Advances in Military Technology
Vol. 21, No. 2, 2026, pp. 375-394
ISSN 1802-2308, eISSN 2533-4123
DOI 10.3849/aimt.02077

Automation in Military Intelligence Workflows: Integrating RPA, Databases, and Offline LLMs

O. Kačmařík*, R. Vašíček, J. Podoba, and J. Říha

Department of Intelligence Support, University of Defence, Brno, Czech Republic

The manuscript was received on 6 January 2026 and was accepted after revision for publication as an **original research article** on 18 August 2026.

Abstract:

This article examines the integration of Robotic Process Automation (RPA), a structured PostgreSQL database, and an offline Large Language Model (LLM) within a military Intelligence Requirements Management and Collection Management (IRM&CM) workflow. Using a comparative case study of two intelligence staffs – one employing intelligent automation and one relying on traditional methods – the study evaluates impacts on processing speed, accuracy, analytical quality, and workload. Results show that automation substantially improves report throughput, reduces procedural errors, and enables faster, more coherent intelligence production while maintaining analyst oversight. The findings demonstrate that secure, offline Artificial Intelligence (AI) combined with RPA can significantly enhance lower-echelon intelligence operations without compromising doctrinal standards or analytical integrity.

Keywords:

Artificial Intelligence, Robotic Process Automation, military intelligence, Large Language Models, automated data processing, intelligence requirements management

1 Introduction

Modern military intelligence operates in increasingly complex, data-rich environments where the speed and heterogeneity of incoming information routinely exceed the capacity of traditional workflows. Intelligence staffs at all echelons must process reports from sensors, digital platforms, human sources, and partner forces while maintaining accuracy, timeliness, and traceability. United States (US) Army doctrine focusing on Intelligence Analysis [1], emphasizes that intelligence analysis is both iterative and labor-intensive, requiring structured reasoning and disciplined information management. Yet many lower-echelon staffs continue to rely on fragmented and largely

* *Corresponding author: Department of Intelligence Support, University of Defence, Kounicova 156/65, CZ-662 10 Brno, Czech Republic. Phone: +420 973 442 574, E-mail: ondrej.kacmaril@unob.cz. ORCID 0009-0000-3023-8149.*

DOI 10.3849/aimt.02077

manual processes that are not well suited to the volumes and tempo of contemporary operations.

These challenges are especially evident within the Intelligence Requirements Management and Collection Management (IRM&CM) cycle, which forms the organizational backbone of intelligence activities (see Fig. 1). IRM&CM tasks such as receiving and triaging reports, extracting metadata, assigning information to PIRs, updating event logs, and producing finished assessments demand precision and consistency [2]. In practice, however, intelligence sections often perform these tasks manually using diverse tools and inconsistent workflows. Observations from the case environment revealed recurring inefficiencies: data fragmentation across formats and platforms, high clerical workload, inconsistent metadata entry, and difficulty maintaining situational awareness during peak reporting periods. Such limitations slow the intelligence cycle, introduce opportunities for error, and reduce the time analysts can devote to higher-order reasoning.

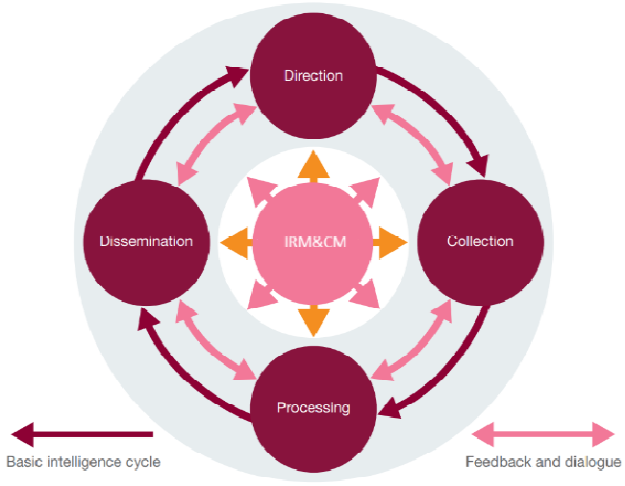


Fig. 1 IRM&CM as a backbone of the Intelligence cycle [3]

Defense modernization strategies recognize these challenges. For example, the Concept for the Development of the Armed Forces of the Czech Republic 2035 [4] calls for increased automation in the exploitation of sensor data and semi-automated intelligence processing. Similar trends exist across NATO, where the growing complexity of operational environments and constraints on human resources have accelerated interest in intelligent automation. At the same time, military settings impose specific requirements on any such technology: strict security and offline operability, full auditability, human-in-the-loop oversight, and adherence to doctrinal analytic standards. These constraints demand systems that enhance intelligence production without compromising analytical rigor or operational security.

The present study responds to this need by evaluating an integrated approach combining Robotic Process Automation (RPA), a structured PostgreSQL intelligence database, and an offline Large Language Model (LLM). Together, these components form the IRM&CM Cell Station, a prototype system designed to automate routine administrative tasks, provide structured data management, and support analysts with AI-assisted summarization and contextual reasoning. The system aims to relieve ana-

lysts of repetitive clerical work, improve data consistency, and accelerate the production of intelligence products while retaining human authority over analytic judgments.

To assess the system's effectiveness, the study compares two intelligence staffs performing identical tasks under controlled conditions: one using the automation-supported workflow, the other relying entirely on traditional manual methods. The comparison focuses on processing speed, error frequency, analytic quality, and perceived workload. By examining the operational and cognitive implications of intelligent automation, the study contributes to the broader discussion of how AI-enabled tools can responsibly enhance military intelligence processes.

2 Theoretical Background and Problem Formulation

The theoretical background of this study relies primarily on unclassified United States (US) Army doctrinal publications and United Kingdom (UK) Joint Doctrine Publications (JDPs) rather than on NATO Allied Joint Publications (AJPs). This choice is necessitated by classification constraints. While US and UK doctrinal materials, including analytical methodologies and intelligence process descriptions, are publicly accessible and authorized for unrestricted academic use, most NATO intelligence-related AJPs are releasable for official use only (NATO UNCLASSIFIED) or higher. These restrictions limit their suitability for open academic research and prevent their citation within unclassified scholarly work.

US Army Intelligence analysis doctrine defines analysis as a structured, iterative process that transforms raw data into actionable understanding for commanders. This doctrine emphasizes four core phases: screening, analysis, integration, and production, each of which relies on accurate information management and consistent analytical discipline. At lower echelons, however, these phases are often constrained by limited technological support, heterogeneous data streams, and a high volume of reports that must be processed rapidly under operational pressure.

Central to this workflow is the Intelligence Requirements Management and Collection Management (IRM&CM) cycle, which ensures that reporting remains aligned with Priority Intelligence Requirements (PIRs) and that collected information is systematically catalogued and assessed. IRM&CM involves triage, metadata assignment, correlation with indicators, and continuous updating of the collection plan. In many units these activities remain largely manual, leading to delays, inconsistent data handling, and reduced situational awareness.

Automated Data Processing (ADP) offers a relevant response to these challenges. ADP systems reduce human error, accelerate data handling, and support large-scale processing of heterogeneous inputs across sectors such as healthcare, finance, and logistics [5]. In military environments, these benefits translate into faster ingestion of sensor reports, more reliable metadata extraction, and more consistent assignment of information to doctrinal categories.

RPA represents a practical subset of ADP particularly well suited to repetitive intelligence tasks. RPA tools simulate human interactions with digital systems, enabling automated file conversion, format normalization, data extraction, and basic classification. Civilian studies consistently show that RPA reduces processing time and increases accuracy while maintaining a transparent audit trail [6]. Within intelligence workflows, RPA can eliminate clerical bottlenecks, allowing analysts to focus on interpretation rather than data preparation.

Recent advances in artificial intelligence have enabled the integration of machine learning and Large Language Models (LLMs) into analytic workflows. LLMs can perform summarization, classification, entity extraction, and multi-document synthesis – capabilities that parallel key functions of intelligence analysis. Their use in defense, however, is constrained by security, traceability, and the doctrinal requirement for human-in-the-loop oversight. Offline, locally hosted models offer a viable compromise, reducing the risk of data exposure while providing analysts with contextualized support during information assessment [7].

Structured databases form another essential component of modern intelligence processing. A relational system such as PostgreSQL provides consistent data storage, searchability, and cross-referencing, which improves analytic reliability and reduces information loss. Prior research highlights that AI systems perform more predictably and with lower uncertainty when supported by structured, validated datasets [8].

Finally, human-machine interaction research underscores the continued importance of analyst agency. Automation should reduce cognitive load and enhance analytic clarity but not replace the judgment of trained professionals. Effective integration therefore requires transparent interfaces, traceable system behavior, and enforceable human validation at all critical decision points [9].

Taken together, these theoretical foundations provide the rationale for evaluating an integrated system combining RPA, structured data management, and offline LLM support within the IRM&CM workflow. Fig. 2 depicts IRM&CM Cell Station as a key driving element of the Collation substep of the Intelligence cycle.

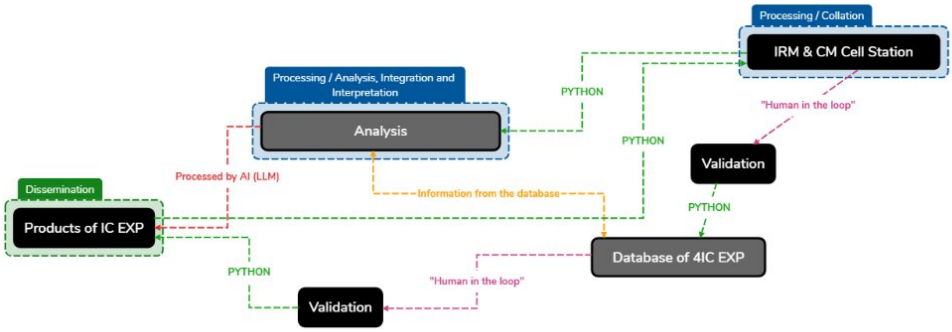


Fig. 2 Integration of the IRM&CM Cell Station within the workflow

3 Methodology

3.1 Research Design

This study adopts a comparative case study design to examine the effects of intelligent automation on military intelligence workflows within the Intelligence Requirements Management and Collection Management (IRM&CM) cycle. The methodological approach combines quantitative and qualitative methods to assess both procedural performance and analytical effectiveness. Two intelligence staffs operating under identical scenario conditions were compared: one employing an automation-supported workflow (Staff A) and one relying on traditional manual methods (Staff B). This

design enables the isolation of workflow-related differences while maintaining analytical realism.

During the experiment, a total of 263 input reports were processed, with both staffs receiving an identical number of reports in accordance with the training scenario design to ensure comparable working conditions. In terms of evaluated output, Staff A generated 157 reports, while Staff B produced 129 reports during the exercise. Throughout the analytical process, 99 relevant cases were identified, with the same number of cases available to both staffs as part of the controlled experimental setting.

3.2 Case Environment and Technical Setup

The study was conducted within a simulated but technically authentic intelligence operations setting within annual student training activity Agile Cub 2025 [10]. The experimental environment was designed to replicate the operational conditions of a lower-echelon military intelligence cell while remaining within unclassified and controlled constraints. All activities were conducted in a closed, offline environment without access to external networks or cloud-based services.

For consistency across data processing and the international character of training activity, the operational language environment including the input data provided to the system, the prompts formulated for the AI model, and the generated outputs, was standardized in English. At the same time, the use of English ensured better compatibility with the offline LLMs used in the experiment, whose performance and language support for Czech remain more limited than for English.

Hardware Configuration

The automation-supported workstation used by Staff A was built on commercially available hardware representative of equipment realistically deployable within military units. The configuration consisted of:

- processor: Intel Core i7-14700F (Raptor Lake Refresh), up to 5.4 GHz,
- graphics processing unit: NVIDIA GeForce RTX 4060 with 8 GB VRAM,
- system memory: 16 GB DDR5 RAM.

This configuration provided sufficient computational capacity for running local automation processes, database operations, and an offline Large Language Model which was informed by the findings of a preceding case study focused on AI powered text categorization. The results demonstrated that transformer-based LLMs with parameter sizes in the range of approximately 14 billion parameters offered a favorable balance between linguistic capability, contextual sensitivity, and computational feasibility (see Fig. 3). The hardware choice reflects a balance between performance and practical deployability in non-specialized military IT environments.

Staff B used standard office-class workstations with no dedicated AI acceleration capabilities, reflecting typical conditions in units operating without automation support.

System Architecture / Processing Workflow

Fig. 4 illustrates the processing chain of the proposed automation framework. Incoming intelligence reports and sensor data represent the input layer, which are subsequently processed through automated data transformation, database storage, and AI-assisted analysis. The resulting outputs consist of structured intelligence products,

DOI 10.3849/aimt.02077

visual dashboards, and formatted reports distributed via an email client or folder structure. A detailed description of the specific software environment and tools used to implement these processes is provided in the following chapter.

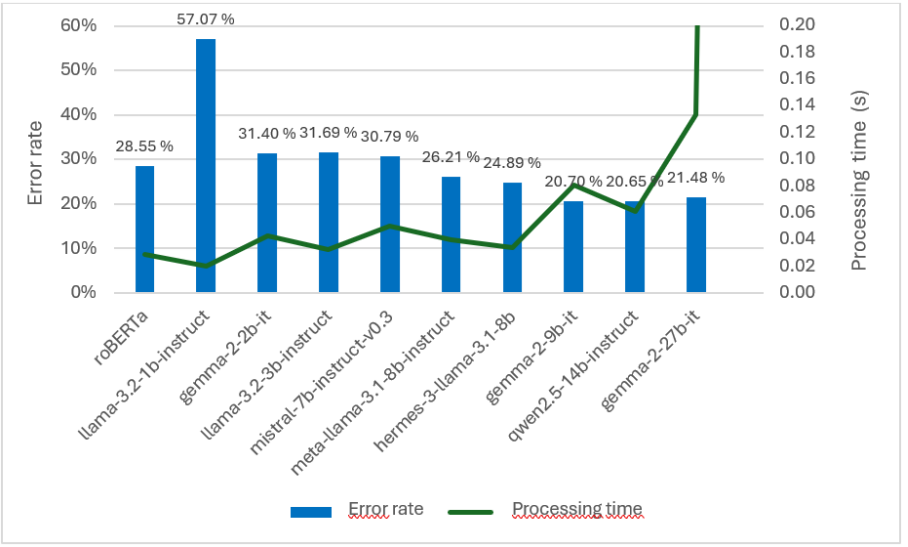


Fig. 3 Offline LLMs performance comparison [11]

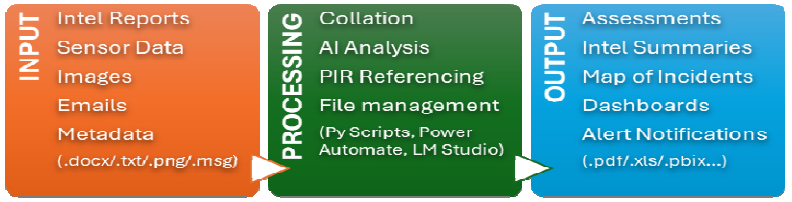


Fig. 4 Overview of the automated intelligence processing workflow (input – process – output model)

3.3 Software Environment and Automation Stack

The study Staff A's workflow was supported by an integrated software ecosystem enabling intelligent automation, analysis, and visualization. The following components were employed:

Robotic Process Automation:

Microsoft Power Automate (Fig. 5) and custom Python scripts were used to automate repetitive administrative and preparatory tasks, including file ingestion, format conversion, metadata extraction, directory management, and task notification.

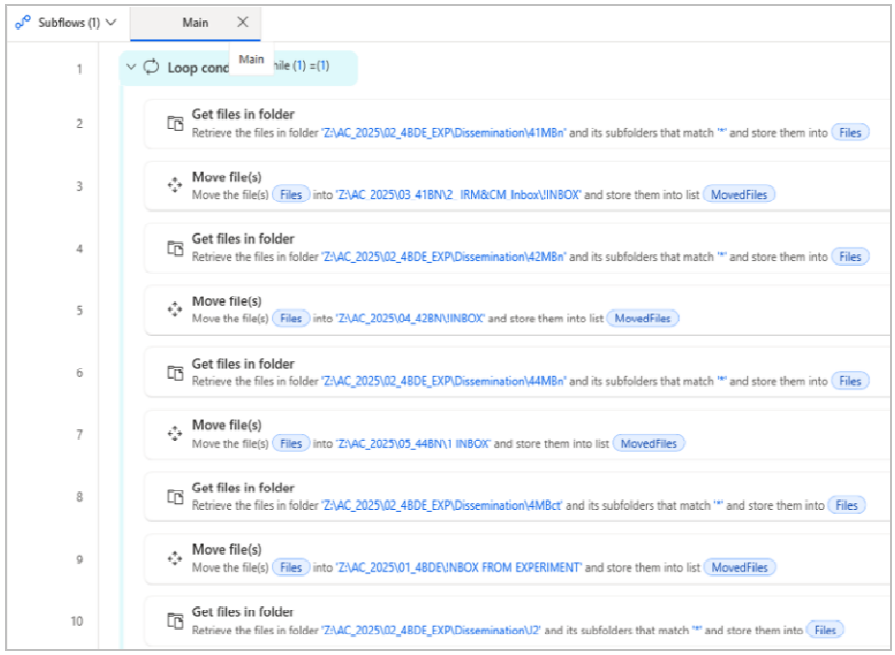


Fig. 5 Example of a "flow" in Microsoft Power Automate supporting directory management within the intelligence collation

Offline Large Language Model:

LM Studio was used to host a locally deployed OpenAI gpt-oss-20b LLM designed for powerful reasoning, agentic tasks, and versatile developer use cases. Fig. 6 illustrates a part of the specific prompt entered in the LM Studio chat window for generating the analytical intelligence assessment. The model was deployed in an offline environment to ensure operational security (OPSEC) and data control. For inference, the model was executed using a quantized configuration (4-bit quantization, GPTQ) to reduce memory requirements and enable efficient deployment on local hardware without significant degradation in performance. The inference pipeline was implemented using a standard transformer-based framework (e.g., Hugging Face Transformers-compatible runtime).

The following inference settings were applied consistently across all experiments:

- temperature: 0.2 (to favor deterministic and stable outputs),
- top-p (nucleus sampling): 0.9,
- max tokens: 1 024,
- repetition penalty: 1.1.

No fine-tuning was performed as the model was used in its pre-trained base checkpoint, with task-specific behavior controlled exclusively through structured prompting.

DOI 10.3849/aimt.02077

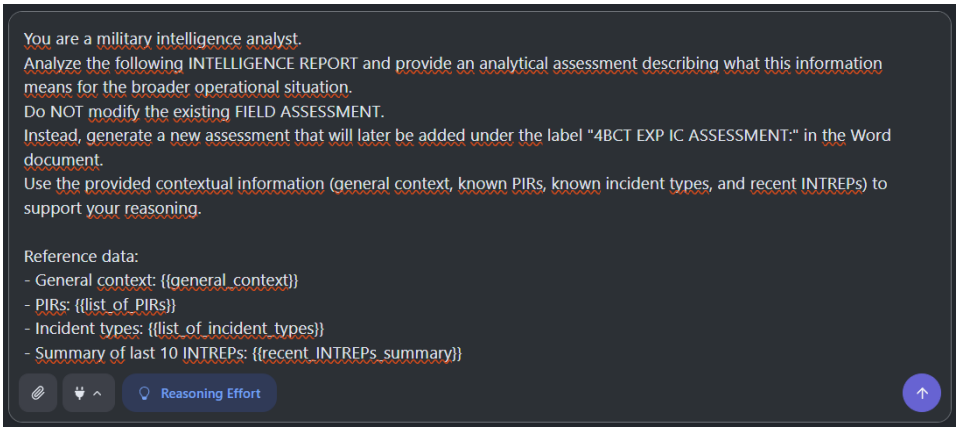


Fig. 6 Prompt used to generate an analytical assessment within the AI-assisted intelligence workflow

Data Management and Analysis

A PostgreSQL relational database (Fig. 7) served as the central knowledge repository, enabling structured storage of reports, metadata, indicators, and links to PIRs and ICPs.

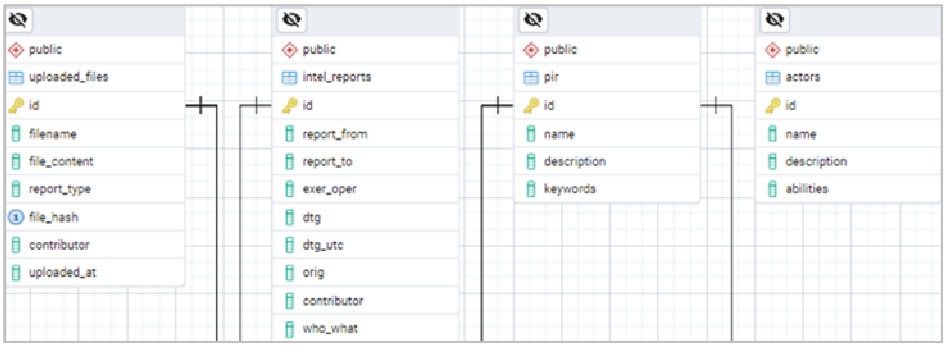


Fig. 7 Entity-Relationship (E-R) diagram of the database. Tables (entities) represent the types of objects (domains) monitored in the database, while attributes describe the properties of each entity

Visualization and Pre-Analysis

Microsoft Power BI was used for analytical visualization (Fig. 8), trend identification, and exploratory analysis of reporting patterns prior to formal intelligence assessment.

3.1 Participants

The study involved two intelligence staffs with deliberately different manning levels. Staff A consisted of five personnel, reflecting the assumption that intelligent automation reduces clerical workload. This staff was composed primarily of IT-enthusiastic participants with no prior experience in automation systems. Only one member pos-

sessed intermediate programming knowledge, while the remaining participants had no formal technical training beyond basic digital literacy.

Staff B was manned with seven members and consisted of participants with comparable intelligence training and experience but no exposure to automation or AI-supported tools. This composition ensured that observed performance differences could be attributed to workflow design rather than specialized technical expertise.

To ensure comparability between experimental groups with unequal staffing levels (Staff A: n = 5; Staff B: n = 7), all key performance indicators were normalized to a per person-hour basis. Specifically, aggregated values were divided by the number of personnel in each team, enabling efficiency comparisons independent of team size.

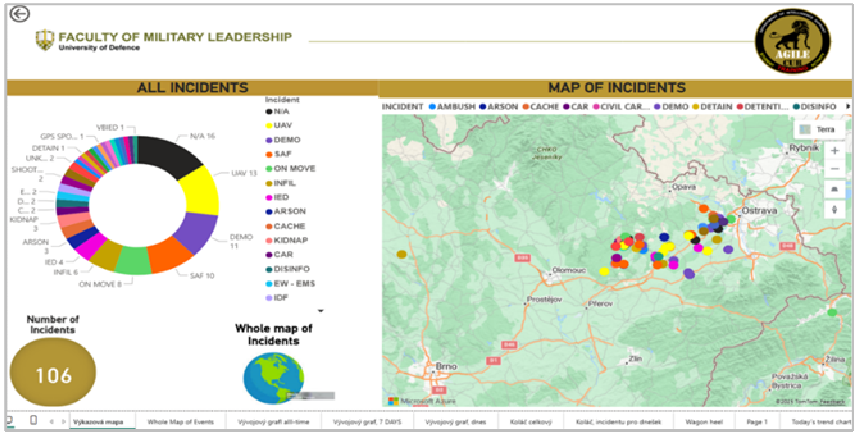


Fig. 8 Analytical visualization in Microsoft Power BI

3.2 Data Sources, Instruments and Evaluation Framework

The methodology employed a combination of quantitative and qualitative data collection instruments to construct a comprehensive baseline for evaluating both Measures of Performance (MOP) and Measures of Effectiveness (MOE).

Five quantitative instruments [12] were used to collect data related to processing speed, error rates, throughput, and analytic performance. These instruments enabled precise and replicable measurements of workflow efficiency for both Staff A (automation-enabled) and Staff B (manual processes). Tab. 1 summarizes the quantitative instruments and their respective measurement elements.

Quantitative findings were supplemented with qualitative insights to capture the human dimension of automation adoption, cognitive burden, workflow perception, and the perceived utility of analytic tools. Qualitative data were obtained from interviews, observations, and an independent peer review of final products.

MOPs quantify the observable, task-level performance of the system and are derived directly from measured experimental variables.

For each experiment i, the *MOP* score is defined as:

$$MOP_i = \sum_{j=1}^{n} w_j \hat{x}_{i,j} \tag{1}$$

where

- $x_{i,k}$ = the raw value of performance metric j in experiment i,

- $\hat{x}_{i,j}$ = the normalized value of metric j (scaled to [0;1]),
- w_j = the weight assigned to metric j, with $\sum w_j = 1$,
- n = the number of performance metrics.

Tab. 1 Quantitative instruments and their respective elements

Instrument	Description	Measured Elements
Time-stamped activity logs	Captured automatically from both Staff A and Staff B workstations.	Report receipt time; processing time per received report; analysis time per report; product creation time; triage delays.
RPA execution logs	Provided detailed traces of automated steps for Staff A.	File conversions; metadata extraction; database operations; notification triggers.
Analyst performance metrics	Collected to evaluate individual and team performance.	Number of reports processed; error counts (incorrect metadata, missed events); number of product revisions.
Automated and manual product comparison	Products evaluated using a predefined 5-criterion rubric.	Language accuracy; completeness; structure; traceability; alignment with PIRs.
Throughput benchmarks	Derived from number of reports processed during the operational window.	Overall processing throughput.

Normalization:

For cost-type metrics (e.g., processing time, delays), normalization is defined as:

$$\hat{x}_{i,j} = \frac{x_{j\max} - x_{i,j}}{x_{j\max} - x_{j\min}} \tag{2}$$

For benefit-type metrics (if applicable), the standard min-max scaling is used.

Weights (example used in study):

- the processing time: 0.30,
- the product creation time: 0.25,
- the collation delays: 0.25,
- the database operations: 0.20.

Range:

$$MOP_i \in [0;1] \tag{3}$$

The peer-evaluation process involved three independent evaluators, each of whom is a former intelligence officer with more than 15 years of professional experience in analytical and operational domains. The evaluation followed a structured scoring procedure based on predefined criteria aligned with analytical standards. Each criterion was assessed using a five-point Likert scale (1 = very low, 5 = very high). Evaluators worked independently to avoid bias, and no prior consultation or calibration discussion was conducted before scoring. To ensure consistency, evaluators were provided with standardized evaluation guidelines, including definitions of each criterion and illustrative examples. After individual assessments, scores were aggregated by calculating mean values across evaluators for each observation. The level of agreement among evaluators was assessed using inter-rater reliability analysis, specifically Krippendorff's alpha, which is suitable for ordinal data and multiple raters. The resulting coefficient ($\alpha \approx 0.78$) indicates substantial agreement, suggesting a satisfactory level of consistency across evaluators. In cases of significant divergence (difference

≥ 2 points), a post hoc reconciliation step was conducted through brief discussion to clarify interpretation, although original scores were retained for analysis. Tab. 2 summarizes the qualitative instruments applied in the study.

Tab. 2 Qualitative instruments and their respective elements

Instrument	Description	Observed or Captured Elements
Semi-structured interviews	Conducted post-exercise to capture analyst experiences and perceptions.	Perceived workload; cognitive burden; confidence in assessments; satisfaction with workflow structure; perceived situational awareness.
Observation	Researchers observed staff behavior and workflow dynamics during the exercise.	User interaction patterns; bottleneck behavior; human–machine collaboration in Staff A; manual coordination behavior in Staff B.
Analytic product peer review board	Independent intelligence instructors reviewed final products without knowing which staff produced them (blind evaluation).	Product quality (clarity, accuracy, and adherence to PIRs).

MOEs reflect higher-level mission effectiveness, incorporating qualitative and outcome-oriented indicators.

For each experiment i:

$$MOE_i = \sum_{k=1}^{m} v_k \hat{y}_{i,k} \tag{4}$$

where

- $y_{i,k}$ = the value of effectiveness metric k,
- $\hat{y}_{i,k}$ = the normalized score of metric k,
- v_k = the weight of metric k, with $\sum v_k = 1$,
- m = the number of effectiveness metrics.

Metrics included:

- perceived workload,
- situational awareness (SA),
- product quality,
- bottleneck behavior.

All MOE variables are benefit-type (higher is better), normalized as:

$$\hat{y}_{i,k} = \frac{y_{i,k} - y_{k\min}}{y_{k\max} - y_{k\min}} \tag{5}$$

Weights (example used in study):

- workload: 0.25,
- SA: 0.25,
- product quality: 0.30,
- bottleneck behavior: 0.20.

Range:

$$MOE_i \in [0;1] \tag{6}$$

Aggregation and Overall Score (to combine performance and effectiveness):

$$Score_i = \alpha \cdot MOP_i + (1 - \alpha) \cdot MOE_i \quad (7)$$

where $\alpha = 0.5$ (equal importance of performance and effectiveness).

The following thresholds were applied for interpretation:

- $Score \geq 0.75$: high performance / effective system,
- $0.50 \leq Score < 0.75$: moderate performance,
- $Score < 0.50$: low performance / requires improvement.

The integration of both quantitative and qualitative evidence ensured that MOP and MOE reflected not only the procedural efficiency of each workflow but also its substantive analytic value and human factors implications.

3.3 Experimental Procedure

The experimental procedure was structured into five sequential phases, each contributing data that fed directly into the baseline used to calculate MOP and MOE. All actions performed by Staff A and Staff B were logged, observed, or recorded in a manner that enabled subsequent quantitative and qualitative analysis.

In the Scenario Deployment phase, both staffs received identical intelligence requirements and access to synthetic reporting streams, ensuring that all subsequent measurements reflected workflow differences rather than scenario variance. The controlled environment established the initial conditions for evaluating both performance and effectiveness.

The Report Injection phase introduced a high-volume, multi-format reporting stream under time-compressed conditions. This phase generated most of the raw performance data – such as ingestion times, conversion delays, and triage activity – from which the MOP baseline was constructed.

During the Analysis phase, both staffs undertook correlation, indicator identification, and interim assessment tasks. Observational data from this phase contributed to the qualitative baseline, informing later MOE calculations [13, 14].

The Production phase provided key evidence for both MOP and MOE. Product preparation times and error frequencies supported performance assessment, while the content and structure of the final analytic products formed part of the effectiveness baseline used by the independent evaluation board.

Finally, in the Dissemination phase, products were submitted under blind conditions to evaluators whose assessments directly informed the MOE baseline. Their evaluations were triangulated with observational notes and interview data to ensure consistent judgement across measures.

Together, these sequential phases produced the full dataset (quantitative and qualitative) used to generate the MOP and MOE that underpin the analytical conclusions of this study.

3.4 Validity, Reliability, and Constraints

To ensure internal validity, both staffs operated under identical scenario conditions with the same reporting inputs and time constraints. Reliability was supported through automated logging, consistent software versions, and repeatable workflows [15, 16].

A key methodological constraint relates to security and classification requirements. The experimental setup deliberately avoided integration with classified systems, focusing instead on technical feasibility and workflow impact. As such, real-

world deployment would require additional certification, accreditation, and compliance testing within classified information infrastructures.

4 Results and Discussion

The evaluation of Staff A (automation-supported) and Staff B (manual workflow) produced a set of results grounded in the quantitative and qualitative instruments described earlier. These results provide insight into how intelligent automation affects operational performance, analytical reliability, and the cognitive experience of intelligence personnel.

4.1 Quantitative Findings

Quantitative results were derived from three core instruments: time-stamped activity logs, analyst performance metrics, and product comparison rubrics. The measurement elements used included processing time, error rate, and total reports processed. Received outcomes indicate that Staff A consistently outperformed Staff B in all three categories.

The quantitative outcomes depicted in Fig. 9 represent time-based indicators only and are presented primarily as a demonstrative showcase of observed performance tendencies. These elements were selected to illustrate representative differences between the two workflows rather than to provide an exhaustive quantitative evaluation. The *x*-axis represents key workflow activities: processing time of received reports, product creation time, triage delays, and database operations, while the *y*-axis shows the corresponding time in minutes. Although automated triage (collation) and database operations are technically executed within seconds by the automation components, the time values presented in this study are expressed in minutes to reflect the human-in-the-loop validation process. The results indicate substantially reduced processing times across all observed activities for Staff A, reflecting the efficiency gains associated with intelligent automation.

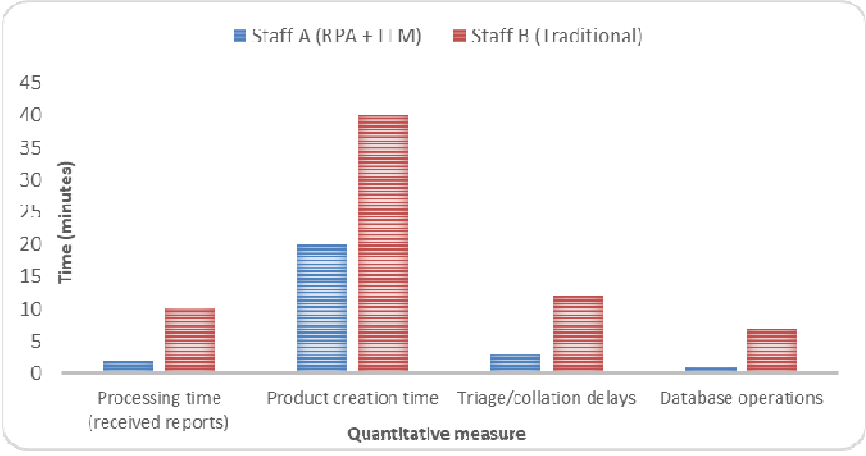


Fig. 9 Quantitative performance comparison based on observed tendencies

Tab. 3 provides the complete numerical data underlying Fig. 9. The table includes explicit sample sizes, raw counts, and all measured values for both Staff A and

Staff B. The results are based on a controlled experimental dataset comprising 20 individual observations per staff.

These findings suggest that automation improves workflow stability and reduces clerical friction, allowing analysts to move more rapidly into higher-value tasks such as interpretation and synthesis.

Normalized results in Tab. 4 indicate that, when adjusted for team size, Staff A demonstrates higher efficiency across operational metrics, particularly in processing time and database operations.

Tab. 3 Experimental Data for Staffs A/B Performance Metrics (n = 20)

Exp.	Processing time		Product creation time		Collation delays		Database operations	
#	Staff A	Staff B	Staff A	Staff B	Staff A	Staff B	Staff A	Staff B
1	1.82	7.80	21.72	37.08	3.69	14.75	0.89	6.40
2	1.70	8.63	16.60	37.42	3.65	9.27	0.93	7.52
3	2.40	7.70	17.16	44.22	2.62	11.96	1.00	5.30
4	2.44	7.84	17.34	35.31	3.37	12.45	1.18	6.30
5	2.37	9.01	22.44	33.86	2.80	13.79	1.09	7.11
6	2.08	8.81	17.35	49.67	2.84	14.15	0.78	6.68
7	2.29	7.56	24.43	33.91	2.85	11.71	0.88	6.82
8	1.75	8.55	23.29	45.58	3.09	10.21	1.11	8.59
9	1.98	9.53	20.33	30.84	2.83	13.73	1.11	7.26
10	2.18	7.68	24.69	38.34	2.51	9.67	0.83	5.89
11	1.95	7.94	18.54	34.28	3.41	13.10	1.05	5.58
12	2.19	8.50	16.33	45.06	2.49	12.92	1.25	6.35
13	2.43	8.15	23.71	35.82	2.74	12.93	1.23	5.75
14	2.30	12.17	18.45	41.33	3.54	10.7	1.18	6.03
15	1.82	10.69	21.03	31.07	2.59	10.21	0.78	5.41
16	2.08	10.79	23.87	30.53	3.41	9.56	0.95	6.05
17	1.90	9.77	20.26	30.21	2.53	11.03	0.88	5.26
18	1.94	10.10	21.60	38.55	2.36	9.30	1.04	8.29
19	2.23	10.88	21.81	31.10	3.25	10.37	1.05	8.55
20	1.77	11.80	21.50	50.83	3.14	9.67	0.83	5.75
avg	**2**	**10**	**20**	**40**	**3**	**12**	**1**	**7**

4.2 Qualitative Findings

Qualitative results were drawn from three instruments: semi-structured interviews, direct observation, and peer review of final products. Fig. 10 presents a comparison of qualitative outcomes between Staff A (using RPA and offline LLM support) and Staff B (using traditional methods), based on a ten-point evaluative scale. The x-axis displays selected qualitative indicators – perceived workload, perceived situational

awareness, product quality, and bottleneck behavior – while the *y*-axis represents subjective scores ranging from 1 to 10 (higher is better unless stated). The indicators presented in Tab. 5 are based on the experimental dataset and reflect aggregated results derived from the measured observations (n = 20); they are intended to capture representative performance patterns between Staff A and Staff B across selected metrics rather than to provide an exhaustive evaluation of all possible qualitative dimensions. The results indicate a lower perceived workload for Staff A, reflecting the benefits of automation. The higher perceived situational awareness reported by Staff B should be interpreted with caution. This outcome was partly influenced by the comparatively high level of confidence placed in AI capabilities by some members of Staff A, which in several cases led to reduced individual engagement and a degree of cognitive complacency during information monitoring. By contrast, the fully manual workflow of Staff B required continuous direct interaction with incoming information, which contributed to a stronger subjective sense of situational awareness.

Tab. 4 Normalized Performance Results (per person-hour)

Metric	Staff A (5 persons)	Staff B (7 persons)
Processing time	0.40	1.43
Product creation time	4.00	5.71
Collation delays	0.60	1.71
Database operations	0.20	1.00

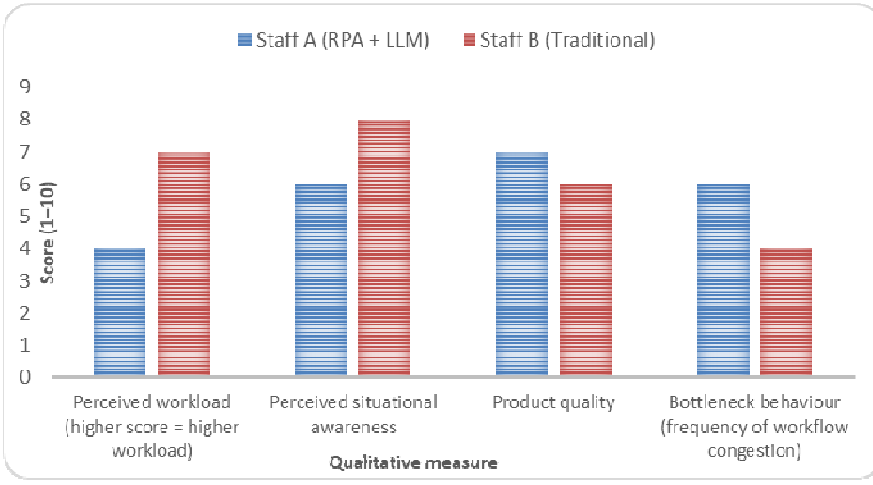


Fig. 10 Comparison of qualitative outcomes between Staff A and Staff B

The higher product quality observed for Staff A must also be understood in the context of a student-based experimental environment, where neither staff possessed extensive professional intelligence experience. In this setting, AI-supported tools enabled Staff A to approximate professional-level outputs and to mitigate errors associated with limited analytical experience and language proficiency. This effect was particularly evident in the production of English-language intelligence products, where AI assistance reduced linguistic barriers and improved structural and stylistic consistency. This effect did not eliminate the importance of expertise; rather, it highlighted that automation can serve as a knowledge amplifier, providing scaffolding that

DOI 10.3849/aimt.02077

helps novices identify relevant patterns, maintain structured reporting, and avoid common analytic pitfalls. Nevertheless, the risk remains that novices may overtrust AI outputs, making continuous training in critical evaluation essential.

Interestingly, product quality remains higher for Staff A, while situational awareness is comparable across teams after normalization computed in Tab. 6. A comparison of intel assessment examples produced by the respective staffs is depicted in Tab. 7.

Tab. 5 Experimental Data for Staffs A/B Performance Metrics (n = 20)

Exp.	Perceived workload		Perceived sit. awareness		Product quality		Bottleneck behavior	
#	Staff A	Staff B	Staff A	Staff B	Staff A	Staff B	Staff A	Staff B
1	4.08	7.69	7.24	8.76	5.17	7.01	4.28	3.94
2	2.82	7.18	7.13	7.73	6.59	5.30	4.63	4.19
3	4.64	4.99	4.63	10.2	5.19	5.94	7.03	4.44
4	4.19	5.15	4.77	9.04	6.34	5.36	6.15	3.94
5	3.86	5.13	6.54	5.76	6.44	7.26	4.88	4.54
6	3.01	5.00	6.61	7.25	5.27	6.16	6.51	3.86
7	5.17	8.77	7.39	8.37	6.88	5.67	7.29	4.28
8	3.25	5.25	5.94	8.49	8.44	5.17	6.50	2.82
9	3.03	8.19	7.23	9.88	8.59	4.81	6.74	2.96
10	3.01	5.58	5.76	6.08	9.03	4.96	6.20	4.54
11	3.56	7.45	6.87	9.36	7.61	6.80	5.04	4.26
12	4.30	8.71	7.51	7.51	6.22	4.90	7.09	3.59
13	2.95	6.09	4.31	9.50	8.73	5.60	6.76	3.89
14	4.57	5.51	5.21	5.92	7.84	6.05	6.55	5.00
15	2.82	8.72	4.65	9.64	7.48	6.09	4.92	5.20
16	4.80	4.92	4.83	8.43	6.14	5.37	6.67	2.84
17	4.61	5.34	7.42	6.98	7.60	4.59	5.43	3.68
18	3.60	7.93	4.51	7.36	5.68	7.58	5.68	3.62
19	3.83	6.40	7.59	7.75	5.23	6.28	4.74	2.91
20	3.31	5.76	5.72	10.10	6.40	7.82	7.32	5.05
avg	**4**	**7**	**6**	**8**	**7**	**6**	**6**	**4**

Tab. 6 Normalized Perception-Based Metrics (per person)

Metric	Staff A (5 persons)	Staff B (7 persons)
Perceived workload	0.80	1.00
Perceived SA	1.20	1.14
Product quality	1.40	0.86
Bottleneck behavior	1.20	0.57

Tab. 7 Comparison of intelligence assessments of an incident in the electromagnetic spectrum

Assessment – Staff A (AI powered)	**Assessment – Staff B**
The spoofing likely represents an intentional electronic warfare action by TPDM aimed at testing and exploiting vulnerabilities in our navigation infrastructure. By masking true positions, adversaries could facilitate covert movement or misdirect forces near the separation zone, potentially enabling massing of units close to the buffer line without detection. This tactic may be coordinated with other observed PMD activities such as engineering works and UAV operations noted in recent reports.	*It is likely that the ECM was being deployed in order to influence FOM of our own troops within 4BCT AOO.*
Score: 8 points	Score: 3 points

The increased occurrence of bottleneck behavior observed in Staff A should be interpreted considering the system's dependence on strict data formatting requirements. Automated processing and database operations relied on precise input structures, particularly in relation to geospatial data formats (e.g., MGRS coordinates). When inconsistencies or deviations from expected formats occurred, automated routines were unable to proceed without manual intervention. Resolving such issues required deeper technical engagement by staff members to identify, correct, and revalidate the affected data, temporarily interrupting the workflow. This sensitivity to formatting constraints increased Staff A's vulnerability to bottlenecks despite overall gains in processing efficiency.

4.3 Analytical Reliability and AI Limitations

Despite the advantages identified, several limitations emerged regarding the analytic use of AI. The offline Large Language Model occasionally produced misleading or insufficiently grounded analytical statements, even when provided with structured contextual data. This is attributable to fundamental characteristics of LLMs: they generate text by predicting statistically probable word sequences rather than understanding operational context or validating causal relationships [17]. As a result, the AI sometimes:

- inferred connections not present in the underlying reports,
- overstated the significance of ambiguous indicators,
- or inserted plausible but unsupported narrative elements.

The above-mentioned limitation affecting analytical reliability arises also from the token limitations inherent to offline Large Language Models. Unlike cloud-based models that can process extended contexts with large token windows, locally deployed LLMs operate under significantly more restrictive memory and context-length constraints due to hardware and security limitations. As a result, only a bounded subset of the available information – such as selected reports, metadata, or database excerpts – can be provided to the model at any given time. This restriction directly contributes to the model's lack of genuine situational comprehension, as it cannot simultaneously consider the full operational picture or dynamically integrate long temporal sequences

of events. Consequently, analytic outputs may reflect locally coherent interpretations while missing broader contextual dependencies, increasing the risk of incomplete or misleading assessments. This constraint reinforces the necessity of human oversight and structured information selection when employing offline LLMs in intelligence analysis.

5 Conclusion

This study examined the integration of intelligent automation – combining Robotic Process Automation (RPA), structured database support, and an offline Large Language Model (LLM) – into a military Intelligence Requirements Management and Collection Management (IRM&CM) workflow. Through a comparative case study of two intelligence staffs, the research demonstrated that automation can significantly improve procedural efficiency and product quality, while also introducing new dependencies and risks that require careful management.

Quantitative findings indicate that automation-supported workflows reduce processing time, accelerate product creation, and decrease clerical error rates. However, these improvements reflect end-to-end task durations rather than pure machine execution time. Automated triage, collation, and database operations were technically performed within seconds, but mandatory human-in-the-loop validation extended overall completion times into minutes to ensure analytical accuracy, traceability, and doctrinal compliance. This confirms that automation in intelligence serves to reallocate human effort rather than eliminate it.

Qualitative results revealed important trade-offs. While Staff A experienced reduced perceived workload and higher overall product quality, perceived situational awareness was higher among Staff B. This paradox can be attributed to the continuous manual engagement required by traditional workflows and to occasional overconfidence in AI-supported tools among Staff A participants, which reduced individual vigilance. These findings underscore the necessity of training and organizational adaptation to prevent cognitive complacency in automated environments.

Bottleneck behavior further illustrates the vulnerabilities of automation-supported workflows. Staff A proved more susceptible to workflow interruptions due to its reliance on strict data formatting requirements, particularly for structured inputs such as geospatial coordinates. When automated processes encountered malformed data, deeper technical intervention was required, temporarily offsetting efficiency gains and highlighting the importance of robust data governance.

The study also confirmed inherent limitations of AI-assisted analysis. Offline LLMs lack genuine situational comprehension and operate under constrained token limits, restricting their ability to integrate broader operational context. As a result, AI-generated outputs may appear linguistically coherent while remaining analytically incomplete. Human oversight therefore remains indispensable, and AI tools should be employed strictly as decision-support mechanisms.

Within the student-based experimental environment, intelligent automation demonstrated a positive capability-bridging effect. AI assistance enabled less experienced analysts to approximate more professional-level outputs, particularly in English-language product creation, by mitigating analytical and linguistic limitations.

In conclusion, intelligent automation offers substantial potential to enhance military intelligence workflows; however, its effective deployment depends on balancing efficiency gains with rigorous human oversight, robust data standards, and compliance

with classified information system requirements. An important finding of the experiment was the observed difference in performance perceptions between the two groups: the results indicate that Staff A reported a lower perceived workload and higher overall product quality, whereas Staff B demonstrated a higher level of perceived situational awareness. This outcome highlights the trade-offs that may emerge when integrating automated tools into analytical processes and underscores the importance of carefully evaluating both efficiency and cognitive performance indicators. Future research should therefore focus on long-term operational validation and on examining human–AI interaction under realistic operational conditions.

Acknowledgement

The work presented in this article has been supported by the Ministry of Defence of the Czech Republic (research project No. OYUOB20240002).

References

[1] *ATP 2-33.4: Intelligence Analysis* [online]. Washington: Department of the Army, 2020 [viewed 2025-11-15]. Available from: https://www.utahmilitia.org/members/lib/adp/atp2-33-4.pdf

[2] *FM 2-0: Intelligence* [online]. Washington: Department of the Army, 2023 [viewed 2025-11-15]. Available from: https://ia902808.us.archive.org/3/items/2023-2024-us-army-fm/ARN39259-FM_2-0-000-WEB-2.pdf

[3] *Intelligence, Counter-Intelligence and Security Support to Joint Operations* [online]. London: MOD, 2023 [viewed 2025-11-15]. Available from: https://assets.publishing.service.gov.uk/media/653a4b0780884d0013f71bb0/JDP_2_00_Ed_4_web.pdf

[4] *Concept of the Build-up of the Armed Forces of the Czech Republic 2035* (in Czech) [online]. 2024 [viewed 2025-11-18]. Available from: https://mocr.mo.gov.cz/assets/informacni-servis/zpravodajstvi/kvacr_2035_final.pdf

[5] *Automated Data Processing: Basics, Methods & Business Applications* [online]. 2025 [viewed 2025-11-18]. Available from: https://infomineo.com/blog/automated-data-processing-basics-methods-business-applications/

[6] FAGBENLE, E., J.Z. KHAMENEH and G. AJAYI. Reducing Administrative Burden in Primary Care Through Intelligent Workflow Automation. *International Journal of Science, Architecture, Technology and Environment*, 2024, **1**(7), pp. 164-184. DOI 10.63680/ijsate032487.012.

[7] KAHN, P., et al. Teacher Agency and Generative Artificial Intelligence: Teaching in Higher Education as a Responsive Cultural Activity. *Learning, Media and Technology* [online]. 2025 [viewed 2025-11-20]. DOI 10.1080/17439884.2025.2575993.

[8] ZHAO, X., Y. YANG, Y. NIU and X. LI. Uncertainty-Aware Multi-Modal Learning for Autonomous Pallet Handling in Dynamic Warehouse Scenarios. *IEEE Access*, 2025, **13**, pp. 168732-168742. DOI 10.1109/ACCESS.2025.3613729.

[9] GREEN, J.M. *Establishing System Measures of Effectiveness* [online]. 2001 [viewed 2026-11-23]. Available from: https://apps.dtic.mil/sti/tr/pdf/ADA405408.pdf

[10] *Intelligence, Drones, and Artificial Intelligence: The Czech Army Enters a New Era of Information* [online]. 2024 [viewed 2025-11-23]. Available from: https://www.czdefence.com/article/intelligence-drones-and-artificial-intelligence-the-czech-army-enters-a-new-era-of-information

[11] STODOLA, P. Utilisation of Artificial Intelligence in Natural Language Processing: Text Categorisation [part of the project report]. *Analysis of the Use of Artificial Intelligence in Open Source Intelligence OYUOB20240002,* 2024-2026 [Research project].

[12] HASTIE, T., R. TIBSHIRANI and J. FRIEDMAN. *The Elements of Statistical Learning*. New York: Springer, 2009. ISBN 978-0-387-84857-0.

[13] HOROWITZ, M.C. and L. KAHN. Bending the Automation Bias Curve: A Study of Human and AI-Based Decision Making in National Security Contexts. *International Studies Quarterly*, 2024, **68**(2), sqae020. DOI 10.1093/isq/sqae020.

[14] HOQUE, A., et al. AI and Machine Learning in Banking: Driving Efficiency and Innovation. *Well Testing*, 2025, **34**(S3), pp. 79-98. ISSN 1004-4388.

[15] HEINONEN, O. *Design and Implementation of a Machine Learning-Based Tool and a Robotic Process Automation Workflow for SAP S/4HANA Item Setusp* [Master's Thesis] [online]. LUT University, 2025 [viewed 2025-12-07]. Available from: https://lutpub.lut.fi/bitstream/handle/10024/170551/Diplomityo_Heinonen_Otto.pdf?sequence=1&isAllowed=y

[16] RUSSELL, S. and P. NORVIG. *Artificial Intelligence: A Modern Approach*. 4th ed. Harlow: Pearson, 2021. ISBN 0-13-461099-7.

[17] *Too Much Data, Too Few Analysts: How AI Offers a 'Force Multiplier' for Intelligence Analysts* [online]. 2025 [viewed 2025-12-07]. Available from: https://defensescoop.com/2025/11/12/too-much-data-too-few-analysts-how-ai-offers-a-force-multiplier-for-intelligence-analysts/