white paper, we primarily focus on the Claim system as an example for this new AI Agent-Driven Insurance Platform.

A. Fundamental Transformations

From code-bound rules to business-owned policy. Business Analysts express changes in plain English, which are transformed into a typed, versioned DSL with unit tests, conflict checks, and approvals. Production evaluates these compiled policies deterministically, ensuring repeatable outcomes and auditability.

From linear queues to agentic orchestration. The Orchestrator delegates work to domain-specific agents, selects appropriate tools (OCR, rules runtime, pricing APIs, vector search), and enforces guardrails including confidence thresholds, human-in-the-loop tiers, and compliance checks.

From opaque decisions to explainable outcomes. Every decision carry rationale, citations, policy IDs and versions, and a durable audit trail, enabling rapid review, coaching, and regulatory response.

B. End-to-End Value Chain Coverage

Intake and Document Understanding: Multi-channel First Notice of Loss (FNOL) with automated extraction from invoices, forms, photos, and reports with quality and confidence scoring.

Classification and Triage: Claim type and severity determination with routing informed by historical patterns and predictive analytics.

Coverage and Benefits: Deterministic verification of eligibility, limits, deductibles, and endorsements/exclusions with explicit citations to policy text.

Fraud and Risk Assessment: Rules combined with machine learning signals including recency analysis, entity link analysis, and external data such as police reports and motor vehicle records.

Adjudication: Tool first decisioning leveraging fee schedules, contracted pricing, and policy constraints, with agents producing decisions, rationale, and confidence scores.

Payments and Recoveries: Net payment calculation incorporating reserves and deductibles, accounts payable and general ledger posting, and subrogation setup.

Communications: Clear explanations of benefits, notices, and status updates rendered from templates, localized, and archived.

C. Safety and Reliability Guarantees

Deterministic core: Pricing and coverage decisions are executed by a Rules Runtime compiling approved policies; large language models orchestrate and draft explanations but do not generate financial outcomes.

Governed change: All policy edits are versioned, test-covered, and deployed via continuous integration and deployment pipelines with canary rollout and instant rollback capabilities.

Assured oversight: Confidence thresholds route exceptions to human-in-the-loop queues. Every interaction prompts, tool calls, and outputs is immutably logged.

D. Business Impact

Speed to change: New coverage interpretations, reserve limits, and fraud thresholds move from ideation to production in hours without developer cycles.

Operational efficiency: Straight-through processing rises for low-complexity claims while adjusters focus on high-severity and nuanced cases.

Leakage and risk reduction: Early detection of duplicates and anomalies reduces overpayment; transparent reasoning lowers denial overturns.

Enhanced experience: Claimants receive timely, plain-language updates with consistent determinations across channels and adjusters.

E. Technology Outlook

The platform integrates with existing policy administration, accounts payable and general ledger, repair and provider networks, and data sources through an API gateway and event bus. It adopts cloud-native components for scale and resilience, and a vector search layer for grounded citations. The modular agent pattern allows incremental adoption—starting with FNOL and document AI, then expanding to coverage, adjudication, and payments as key performance indicators meet thresholds. Over time, the agentic layer becomes the effective claims core, with legacy systems acting as tools and data providers.

This same architectural blueprint is equally applicable to Policy Administration and Billing workflows, where it can automate underwriting support, policy issuance, premium calculation, invoice reconciliation, and payment execution—seamlessly connecting with existing PAS, billing engines, and payment gateways. By applying agent-based orchestration to these domains, insurers gain consistent, auditable automation across the insurance value chain—reducing manual intervention, minimizing errors, and accelerating end-to-end policy lifecycle and financial operations.

This architecture unifies automation, safety, and business agility. By aligning natural-language intent (what the business wants) with deterministic execution (what production does), insurers can materially reduce cost-per-claim and cycle time while improving compliance, explainability, and customer outcomes—across claims, underwriting, policy service, and finance.

II. CURRENT MAINSTREAM CLAIMS SYSTEMS

Mainstream claims administration platforms used across property and casualty (P&C), health, life, and specialty insurance are typically built on rules-driven workflow engines, deterministic processing logic, and enterprise integrations with policy, billing, document management, and financial systems. These platforms were designed to enforce consistency, regulatory compliance, and operational efficiency, but most trace their lineage to legacy architectures that emphasize sequential, queue-based processing over dynamic automation or adaptive decision.

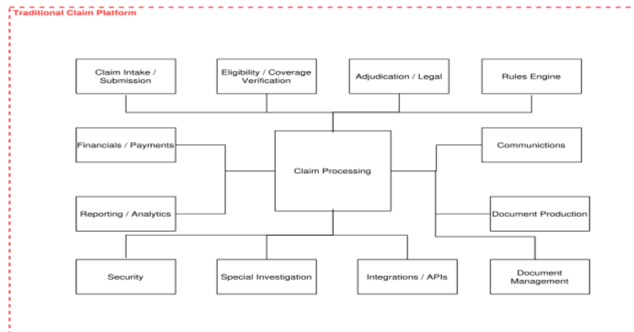

Fig. 1. Traditional Claims System Architecture Overview.

At their core, these systems rely on the following foundational pillars:

A. Workflow Queues and Case Management

Most traditional claims systems organize work using linear workflow queues that route tasks to adjusters or examiners based on claim type, geography, severity, or workload. These queues enforce a structured flow—FNOL, investigation, coverage determination, adjudication, payment authorization—but are often static and inflexible, making it difficult to reconfigure processes or support real-time, context-aware routing.

Characteristics of workflow-centric systems: Human-assigned tasks following rigid workflows; limited dynamic rerouting based on confidence, risk, or completeness; heavy reliance on manual triage and adjuster intervention; limited ability to leverage historical context or emerging patterns.

This creates bottlenecks for even simple claims, which today could be automated with AI-led processing.

B. Rules Engines for Business Logic and Adjudication

Traditional platforms embed business logic inside fixed, code-based rules engines. These engines power coverage verification, benefit calculations, reserve setting, medical coding, repair pricing, and eligibility checks. While deterministic and audit-friendly, they present significant limitations: hard to modify—rule updates often require IT intervention, regression testing, and deployment cycles; siloed from context—rules operate only on structured data provided by the claim system, without the ability to ingest or reason over unstructured documents, narrative text, or external signals; static—difficult to adapt to evolving fraud patterns, new products, or jurisdictional changes.

 ISBN: 978-1-68558-400-9

This rigidity slows business teams and increases time-to-market for new policies or operational improvements.

C. Enterprise System Integration

Policy Administration Systems (PAS): Provide essential data such as coverage limits, deductibles and coinsurance, effective dates and endorsements, and named insured and additional party information. However, PAS integrations are frequently batch or cache-based, leading to latency and mismatch with real-time decisioning needs.

Financial Systems (General Ledger): Used for payment authorization, reserve management, reinsurance recoveries, and subrogation and salvage accounting. These integrations typically rely on SOAP/REST APIs or file exchanges, which often create manual reconciliation overhead.

Document Management Systems (DMS): Support storing PDFs, forms, photos, police reports, and medical records with indexing via metadata and version control. Yet traditional DMS platforms cannot understand the content of documents, requiring adjusters to manually read and interpret them.

D. Core Functional Modules

Although implementations vary by line of business and vendor, mainstream systems usually include these functional components:

Claim Intake and Registration: Multi-channel FNOL ingestion (phone, email, portal), manual validation of required information, initial claim creation and assignment.

Eligibility and Coverage Verification: Lookup of policy terms, limits, and endorsements; date-based eligibility checks; exclusion detection using static rules.

Adjudication via Business Rules: Fee schedule calculations, coding validations (CPT, ICD, repair codes), duplicate claim checks, benefit calculations or repair estimates.

Adjuster Workbenches: Case notes, task lists, and work queues; manual investigation workflows; diary tasks and follow-up reminders.

Payments, Recoveries, and Financials: Reserve creation and adjustments, payment approval workflows, subrogation and salvage processing, ledger updates and reconciliation.

Communications: Letters, explanations of benefits, emails, and notices using templates—often manually triggered with limited personalization.

Fraud and Compliance: Basic rule-based red flags, checks against watchlists or known fraud indicators, regulatory validation and documentation.

Reporting and Analytics: Operational dashboards, compliance reporting, and basic historical insights.

E. Limitations of Mainstream Architecture

While mainstream platforms have been foundational for decades, they are not optimized for the dynamic, high-velocity environment insurers face today: manual-heavy processes persist even for simple claims (glass, towing, small losses); limited automation due to dependence on structured data and static rules; slow to change—business rule updates require developer involvement, long QA cycles, and deployment windows; lack of semantic understanding—systems cannot reason over documents, images, or contextual information; siloed decision logic across multiple systems without a unified reasoning engine; inconsistent customer experience because outcomes vary by adjuster expertise.

These limitations underscore the need for a new, flexible, intelligent claims architecture—one built on collaborative AI agents capable of understanding context, automating routine work, and ensuring decisions remain safe, deterministic, and governed.

III. PAIN POINTS IN EXISTING SYSTEMS

Despite decades of incremental improvements, most mainstream claims administration systems still reflect the architectural constraints of the eras in which they were built—rigid workflows, static rules engines, and siloed integrations. As a result, carriers face persistent operational inefficiencies, limited adaptability, and rising cost pressures.

A. Hard-Coded Rules and Developer Dependency

In traditional claims platforms, critical decision logic, such as coverage conditions, reserves, benefit calculations, fraud flags, communication templates, is embedded directly in hard-coded rules managed by IT teams. Even seemingly simple changes, such as adjusting a deductible threshold or updating a reserve formula, would cause requirements translation from business to IT, coding and/or rules-authoring cycles, regression testing across interconnected rulesets, and deployment through formal release cycles.

These steps create weeks-to-months delays for even minor business updates. As a result, business agility suffers as product teams cannot quickly respond to regulatory changes, competitive pressure, or emergent risk patterns. Operational rigidity increases as adjusters must compensate manually for outdated system logic, introducing inconsistent decisions and leakage. Innovation slows because new ideas—such as micro-rules for fraud detection or contextual routing—cannot be tested rapidly.

B. Siloed Workflows and Brittle Integrations

Most legacy claim systems rely on point-to-point integrations and linear workflow queues, which were adequate when claims were handled sequentially by adjusters. Today, these architectures struggle under the complexity and speed required for real-time automation.

Common issues include fragmented data flows between claim admin, policy admin, billing, and DMS systems; duplicated logic across lines of business leading to unnecessary repetitive efforts; manual handoffs between departments due to limited cross-system context; and high maintenance overhead for each new integration.

When workflows are siloed, the system cannot dynamically adapt to risk indicators, claim complexity, or missing information because every task is pre-determined by rigid workflow maps rather than context-aware orchestration.

C. Extensive Manual Handling and Low Straight-Through Processing

Mainstream systems depend heavily on adjuster intervention, even for low-complexity claims such as auto-glass repairs, towing reimbursements, simple property damage, and outpatient medical claims. This occurs because information extracted from documents must be read manually, coverage rules require manual interpretation, suspicious patterns lack automated scoring, and workflows cannot progress without human validation of each step.

Low straight-through processing causes high cost per claim, longer cycle times and poor customer satisfaction, increased rework and leakage due to inconsistent decisions, and adjuster burnout with higher turnover. Carriers are left with high overhead for work that, in today's AI-capable environment, could be fully automated.

D. Limited Explainability and Auditability

Traditional systems provide deterministic outputs but often lack transparent reasoning. Adjusters and auditors see what decision was made, but not clearly why. Challenges include no granular visibility into which rule fired, in what order, or based on what evidence; no direct citations of support documents, contract language, or historical cases; difficulty reconstructing decision pathways during disputes or regulatory audits; and inconsistent rationales across adjusters and channels.

This limited explainability reduces trust in system outputs, increases overturn rates on appeals, and introduces regulatory risk—particularly in jurisdictions demanding transparent, consumer-friendly decisioning.

E. Inflexibility in Piloting New Approaches

Fraud patterns, regulatory expectations, customer communications standards, and market variables change frequently. Traditional claim systems struggle to adapt because integrating new fraud indicators requires model deployment plus rules updates, modifying pricing or coverage logic requires expensive IT cycles, A/B testing is nearly impossible in rigid workflow structures, and legacy data architectures cannot easily incorporate new signals such as telematics, unstructured text, or automated image-based damage analysis.

This leads to slow reaction to fraud trends, increasing leakage, stale pricing or reserve logic worsening loss ratio, inconsistent claimant experience especially across digital and human channels, and missed opportunities to test and optimize operational strategies quickly.

These pain points: rigid rules, siloed workflows, manual reliance, poor explainability, and limited agility, form the core motivation for moving to an AI agent-driven claims platform where business owns logic rather than IT; agents collaborate dynamically rather than follow static workflows; decisions are transparent, explainable, and grounded in policy language, automation covers the majority of low-complexity claims, and new fraud patterns, routing rules, and communication patterns can be piloted instantly.

IV. AI AGENT-DRIVEN FUTURE STATE ARCHITECTURE

The AI Agent-Driven future state represents a fundamental shift from traditional rules-bound, workflow-centric insurance platforms toward a dynamic, context-aware, and continuously improving intelligence layer. Rather than a monolithic system encoded with thousands of brittle rules, the next-generation platform is composed of specialized, collaborative AI agents, each responsible for a narrow domain of expertise, working together under the guidance of an Orchestrator Agent that governs task delegation, tool usage, and confidence thresholds.

This new architecture enables insurers to automate the majority of low-complexity insurance tasks, scale human expertise through intelligent assistance, and continuously keep business logic aligned with regulatory requirements and strategic objectives.

The following diagram illustrates an enterprise-grade, multi-layered architecture centered around an AI agent orchestrator. Designed to be scalable, secure, and human-aligned, it enables flexible instruction ingestion, intelligent orchestration, autonomous execution, and crucially, maintains robust governance and control.

1. Ingestion Layer: Instructions enter the system via diverse modalities: Web & Mobile User Interfaces, API Gateway, Interactive Voice Response (IVR) systems, Agent-native natural language prompt interfaces, Chat interfaces, email auto-ingestion, and more. All inputs are normalized into a standardized intermediate representation for downstream processing.

2. Orchestration Layer: The orchestrator serves as the "brain" of the system responsible for reasoning, planning, and coordination. Key components include the Instruction Analyzer, Workflow Generator, Memory Store, Tools Registry, and Task Distributor.

Critical Safeguard: All generated workflow plans undergo mandatory review by the Workflow Plan Validator, which verifies compliance with business rules & regulatory constraints, logical consistency and safety, alignment with ethical guidelines. Only approved workflows proceed to execution.

3. Agent Layer: A heterogeneous ecosystem of specialized AI agents, each optimized for specific domains or tasks (e.g., data extraction, decision support, multi-step problem solving). executes validated workflow steps. Agents operate as autonomous yet coordinated units, leveraging LLMs, rule-based engines, retrieval-augmented generation (RAG), and symbolic reasoning modules.

4. Human-AI Collaboration & Governance Layer: A continuous quality loop ensures transparency, accountability, and control embedded across all stages, including Execution Monitor, Circuit Breaker, Escalation Engine, and Human-in-the-Loop (HITL).

5. Empowerment Layer: A key innovation: no-code/natural-language workflow co-creation. Business users define rules, templates, and exceptions in plain language and they are automatically transformed into validated artifacts by the system. This democratizes automation while preserving governance.

Unlike legacy RPA or rigid BPM systems, this architecture is adaptive, explainable, and continuously learning—blending AI agility with enterprise-grade reliability.

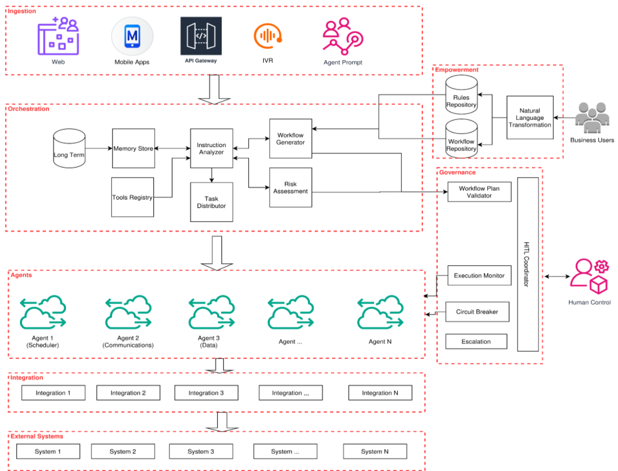

Fig. 2. Enterprise-Grade Multi-Layered Architecture.

A. From Monolithic Logic to Multi-Agent Collaboration

In mainstream systems, business logic is centralized and rigid—often buried in rules engines, code, or workflow definitions. The AI agent-driven model decentralizes functionality into narrow, accountable agents: Intake Agent, Document/OCR Agent, Classification and Triage Agent, Coverage and Benefits Agent, Fraud and Risk Agent, Adjudication Agent, Payment Agent, and Communications Agent.

Each agent has clear authority, strict tool access, and well-defined responsibilities, enabling modular improvement without destabilizing the entire system.

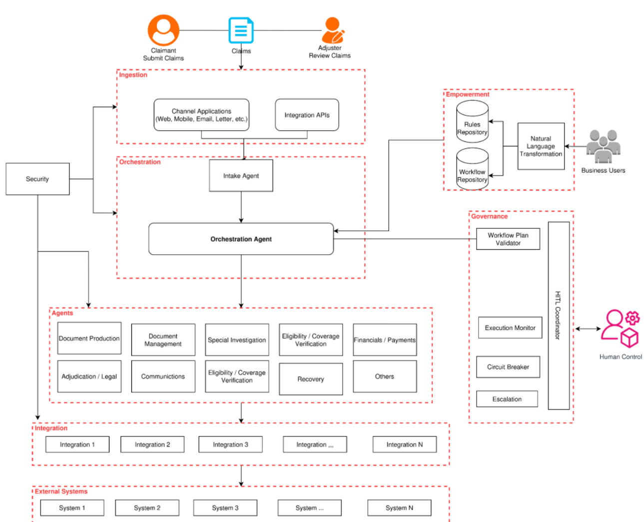

Fig. 3. Multi-Agent Collaboration Architecture for Claim Processing.

Agents do not compete or duplicate work. Instead, the Orchestrator Agent coordinates their interactions by determining which agent should handle which part of a claim, invoking tools (OCR, rules engine, pricing APIs, vector search), enforcing confidence thresholds, deciding when to escalate to a human, and ensuring all decisions remain explainable and grounded in evidence.

B. Business-Owned Rules Using Natural Language to DSL Transformation

Traditionally, business rule maintenance requires developers and long release cycles. The future state eliminates this dependency through a policy/rules-as-content approach.

How It Works:

Step 1: Business Analysts write rules in plain English. Example: "For auto-glass claims in NY with no injuries, set reserves to $300; if deductible > $300, use deductible."

Step 2: The Rule/Policy Authoring Agent parses the natural-language rule, generates structured machine-runnable DSL (YAML/JSON), creates unit tests to validate expected behavior, performs conflict analysis against existing rules, runs simulation/backtesting against historical claims, and provides an impact assessment (reserve changes, STP shifts).

Step 3: Governance workflows enforce human approval, versioning, separation of duties, canary deployment, and rollback capability.

Step 4: The approved DSL is compiled and deployed to the Rules Runtime.

Step 5: Agents call the runtime at execution time to apply the new policy automatically.

Sample DSL (YAML Format):

```
meta:   id: reserve_auto_glass_ny
effective_from: 2026-02-10 when:   all:
- claim.type == "AUTO_GLASS"    -
claim.jurisdiction == "NY"    -
claim.injuries == false then:   - if:
policy.deductible > 350     do:
set_reserve(policy.deductible)   - else:
do: set_reserve(350)
```

This fundamentally changes how insurers evolve their business logic—turning claims operations into a rapidly configurable, business-owned system with zero developer involvement in day-to-day policy changes, rule updates that go live in hours versus months, full traceability from business intent to tests to production behavior, and consistent governance that reduces operational and compliance risk.

C. Deterministic, Safe Decisioning Through Compiled Rules Runtime

A central requirement for insurance operations is determinism: decisions must be repeatable, defensible, and compliant. In the future-state architecture, the Rules Runtime is the deterministic backbone of the platform.

Key characteristics include all business logic running through compiled DSL rules rather than free-form LLM generation. The runtime evaluates coverage eligibility, deductible/limit calculations, reserve formulas, pricing constraints, fraud thresholds, and compliance checks. Every execution produces decision values, rule IDs and versions, explanation metadata, and audit-ready logs.

Agents never "guess" at financial outcomes. They orchestrate actions, compose evidence, and generate rationales—but the Rules Runtime produces final, safe, and consistent determinations. This design ensures regulatory compliance, consistency across adjusters and channels, reproducibility for audits and disputes, and predictable operational performance.

D. Runtime Safety, Guardrails, and Human-in-the-Loop Assurance

Every agent action adheres to a strict set of controls:

Confidence thresholds: Low-confidence agent outputs route to human adjusters.

Criticality tiers: Higher-risk decisions (bodily injury, potential fraud) require human review.

Evidence binding: All decisions must cite supporting documents, policy text, or historical precedent.

Safety filters: PHI/PII redaction, prompt injection protection, content safety, and regulatory constraints.

Immutable audit logs: Prompts, tool calls, decision artifacts, agent chain-of-thought summaries (sanitized), and policy versions used.

This ensures the system is safe, governable, and aligned with regulatory expectations, even as automation increases.

E. Continuous Optimization Through Learning Loops

Unlike static legacy platforms, the agent-driven system improves over time. Agents capture telemetry on errors, escalations, adjuster overrides, fraud false positives/negatives, and time-to-decision. These signals feed evaluation datasets for refined policies, updated models, improved routing logic, better fraud thresholds, and enhanced communications.

Because the DSL is versioned and test-backed, improvements can be safely deployed without destabilizing production.

F. Strategic Benefits of the Future State

Operational Efficiency: Higher straight-through processing, less manual data entry, fewer rework loops, and faster cycle times.

Financial Outcomes: Reduced leakage, better reserve adequacy, improved fraud detection, and higher payment accuracy.

Customer and Adjuster Experience: More consistent outcomes, clearer explanations, faster responses, and reduced adjuster burnout.

Enterprise Agility: Rapid policy changes, faster regulatory compliance, easier experimentation and A/B testing, and modular expansion of agent capabilities.

Together, these benefits create an insurance platform that is intelligent, explainable, adaptable, and safe—positioning carriers to compete in a fast-moving digital market.

V. COMPREHENSIVE BENEFITS ANALYSIS

The AI Agent-Driven Architecture introduces substantial improvements over traditional claims systems by combining automation, deterministic rule execution, explainability, and business-owned rules control. The result is a platform that accelerates operational performance, reduces loss costs, improves compliance, and provides a unified, intelligent claims experience.

A. Speed: Rapid Business Logic Changes

Traditional claims platforms depend on developer-coded rules, often requiring multi-week release cycles for simple updates. The agent-driven system replaces this with a policy/rules-as-content operating model in which Business Analysts author changes directly in plain English without involving IT, the Policy Authoring Agent converts natural-language instructions into structured DSL complete with tests, conflict checks, and simulations, approved changes are deployed via CI/CD pipelines enabling updates in hours rather than weeks, and canary rollouts allow changes to be validated safely in production on a small subset of claims.

This dramatically increases business agility carriers can respond quickly to regulatory updates, new coverage products, supply-chain constraints, fraud trends, and pricing adjustments.

B. Automation: Multi-Agent Orchestration Drives Higher STP

Using claim processing as an example, low-complexity claims (auto glass, towing, low-severity property, minor medical) still consume a disproportionate amount of adjuster effort because legacy systems lack contextual automation. The multi-agent system introduces specialized agents such as Intake, OCR, Classification, Coverage, Fraud, Adjudication, Payment, and Communication, that collaborate to automate most of the claim lifecycle.

Documents are automatically extracted and normalized. Coverage eligibility and deductible calculations execute deterministically. Fraud-scoring models and rules are invoked in real time. Adjudication uses contracted pricing or fee schedule logic. Payments and communications are automatically prepared. This increases straight-through processing (STP) dramatically, allowing adjusters to focus on high-severity or disputed claims rather than routine administrative work.

C. Determinism and Safety: Reliable, Compliant Outputs

A hallmark of insurance operations is the need for repeatable, defensible, and audit-ready decisions. Freeform LLM responses alone cannot guarantee this. The AI agent-driven system ensures safety by compiling all approved business policies into a deterministic Rules Runtime, ensuring agents never guess at financial outcomes, restricting agents to orchestrate tools rather than generating outcomes directly, and capturing every execution with complete traceability inputs, rules fired, tool calls, decisions, and rationale.

D. Explainability: Clear Rationale with Citations

Traditional claims systems often lack transparency into why a particular decision was made. The AI agent-driven approach introduces built-in explainability through automatic generation of decision rationales, inclusion of policy/rule IDs and version numbers used in the Rules Runtime, direct citations to policy wording and endorsement text via vector-search grounding and supporting documents (police report sections, invoice line items, medical records), and immutable audit trails storing every step of agent reasoning, tool usage, and final decision outputs.

This reduces dispute frequency, improves regulatory defensibility, and supports fairer, more consistent outcomes for customers.

E. Scalability: Event-Driven, Serverless Architecture

Legacy claims systems struggle under high volume because workflows and rules engines scale poorly. The agent-driven platform is inherently elastic with event-driven pipelines that decouple system components to avoid bottlenecks, serverless or autoscaling computing that expands extraction and reasoning capacity instantly, a Rules Runtime that can process thousands of parallel evaluations per second, and each agent being independently deployable and horizontally scalable.

This enables carriers to handle catastrophic events (storms, wildfires, hail events) without operational backlogs or degraded performance.

F. Compliance: Built-In Controls

Insurance operations require strict adherence to regulatory frameworks (Health Insurance Portability and Accountability Act (HIPAA), System and Organization Controls 2 (SOC2), General Data Protection Regulation (GDPR), and state Department of Insurance (DOI) regulations). The platform incorporates compliance as a first-class architectural principle with: PHI/PII detection and redaction via built-in content filtering; Role-Based Access Control (RBAC) ensuring agents and users only access permitted data; audit logging of every agent invocation, model call, rule evaluation, and tool interaction; versioning with full lineage enabling reconstruction of historical decisions; data residency controls and encryption in transit and at rest; and HITL thresholds for high-risk decisions [5].

VI. HIGH-LEVEL TECHNICAL ARCHITECTURE

The AI Agent-Driven Claims Platform is structured into three cohesive architectural layers, Authoring and Governance, Runtime, and the Agent Layer. Together, these layers create a modular, governable, and highly automated ecosystem that replaces rigid legacy workflows with adaptive intelligence while ensuring full compliance, determinism, and auditability.

A. Authoring and Governance Layer

This layer defines how business intent becomes production logic—safely, quickly, and with complete traceability. It empowers business teams to modify rules, test them, and deploy them without developer intervention, while maintaining strict change-control and regulatory compliance.

Business Console (User Interface): A natural-language interface where Business Analysts author or modify claim rules, fraud checks, reserve logic, communication rules, or workflow policies using plain English. Examples include "Increase initial reserve for NY auto-glass claims to $300" or "Flag claims with 2+ prior glass losses in 12 months for manual review."

Policy/Rule Authoring Agent: Transforms natural-language policies into a structured, machine-runnable Domain Specific Language (DSL). This agent parses, validates, and interprets business instructions; generates typed DSL rules; builds unit tests automatically; runs conflict detection against existing policies; simulates policy impact using historical claims; and produces human-readable change summaries for review.

Policy/Rules Store: A versioned repository (Git/Dataverse) that maintains DSL rule files, unit tests, metadata (owners, effective dates, change rationale), approval records, and rollback history. This acts as the single source of truth for all operational business logic.

Governance Service: Applies controls required for compliance and change management: RBAC and separation of duties, multi-step approvals (BA → Manager → Compliance), policy versioning and lineage tracking, canary and progressive rollouts, and rollback triggers with audit logs.

Simulator and CI/CD Pipeline: Before deployment, each policy is validated through back testing on historical claims, forecasted operational impact (STP changes, reserve shifts, fraud detection sensitivity), automated policy unit tests, and integration checks. After validation, CI/CD pipelines compile DSL rules and push them into the production runtime with full auditability.

B. Runtime Layer

The Runtime Layer is the deterministic execution engine of the platform. While agents orchestrate processes and reasoning steps, all final decisions are produced here, ensuring consistency, compliance, and repeatability.

Rules Runtime (Deterministic Execution): The core engine that executes compiled DSL rules for coverage verification, deductible and limit calculations, reserve formulas, pricing constraints, fraud rule triggers, and compliance validations. Agents never "guess" outcomes this runtime guarantees predictable and consistent decisions.

Vector Search (RAG Knowledge Layer): A retrieval system used by agents (not the Rules Runtime) to ground explanations and reasoning in authoritative sources such as policy wordings, endorsement texts, repair guidelines, medical policy bulletins, and historical decision snippets.

Document Store: Stores all claim artifacts: images (damage photos), invoices and repair estimates, medical

 ISBN: 978-1-68558-400-9

records, police reports, and email and communication history. Agents (such as the OCR Agent) retrieve items from here to extract structured data.

Observability: Provides complete visibility into the system: telemetry and performance metrics, policy hit-rates, fraud rule activation rates, agent confidence scores, escalation pathways, and full prompt/tool logs for audits.

Event Bus: An asynchronous messaging backbone enabling decoupled agent orchestration, high-volume parallel processing, real-time streaming of events (FNOL received, OCR completed, rule fired, fraud flagged), and backpressure protection during catastrophes.

C. Agent Layer

Agents are specialized, domain-specific LLM-based modules that perform reasoning, call tools, compose decisions, and orchestrate claim processes under the guidance of the Orchestrator.

Orchestrator Agent: The central controller responsible for delegating tasks to the appropriate agents, evaluating confidence thresholds, deciding HITL escalation, ensuring all actions follow policy constraints, enforcing tool-use safety rules, and composing final decisions and rationales. This is the "brain" that coordinates the multi-agent ecosystem.

Specialized Agents: Intake Agent collects FNOL and validates required fields, requesting missing information dynamically. Document/OCR Agent extracts structured data from images, PDFs, forms, and emails using OCR/IDP. Classification and Triage Agent determines claim type, severity, routing path, and urgency based on extracted data and historical patterns. Coverage and Benefits Agent checks eligibility against policy terms, verifies limits and exclusions, and calculates deductible/coinsurance logic using the Rules Runtime. Fraud/Risk Agent runs fraud rules and ML scoring models, retrieves external risk indicators (MVR, police data), and determines whether HITL review is necessary. Adjudication Agent evaluates pricing, applies constraints, verifies coding rules, and produces a structured decision proposal with rationale and citations. Financial Agent calculates payment amounts, updates reserves, posts financial transactions, and initiates subrogation. Communications Agent drafts EOBs, letters, and status updates, ensures regulatory formatting, localizes content, and archives communications.

VII. REPRESENTATIVE USE CASES

The AI Agent-Driven Claims Platform transforms the end-to-end claims lifecycle across claim types and complexity tiers. The following representative scenarios illustrate how the architecture performs in both straight-through and human-in-the-loop contexts, how business rules are authored and deployed, and how proactive fraud detection is embedded as part of routine operations.

A. Auto Glass Claim: Straight-Through Processing

Use Case Type: High-volume, low-complexity claim

Goal: Near-instant decisioning; maximize straight-through processing (STP)

Auto glass claims are ideal candidates for end-to-end automation due to predictable repair paths, well-understood coverage rules, and stable pricing structures. The platform executes the following steps:

FNOL Intake: The Intake Agent collects claim details (vehicle info, loss cause, location, date/time) via mobile, portal, or API. Missing fields are automatically requested through conversational prompts.

Document and Invoice Extraction: The Document/OCR Agent processes uploaded images or vendor invoices, extracting line items such as labor hours, glass type, moldings, and calibration charges.

Claim Classification: The Classification Agent identifies the claim as "Auto-Glass / Low-Severity" and predicts likely STP eligibility using historical patterns.

Coverage Determination: The Coverage and Benefits Agent queries the policy administration system, retrieves deductibles and endorsements, and applies deterministic rules (e.g., "zero glass deductible" rules in certain states).

Fraud Check: The Fraud/Risk Agent runs anomaly detection, recency checks, duplicate claim detection, and MVR signals. No risk flags means the claim remains in the STP path.

Adjudication: The Adjudication Agent validates pricing against contracted network rates or fee tables, applies deductible math, and computes the net payable amount.

Payment and Communications: The Payment Agent posts the transaction to AP/GL; the Communications Agent generates an Explanation of Benefits with citations to policy text and sends it to the claimant.

Outcome: Claim resolved in minutes rather than days, high STP (often >70-90% of glass claims), and adjusters handle only exceptions rather than routine cases.

B. Bodily Injury Claim: HITL Tier 2 Escalation

Use Case Type: Complex, high-exposure claim

Goal: Human-in-the-loop (HITL) augmentation, not replacement. Bodily injury claims involve medical evidence, legal implications, and higher fraud exposure. The platform amplifies adjusters' capabilities by assembling evidence and surfacing intelligent insights.

Data Aggregation: The Document Agent extracts structured data from medical reports, treatment plans, billing statements, and police narratives.

Policy Context: Coverage Agent gathers applicable liability limits, medical payments coverage, and exclusions relevant to injury scenarios.

Fraud and Risk Analysis: The Fraud/Risk Agent identifies inconsistencies such as treatment gaps, disproportionate billing patterns, or medical provider anomalies.

Orchestrated Escalation: Due to exposure, uncertainty, or elevated fraud signals, the Orchestrator Agent routes the claim to a specialist adjuster.

Compiled Evidence Package: The system produces an evidence brief containing extracted medical facts and timelines, policy citations and limits, similar historical claim patterns, reasoning steps from coverage, fraud, and classification agents, and suggested next actions.

Specialist Decisioning: The adjuster receives a complete, transparent dossier, enabling faster, more informed evaluation and negotiation.

Outcome: Specialists focus only on high-value insights, faster cycle times for complex claims, and improved accuracy with reduced leakage and stronger SIU collaboration.

VIII. BUSINESS IMPACT AND ROI: A CLEAR PATH TO VALUE

The adoption of an AI-Agent driven claims platform is not just a technology upgrade; it is a strategic investment that delivers a clear and compelling return on investment (ROI). By automating manual tasks, improving decision accuracy, and accelerating cycle times, our solution drives measurable improvements across the claims value chain, backed by robust industry data.

TABLE I. INDUSTRY BENCHMARKS

KPI Category	Metric	Target Improvement	Industry Benchmark
Operational Efficiency	Claims cycle time (end-to-end)	40-60% reduction	McKinsey cites up to 70% cycle time reduction [13]
	Cost per claim processed	25-35% reduction	McKinsey and BCG cite 20-30% cost reduction [14] [15]
Decision Quality	Decision accuracy	95%+ accuracy	Can reach 99.9% transaction accuracy [12]
Risk & Fraud	Fraud detection hit rate	2-3x improvement	65% improvement in detection capabilities [12]
	Fraud detection hit rate	2-3x improvement	65% improvement in detection capabilities [12]
Customer Experience	NPS / CSAT	15–25 point improvement	GenAI adopters report 45% higher NPS [12]

Illustrative ROI Projection: For a mid-size insurer processing 500,000 claims per year at an average processing cost of $150 per claim, a 30% cost reduction translates to $22.5 million in annual savings. This is against a typical implementation investment of $1.5 million to $3 million, delivering a payback period of less than 12 months, aligning with industry benchmarks that suggest a 12–18 months payback window is common for high-volume claim automation [16].

A. Cost Reduction Drivers

Increased Straight-Through Processing: Moving from typical industry STP rates of 10-30% to 70-90% for low-complexity claims dramatically reduces manual handling costs.

Reduced Cycle Time: Faster claims resolution reduces administrative overhead and improves customer satisfaction, reducing churn.

Leakage Prevention: Improved fraud detection and consistent adjudication reduce overpayments and missed recoveries.

IT Cost Reduction: Business-owned rule changes eliminate developer dependency for routine policy updates.

B. Revenue and Retention Impact

Beyond direct cost savings, the platform drives revenue through improved customer retention from faster, more transparent claims experiences; competitive differentiation enabling premium pricing; and capacity for growth without proportional headcount increases.

C. Implementation Considerations

Organizations should consider phased implementation starting with high-volume, low-complexity claim types, establishing clear KPI baselines before implementation, building internal capabilities for ongoing optimization, and planning for change management across claims operations.

IX. CONCLUSION

The shift from traditional insurance platforms to an AI agent-driven architecture represents a fundamental evolution in how insurance organizations operate. By unifying deterministic business logic with intelligent, orchestrated agents, carriers can achieve levels of speed, consistency, and transparency that are unattainable with legacy workflow- and rules-based systems.

Modernization begins by separating business intent from system behavior. Business teams express policies in plain English, while the platform transforms that intent into a structured, versioned DSL that executes deterministically through the Rules Runtime. This single architectural decision eliminates the long-standing dependency on developers for rule changes, empowering operations to adapt rapidly to market forces, regulatory shifts, and fraud trends. As a result, changes that once took weeks or months can now be deployed safely within hours.

An AI agent approach also compresses the claim lifecycle, automating low-complexity flows end-to-end and augmenting human expertise for complex cases. Orchestrated agents handle intake, document understanding, classification, coverage evaluation, fraud screening, adjudication, and communication—each specializing in its domain but working together through a governed, explainable workflow. This reduces manual touchpoints, shortens cycle times, and lowers operational cost per claim.

At the same time, the platform strengthens governance and reduces leakage by enforcing deterministic execution, robust audit trails, and policy/rule versioning. Every decision is backed by rule IDs, policy citations, and documented rationale, providing unmatched transparency for regulators, auditors, and customers. The architecture embeds privacy and security controls such as PHI/PII redaction, RBAC, and immutable logging, ensuring compliance as automation increases.

Ultimately, an AI agent-driven platform does more than modernize insurance processing—it redefines it. It enables insurers to operate with the speed of a digital-first organization, the discipline of a regulated financial institution, and the intelligence of a continuously learning system. By aligning business intent with safe, deterministic execution, insurers achieve improved customer outcomes, reduced leakage, rapid adaptability, and a scalable foundation for the next decade of claims innovation.

X. FUTURE RESEARCH DIRECTIONS

While this white paper presents a comprehensive architecture for AI agent-driven insurance processing, several areas warrant further research and development:

A. Advanced Agent Capabilities

Future work should explore more sophisticated reasoning capabilities including multi-modal understanding that combines text, images, audio, and video for comprehensive claim assessment; predictive analytics for proactive claims management and loss prevention; and natural language generation improvements for customer-facing communications.

B. Cross-Industry Applications

The multi-agent architecture described here has potential applications beyond insurance, including healthcare, financial services compliance, legal document review, and supply chain dispute resolution, just to name a few.

C. Regulatory Framework Evolution

As AI adoption in insurance claims increases, research is needed on evolving regulatory requirements for AI transparency, standardization of explainability requirements across jurisdictions, and frameworks for AI governance in regulated industries.

D. Future State Architecture Illustartive

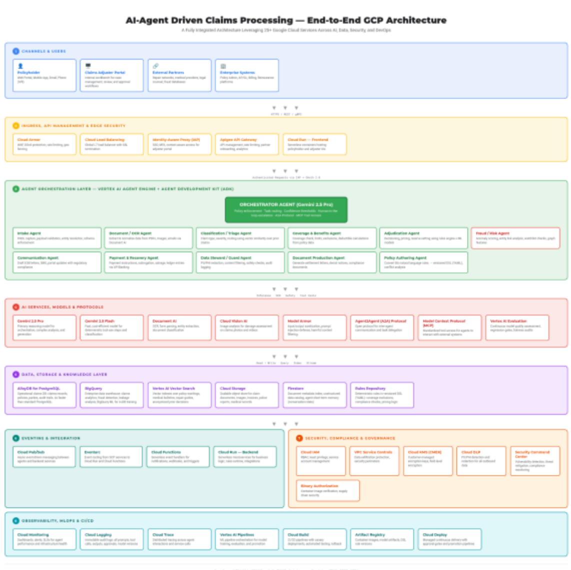

Fig. 4. *GCP* Based Architecture Variant – Illustrative

REFERENCES

[1] Internal analysis and projections based on industry benchmarks, 2025. Unpublished

[2] Coalition Against Insurance Fraud, "The Impact of Insurance Fraud," Coalition Against Insurance Fraud, Washington, DC, 2025.

[3] Verisk Analytics, "The State of Auto Insurance Claims," Verisk Analytics, Jersey City, NJ, 2025.

[4] PwC, "Insurance Consumer Intelligence Series," PricewaterhouseCoopers LLP, 2025.

[5] BCG, "Agentic AI Can Power Core Insurance IT Modernization," Boston Consulting Group, Jan. 12, 2026. [

[6] McKinsey & Company, "The Future of AI for the Insurance Industry," McKinsey & Company, Jul. 15, 2025. [Online].

[7] DRUID AI, "Automated Insurance Claims: How AI Is Transforming the Industry," DRUID AI, Jan. 28, 2026.

[8] Gartner, as cited in "The True State of AI Agents in the Insurance Industry," LinkedIn, Apr. 21, 2025.

[9] Google Cloud, "MediConCen Case Study," Google LLC, 2024.

[10] Google Cloud, "AlloyDB for PostgreSQL Product Page," Google LLC, 2024.

[11] Agentech, "How Is AI Transforming the Insurance Claims Process?," Agentech, Oct. 23, 2025.

[12] Datagrid, "42 AI Agent Statistics for Insurance (Adoption and Impact)," Datagrid, Nov. 6, 2025.

[13] D. R. Kumar, "Automation in Claims Processing: Reducing Turnaround Time," in Proc. Int. Conf. Business and Technology Innovation, Jun. 2025.

[14] McKinsey & Company, as cited in VCA Software, "Automated Claims Processing: The Definitive Guide for Insurers," VCA Software, Oct. 17, 2025.

[15] BCG, "To Win with AI, Insurers Must Go Beyond the Algorithm," Boston Consulting Group, Oct. 28, 2025. [Online].

[16] VCA Software, "Automated Claims Processing: The Definitive Guide for Insurers," VCA Software, Oct. 17, 2025.

[17] J. Zhang, "Treat AI as Human (TAAH): A Governance Framework for Autonomous Agent Systems Based on Institutional Wisdom from Human Societies," Medium, 2026. Unpublished

Copyright (c) IARIA, 2026. ISBN: 978-1-68558-400-9

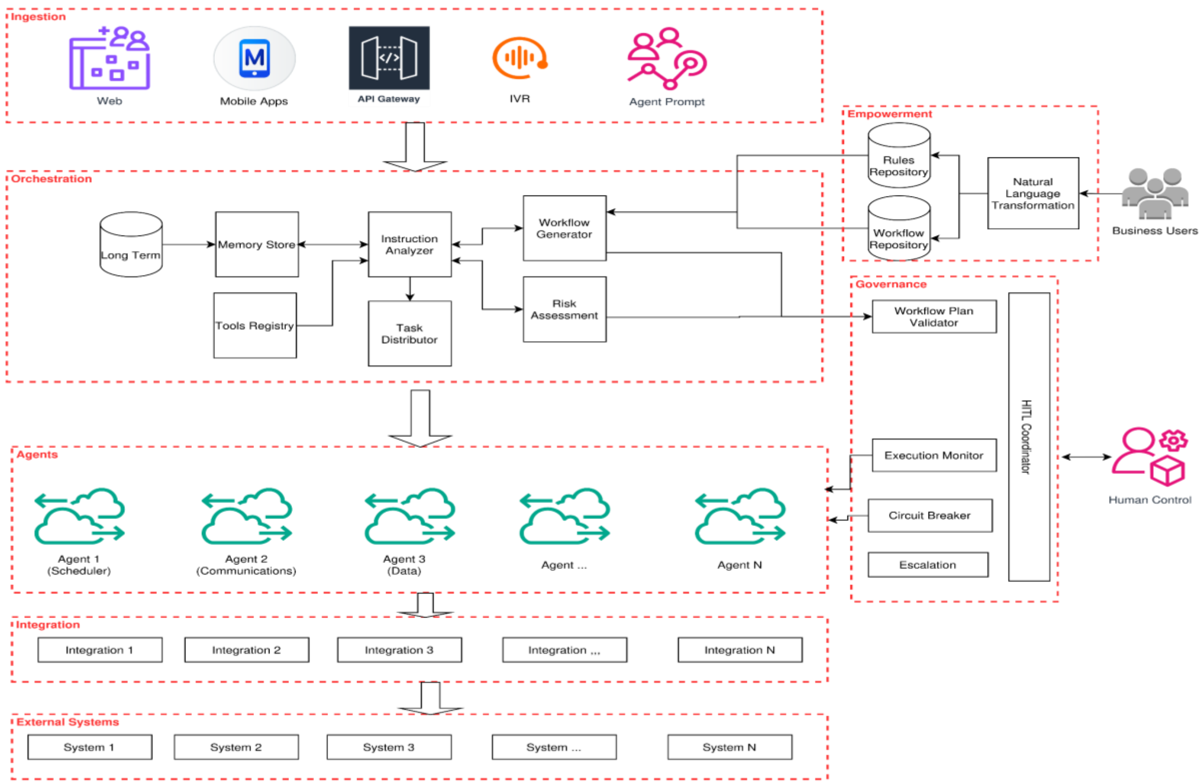

Figure 2. Enterprise-Grade Multi-Layered Architecture.

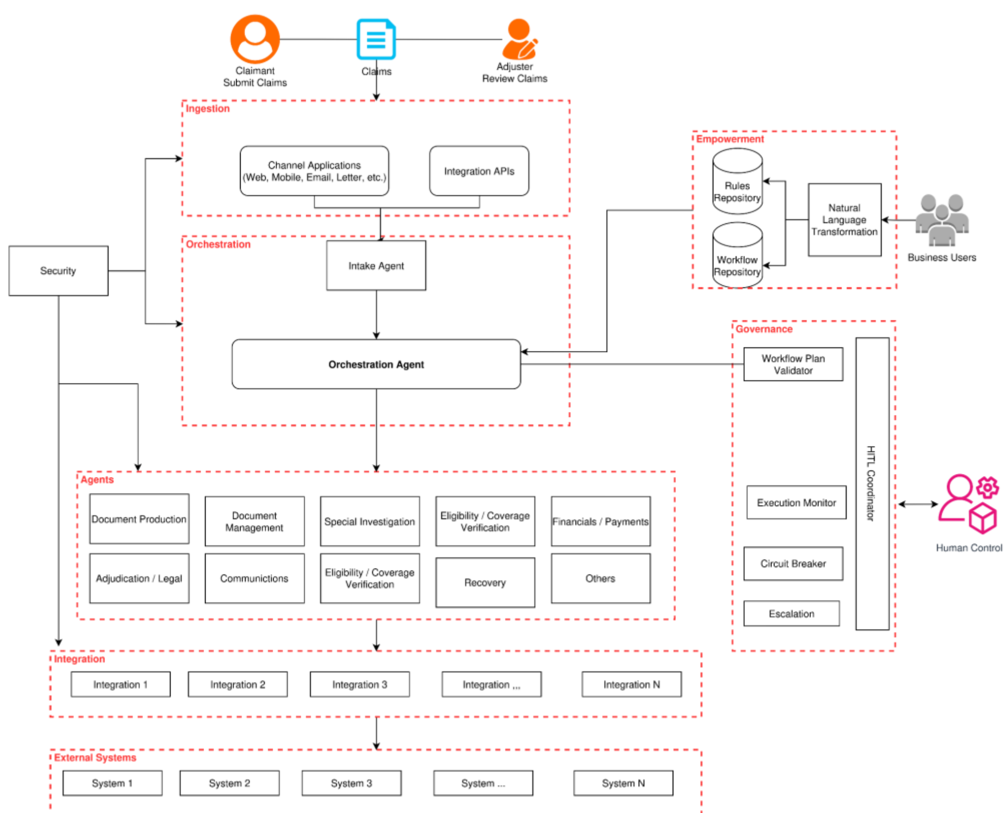

Figure 3. Multi-Agent Collaboration Architecture for Claim Processing.

Using AI-Generated Fake News to Teach Source Evaluation to Norwegian Upper Secondary School Students

Andrea Marie Sæter
Department of Microsystems
University of South-Eastern Norway
Borre, Norway
e-mail: andreamseter@gmail.com

Jesper Edvin Kraft
Department of Microsystems
University of South-Eastern Norway
Borre, Norway
e-mail: kraftjesper@gmail.com

Lasse Berntzen
School of Business
University of South-Eastern Norway
Hønefoss, Norway
e-mail: lasse.berntzen@usn.no

***Abstract*—This study investigates how an AI-based learning platform can support upper secondary school students in distinguishing between authentic and fake news. A social media-inspired platform was developed in which students interacted with both real and AI-generated news articles within a realistic feed environment. The platform also included gamification elements designed to increase engagement and motivation. Using a Design Science Research approach, the study combined system development, classroom testing, and analysis of behavioral log data from 91 students. The results indicate that students were often able to identify manipulated content, although realistic AI-generated fake news remained challenging to detect. Students frequently relied on headlines, images, and perceived source credibility when evaluating information. The study suggests that AI-supported and gamified learning environments may strengthen source evaluation, critical thinking, and digital resilience among young users.**

Keywords-fake news; source evaluation; teaching; generative AI; upper secondary school students.

I. INTRODUCTION

The rapid digitalization of society has fundamentally transformed how information is produced, distributed, and consumed [1]. News and public information are no longer primarily accessed through traditional editorial channels such as newspapers, television, and radio. Instead, social media platforms and algorithm-driven digital feeds increasingly shape how people encounter information in their daily lives. Platforms such as TikTok, Instagram, Snapchat, Facebook, and X provide users with a continuous stream of content in which journalism, entertainment, advertising, opinion, and user-generated material are often presented side by side [2]. While this development has increased access to information and enabled broader participation in public discourse, it has also created significant challenges related to misinformation, disinformation, and source evaluation.

The spread of false or misleading information is not a new phenomenon, but the scale and speed of dissemination have changed considerably in the digital era. Social media platforms enable information to spread rapidly across large networks without traditional editorial review or fact-checking processes. Research has shown that false news often spreads faster and reaches larger audiences than information, particularly when the content is emotionally engaging, sensationalized, or controversial [3]. At the same time, modern digital platforms are designed to maximize engagement and user interaction, which may unintentionally amplify misleading or manipulative content.

Recent developments in generative artificial intelligence (AI) have further intensified these challenges. Modern language and image generation models can produce highly realistic text, images, audio, and video content that may be difficult for ordinary users to distinguish from authentic material. AI-generated misinformation can imitate the visual style, language, and structure of professional journalism, making manipulated content appear highly credible. As generative AI tools become more accessible and easier to use, producing large-scale disinformation campaigns requires significantly fewer technical skills and resources than before. This development has heightened concerns about digital trust, democratic participation, cybersecurity, and the public's ability to critically evaluate online information [4].

These challenges are particularly relevant for adolescents and young adults. According to the Norwegian Media Authority, children and young people spend a large portion of their daily lives on digital platforms, where social media functions as a major source of entertainment, communication, and news consumption [5]. Although many young users are aware that false information exists online, research indicates that they still struggle to identify misinformation in practice, especially when manipulated content appears professional, visually convincing, or emotionally engaging. Young users often encounter information in fast-moving feeds where rapid impressions, headlines, images, and perceived source familiarity influence credibility judgments more strongly than systematic fact-checking or deeper source evaluation. As a result, source evaluation, critical thinking, and digital judgment have become increasingly important competencies in modern education.

Educational institutions are expected not only to teach technical digital skills but also to prepare students to critically evaluate information, reflect on credibility, and participate responsibly in digital society. Previous research suggests that practical, interactive learning activities may be more effective than purely theoretical instruction for teaching media literacy and source evaluation [6]. Students benefit from realistic environments where they actively

 ISBN: 978-1-68558-400-9

engage with information and practice identifying manipulative or misleading content.

The study has two objectives: (1) to examine students' ability to distinguish between authentic and AI-generated fake news, and (2) to explore the cues and evaluation strategies they use when assessing the credibility of digital news content.

The project involved the development of a social media-inspired digital platform in which students were presented with both real news articles and AI- manipulated news in an interactive feed. The platform also incorporated gamification elements such as points, streaks, badges, and progression systems to increase engagement and motivation.

The study was conducted using a Design Science Research approach, combining technical system development with empirical user testing and analysis of behavioral log data. A total of 91 students from four upper secondary school classes participated in the testing process. The objective of the study was not only to examine whether students could correctly identify fake news, but also to explore how realistic digital learning environments, AI-generated content, and gamification could help strengthen critical thinking and source evaluation in practice. In addition, the project explores how generative AI can simultaneously represent both a societal challenge and a pedagogical opportunity within education related to media literacy, digital resilience, and cybersecurity.

II. Related Work

The growing prevalence of misinformation and disinformation in digital environments has become a major research topic across media studies, education, information systems, and cybersecurity. The concept of fake news is commonly used to describe false or misleading information presented in a format resembling legitimate journalism or trustworthy reporting. Lazer et al. describe how modern digital media ecosystems enable misinformation to spread rapidly through social platforms where content distribution is largely driven by user engagement rather than editorial verification [7]. Unlike traditional media, social networking platforms allow virtually anyone to publish and distribute information instantly to large audiences, making it increasingly difficult for users to distinguish between verified journalism, advertising, opinion-based content, and intentionally manipulative information.

Researchers frequently distinguish between misinformation, disinformation, and malinformation. Misinformation refers to false information shared without the intention to deceive, while disinformation involves deliberately fabricated or manipulated information distributed to influence or mislead others. Malinformation, in contrast, refers to authentic information used in misleading or harmful contexts. These distinctions are important because they highlight different motivations and mechanisms behind the spread of misleading information and suggest that different educational and technological approaches may be required to address them effectively.

The rise of generative artificial intelligence has significantly intensified concerns related to misinformation. Modern AI models can generate highly realistic text, images, and multimedia content that imitate professional journalism both visually and linguistically. According to the European Union Agency for Cybersecurity (ENISA), AI-generated misinformation increases the risk of scalable and highly convincing influence operations, making digital source evaluation increasingly important in modern society. AI systems can produce manipulated content quickly and at low cost, lowering the barrier for creating large quantities of believable fake news. At the same time, researchers have emphasized that AI should be viewed not only as a threat but also as a potential educational resource that supports learning and critical reflection [6].

Research concerning adolescents’ digital media habits further illustrates the relevance of this challenge. The Norwegian Media Authority reports that children and young people spend substantial time on digital platforms, where social media serves as a central source of communication, entertainment, and news [5]. Information presented through algorithm-driven feeds is often consumed rapidly and evaluated based on visual impressions, emotional reactions, or familiarity with the source rather than systematic verification. Pérez-Escoda et al. [8] found that although young users are generally aware that misinformation exists online, they continue to rely heavily on social media for information and news. This creates a paradox in which users are simultaneously skeptical of digital information, yet still depend on the same platforms for everyday access to information.

Several studies have shown that users often rely on heuristics when evaluating online information [4]. Flintham et al. [9] demonstrated that users frequently assess credibility based on headlines, images, source familiarity, and emotional framing rather than reading full articles or conducting deeper fact-checking. Similarly, Jones-Jang et al. [10] found that professionally designed misinformation containing realistic visuals and persuasive headlines may be particularly difficult for young users to identify. These findings indicate that digital literacy requires more than awareness of misinformation; users must also develop practical evaluation skills and reflective decision-making strategies.

Media literacy, source evaluation, and critical thinking have therefore become increasingly important concepts in educational research. López-Meri et al. [11] argue that digital competence involves not only technical skills, but also reflection, analysis, and the ability to critically evaluate information in digital environments. Research further suggests that active, practice-oriented learning approaches are more effective than purely theoretical instruction for developing such competencies. Adjin Tettey demonstrated that participants receiving practical media literacy training became significantly better at identifying fake news and less likely to share misleading information online [6]. These findings emphasize the importance of realistic learning environments in which users actively engage with manipulatives and authentic content.

Gamification has also received increasing attention in research on digital learning environments. Gamification refers to the use of game-inspired elements such as points, badges, progression systems, streaks, and rewards to increase motivation and engagement [10]. Capecchi et al. found that gamified learning activities addressing misinformation and fake news increased user engagement and encouraged reflection on information credibility and online behavior. Interactive learning environments may therefore contribute to more active participation and sustained motivation compared to traditional lecture-based approaches.

This study builds on previous research by combining AI-generated fake news, gamification, and a realistic, social media-inspired interface into a single learning platform. While earlier studies have often focused on surveys, theoretical discussions, or isolated experiments, this project investigates how students interact with both authentic and manipulated news content in a realistic digital environment that resembles their everyday media experiences. The study, therefore, contributes to existing research on digital literacy, educational technology, and cybersecurity awareness by exploring how generative AI can be used responsibly to strengthen critical thinking and source evaluation among young users.

III. METHODOLOGY

This study was conducted using a Design Science Research (DSR) approach, with the primary objective of designing, developing, and evaluating a technological artifact in a realistic educational context [13]. Design Science Research focuses on the creation and assessment of practical technological solutions while simultaneously contributing knowledge about how such solutions function in practice. The study, therefore, combined system development, user testing, and empirical analysis to investigate how an AI-based learning platform could support source evaluation and critical thinking among upper secondary school students.

The project also adopted characteristics of a mixed methods approach by combining quantitative and qualitative data collection. Quantitative data primarily consisted of automatically generated logs from student interactions with the platform. These data included classifications of news articles as real or fake, response accuracy, selected justifications, progression, streaks, time usage, and navigation patterns. Qualitative data consisted of classroom observations during testing sessions and anonymous student feedback collected through the Mentimeter survey tool after the tests. The combination of behavioral log data and qualitative reflections provided a broader understanding of how students interacted with the platform and evaluated digital news content.

The developed platform served as a social media-inspired news feed, presenting students with a mix of authentic and AI-manipulated news articles. The system architecture consisted of a web-based frontend developed using HTML (Hypertext Markup Language), CSS (Cascading Style Sheets), and JavaScript, combined with Python-based backend scripts and API (Application Program Interface) integrations for data processing and AI-generated content creation. Firebase Firestore was used as the primary database solution for storing news articles, user activity, and log data.

Authentic news articles were automatically collected from RSS feeds provided by established news organizations through a Python script. The original articles were stored in the database and then processed using the Google Gemini language model to generate modified versions of the same content. The AI-generated articles were stored separately from the authentic articles, enabling the platform to dynamically create realistic news feeds containing both real and manipulated news. To further enhance realism, selected articles also included AI-generated or manipulated images produced by generative image models. All generated content was manually reviewed before publication within the platform to remove unrealistic, offensive, or otherwise inappropriate material. Figure 1 illustrates the system workflow. First, authentic news articles are retrieved from RSS (Really Simple Syndication) feeds (1). Selected articles are then manipulated using the Gemini model (2) and stored alongside authentic articles in Firebase Firestore (3). The news feed is presented to users through the frontend interface (4), while user interactions and classifications are logged for later analysis (5).

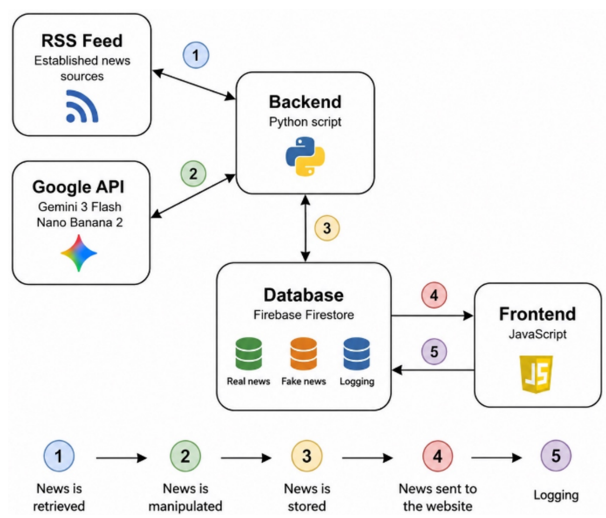

Figure 1. Simplified overview of the platform architecture.

The platform interface was designed to resemble modern social media environments familiar to adolescents. Inspired by platforms such as TikTok, Instagram, Snapchat, and X, the interface presented articles in a continuous feed that users could navigate through scrolling or swipe-based interactions. Gamification elements, including points, streaks, badges, stars, and progression systems, were integrated into the platform to increase engagement and motivation. The user interface is shown in Figure 2.

User testing was conducted with four upper secondary school classes involving a total of 91 students. The tests were conducted in classroom settings and lasted approximately 60 minutes per session. Students first received a shared introduction to the platform and testing procedure before interacting with the system individually. In addition, source evaluation and digital literacy are already part of the

Norwegian school curriculum. Students were encouraged to verify information using external sources and evaluate factors such as source credibility, language, and supporting evidence rather than relying solely on intuition. During the sessions, students evaluated news articles as either authentic or fake and selected two or three justifications supporting their decisions.

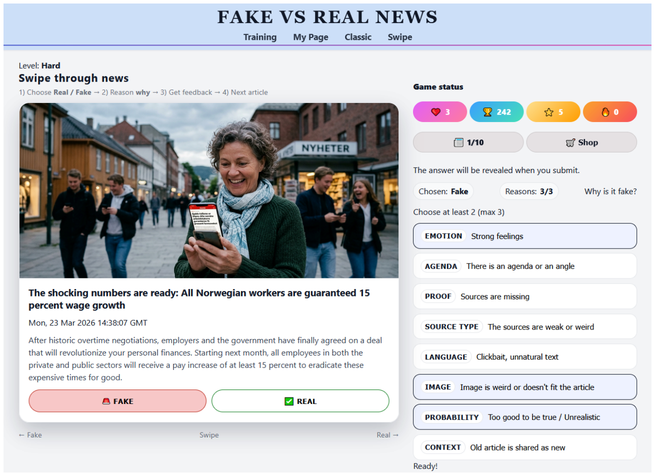

Figure 2. The user interface. Originally in Norwegian, but translated for this paper.

Ethical considerations and privacy protection were central throughout the project. No personally identifiable information was collected, and all participants received randomly generated anonymous user IDs. The platform was password-protected and used exclusively within controlled classroom environments to minimize the risk of distributing AI-generated misinformation outside the educational setting. In addition, all participants were informed that the platform contained both authentic and manipulated news articles as part of a controlled learning activity.

IV. RESULTS

The platform was evaluated through classroom testing involving 91 upper secondary school students from four different classes. The testing sessions were conducted in controlled classroom environments and lasted approximately 60 minutes each. During the sessions, students interacted individually with the platform, where they evaluated news articles presented in a social media-inspired feed containing both authentic and AI-generated fake news articles. User interactions, decisions, progression, and selected justifications were automatically logged for later analysis.

TABLE I. FIELD TRIALS

	School	Date	Students	Testing time
1	Sande High School	12.03.26	16	60 min.
2	Sande High School	12.03.26	27	60 min.
3	Thor Heyerdahl High School	24.03.26	26	40 min.
4	Sande High School	13.04.26	22	60 min.

Overall, the results indicate that students were often able to distinguish between authentic and manipulated news articles, although the difficulty varied considerably depending on the content's design and realism. Articles containing emotionally engaging headlines, realistic images, or references to apparently credible sources were significantly more challenging for students to classify correctly. In several cases, AI-generated fake news articles were perceived as authentic because they imitated the visual style and language typically associated with professional journalism. Authentic news articles were correctly identified in weighted average at approximately 75% of the cases, while AI-generated fake news articles were correctly identified in weighted average at approximately 70% of the cases. Conversely, some authentic news articles were incorrectly classified as fake when the topics appeared unusual, sensational, or emotionally provocative.

The analysis of student justifications revealed several recurring evaluation strategies. Many participants based their decisions on visual and contextual elements such as headlines, language style, source familiarity, and the perceived professionalism of images and layout. Known or recognizable sources were often interpreted as indicators of credibility, while emotionally charged headlines, exaggerated wording, and missing references frequently triggered suspicion. Students also commonly referred to whether the article appeared "realistic" or "trustworthy" rather than relying on systematic fact-checking or external verification. These findings suggest that students frequently use heuristic reasoning and rapid impressions when evaluating information in digital environments.

Figure 3 presents the confusion matrix of students' classifications. The green cells represent correct classifications (true positives and true negatives), while the red cells represent incorrect classifications (false positives and false negatives).

		Student's choice (prediction): Chose real	Student's choice (prediction): Chose fake	Row total (actual total)
Actual truth	Actually real	1859 (True Positive) Correctly identified as real	793 (False Negative) Incorrectly identified as fake	2652
Actual truth	Actually fake	404 (False Positive) Incorrectly identified as real	753 (True Negative) Correctly identified as fake	1157
Column total (predicted total)		2263	1546	3809

Figure 3. Confusion matrix of students' classifications of real and fake news.

The results also showed variation in how students approached the evaluation tasks. Some participants spent considerable time reading articles carefully and reviewing details before making decisions, while others relied primarily on headlines and images, rapidly classifying the content. Log data indicated that longer interaction times were often

 ISBN: 978-1-68558-400-9

associated with more accurate classifications, suggesting that reflective evaluation strategies may improve source evaluation performance. At the same time, some students demonstrated consistent patterns of either skepticism or trust toward nearly all articles, regardless of authenticity. The gamification system included lives, points, stars, and streaks to encourage engagement and continued interaction with the platform. Lives limited the number of incorrect classifications, while points were awarded for correctly identifying both real and fake news articles. Students earned stars for successfully identifying fake news, and streaks tracked consecutive correct classifications to encourage consistent performance and engagement. The implemented gamification elements are illustrated in Figure 4.

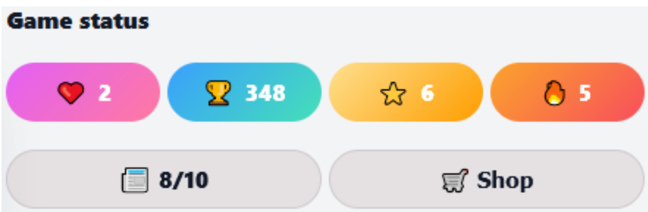

Figure 4. Selected gamification elements used in the platform.

The gamification elements integrated into the platform appeared to contribute positively to student engagement during the testing sessions. Features such as points, streaks, badges, and stars encouraged active participation and motivated students to continue interacting with the platform. Observations from the classroom sessions indicated that many students became increasingly competitive and invested in improving their performance over time. Several students also demonstrated curiosity regarding why particular articles were classified as fake or authentic after receiving feedback from the system.

Student feedback collected after the sessions further supports the value of realistic and interactive learning environments. Many participants reported that the platform was engaging. Several students stated that the activity made them more aware of how difficult it can be to distinguish between authentic and manipulated information online. Others highlighted how realistic AI-generated content challenged their assumptions regarding source credibility and visual trustworthiness.

At the same time, the results also revealed limitations and challenges. Some students appeared to focus more on gameplay elements and progression systems than on careful reflection and source evaluation. In addition, certain AI-generated articles were considered either too unrealistic or too easy to identify, indicating that the quality and balance of generated misinformation remain important for maintaining educational realism. The findings, therefore, suggest that AI-generated educational content requires careful human supervision and iterative refinement to ensure both pedagogical value and ethical responsibility.

Overall, the results indicate that AI-supported and gamified learning environments may offer valuable opportunities to strengthen source evaluation and critical thinking among young users. The findings also highlight the growing importance of practical training environments where students actively engage with realistic misinformation rather than learning about the phenomenon only theoretically.

V. Discussion

The findings from this study suggest that realistic and interactive digital learning environments may contribute positively to students' ability to critically evaluate online information [4]. Most participants correctly identified a substantial number of both authentic and AI-generated fake news articles, indicating that many students already had some awareness of misinformation and source evaluation. At the same time, the results demonstrated that realistic fake news articles remained difficult to detect when they included convincing headlines, emotionally engaging language, realistic imagery, or references to seemingly credible sources. These findings reflect broader societal challenges associated with modern digital media environments, where manipulated information is increasingly sophisticated and visually convincing.

One important observation from the study is that students frequently relied on rapid impressions and heuristic decision-making when evaluating content. Headlines, images, language style, and perceived source familiarity appeared to influence credibility assessments more strongly than detailed analytical reasoning or systematic fact-checking. This behavior is consistent with earlier research showing that users often evaluate information quickly in social media environments characterized by high information volume and continuous scrolling. The platform's feed-based interface, therefore, appears to have successfully simulated the digital environment in which young people commonly encounter information in everyday life. By reproducing realistic social media interaction patterns, the project increased ecological validity and provided insight into how students evaluate information in practice rather than under highly artificial testing conditions.

The study further indicates that practical, experience-based learning activities may strengthen reflection and engagement in source evaluation. Instead of receiving purely theoretical instruction about misinformation, students actively interacted with authentic and manipulated news content and were required to justify their decisions. This aligns with earlier studies suggesting that media literacy training is more effective when users actively practice evaluating information in realistic scenarios. The requirement that students provide justifications for their classifications may have encouraged more reflective decision-making and made students more conscious of the criteria they used when evaluating credibility.

The gamification elements implemented in the platform also appeared to contribute positively to engagement and motivation. Features such as points, streaks, badges, stars, and progression systems encouraged active participation throughout the testing sessions. Observations during the classroom sessions suggested that many students became increasingly invested in the activity and demonstrated curiosity regarding why some articles were authentic while others were manipulated. Gamification may therefore help transform source evaluation from a purely academic exercise into a more engaging and interactive learning activity. However, the study does not isolate the specific effects of gamification from the broader platform design, so the motivational impact cannot be measured independently of the realistic feed environment and AI-generated content.

An important contribution of the project is the demonstration that generative AI can function simultaneously as both a societal challenge and an educational opportunity. Modern AI models can generate highly realistic misinformation at scale, representing an increasing concern for democratic processes, digital trust, and cybersecurity. However, the same technology can also be used constructively in controlled educational settings to create dynamic and realistic learning materials. In this project, AI enabled the automatic generation of manipulated news articles based on authentic news stories, making it possible to rapidly produce varied and realistic training scenarios. This would have been significantly more time-consuming using fully manual approaches.

At the same time, the study highlighted the importance of human oversight and ethical control when using generative AI in educational contexts. All AI-generated articles and images were manually reviewed before publication to remove inappropriate, unrealistic, or potentially harmful material. This human-in-the-loop approach proved important both for quality assurance and for reducing the risk of unintentionally distributing harmful misinformation. The findings therefore support previous arguments that AI systems should augment rather than replace human judgment in educational and information-related contexts.

The study also has several limitations. First, the testing sessions were relatively short and focused primarily on short-term interactions rather than long-term learning outcomes. It therefore remains uncertain whether repeated exposure to similar learning activities would produce lasting improvements in critical thinking and digital source evaluation. Second, the participant group consisted of only 91 students from four upper secondary school classes, limiting the generalizability of the findings. The platform was also limited to text-based articles and static images, while contemporary misinformation increasingly includes video, audio, deepfakes, and short-form social media content.

Despite these limitations, the study demonstrates the potential of combining AI-generated content, gamification, and realistic digital interfaces to support critical thinking and source evaluation among young users. The findings further suggest that digital literacy and cybersecurity are becoming increasingly interconnected, as the ability to critically evaluate information may represent an essential component of digital resilience in modern societies.

VI. Conclusion

This study investigated how an AI-based learning platform can help upper secondary school students distinguish between authentic and fake news in digital media environments. The project combined artificial intelligence, gamification, and social media-inspired design to create a realistic and interactive learning environment where students actively evaluated news credibility and reflected on their decisions.

The findings indicate that many students were able to identify both authentic and manipulated news articles, although realistic AI-generated misinformation could still be difficult to detect. Students frequently relied on visual impressions, headlines, language style, and perceived source credibility when evaluating content. These findings illustrate how information is often consumed in modern digital environments and highlight the importance of strengthening critical thinking, source evaluation, and digital judgment among young users.

The study also demonstrates that generative AI can serve as both a societal challenge and a pedagogical opportunity. While AI increases the ability to produce convincing misinformation at scale, the same technology can be used in controlled educational settings to create realistic training scenarios that support active learning and reflection. The project further showed that gamified and interactive learning activities may increase engagement and motivation compared to traditional theoretical approaches to media literacy education.

From a methodological perspective, the project demonstrates the usefulness of a Design Science Research approach for developing and evaluating educational technology solutions in realistic contexts. By combining technical development, classroom testing, log-data analysis, and qualitative feedback, the study provided insights into both user behavior and system design.

At the same time, the study has limitations. The testing sessions were short and focused on short-term interactions rather than long-term learning outcomes. In addition, the participant group was relatively small and limited to upper secondary school students. Future research should therefore investigate long-term educational effects, broader participant groups, and additional media formats such as video, audio, and short-form social media content.

Overall, the study suggests that realistic digital learning environments, supported by AI and gamification, may contribute to improved reflection and awareness regarding misinformation and digital source evaluation. As misinformation and AI-generated content continue to evolve, educational approaches that combine technology, critical thinking, and practical training may become increasingly important for strengthening digital resilience and informed participation in society.

Acknowledgment

Co-Funded by the European Union through the ERASMUS-EDU-2025-PI-FORWARD Call under Grant Agreement no. 101259941

The authors would like to thank the teachers and students who contributed to the study.

References

[1] A. Marwick and R. Lewis, “Media manipulation and disinformation online,” Data & Society Research Institute, 2017.

[2] The Norwegian Media Authority (Medietilsynet), “Barn og medier 2024 (Children and Media 2024),” 2024.

[3] S. Vosoughi, D. Roy, and S. Aral, “The spread of true and false news online,” Science, vol. 359, no. 6380, pp. 1146–1151, 2018.

[4] ENISA, “ENISA Threat Landscape 2023,” European Union Agency for Cybersecurity, 2023.

[5] The Norwegian Media Authority (Medietilsynet), “Barn og medier 2022 (Children and Media 2022),” 2022.

[6] T. D. Adjin-Tettey, “Combating fake news, disinformation, and misinformation,” Cogent Arts & Humanities, vol. 9, no. 1, 2022.

[7] D. M. J. Lazer et al., “The science of fake news,” Science, vol. 359, no. 6380, pp. 1094–1096, 2018.

[8] A. Pérez-Escoda, J. M. Pedrero-Esteban, J. Rubio Romero, and C. Jimenez-Narros, “Fake news reaching young people on social networks: Distrust Challenging Media Literacy,” Publications, vol. 9, no. 2, 2021.

[9] M. Flintham et al., “Falling for fake news,” Proc. CHI Conference on Human Factors in Computing Systems, pp. 1–10, 2018.

[10] S. M. Jones-Jang, T. Mortensen, and J. Liu, "Does media literacy help identification of fake news? Information literacy helps, but other literacies don’t," American Behavioral Scientist, vol. 65, no. 2, pp. 371–388, 2021.

[11] A. López-Meri, H. Doménech-Fabregat, and S. Marcos-García, "Digital competencies in verifying fake news: Assessing the knowledge and abilities of journalism students," Societies, vol. 14, no. 5, Art. no. 66, 2024.

[12] S. Capecchi et al. “A gamified platform to support educational activities about fake news in social media”, IEEE Transactions on Learning Technologies, vol. 17, pp. 1765-1779, 2024.

[13] A. R. Hevner et al., “Design science in information systems research,” MIS Quarterly, vol. 28, no. 1, pp. 75–105, 2004.

 ISBN: 978-1-68558-400-9