UDC 378.016:004
DOI https://doi.org/10.32782/2663-5941/2025.6.2/39

Palonyi A.S.
Ukrainian State Flight Academy

Zienov D.O.
Ukrainian State Flight Academy

AUTOMATED ASSESSMENT OF AIR TRAFFIC CONTROLLER TEAM COMPETENCE: AN ADAPTIVE TRAINING APPROACH

The subject matter of the article addresses the development of a formalized system of teamwork indicators for air traffic controllers (ATCOs) and a comprehensive mathematical model for automated assessment of team competence in adaptive simulator training. The goal of the article is to develop objective indicators of ATCO teamwork and an automated assessment model suitable for adaptive simulator training systems under conditions of limited real training data. The tasks of the article: to formalize measurable indicators obtainable from simulator logs, controller working positions, and communications; to develop a dual indicator system covering Safety-Performance Indicators (SPI) and Good-Practice Indicators (GPI); to design a three-level assessment architecture integrating ANFIS, Dempster-Shafer Theory, and Bayesian networks; to incorporate gradient-boosting models with SHAP interpretability; to integrate competence tracking through IRT/BKT models. The methods used: fuzzy logic and ANFIS, Dempster-Shafer Theory, Bayesian networks, gradient-boosting algorithms (XGBoost/EBM) with SHAP analysis, Item Response Theory and Bayesian Knowledge Tracing, agent-based simulation. The article presents a five-dimensional team competence model spanning coordination/communication, shared situational awareness, joint decision-making, workload management, and tactical safety management. Each component includes SPI metrics quantifying problematic behaviors and GPI metrics capturing positive practices. The assessment architecture employs ANFIS for fuzzy integration, DST for multi-source fusion, and Bayesian networks for causal risk analysis. XGBoost/EBM models with SHAP provide interpretable alternatives. The system operates through closed-loop processing: data collection, ASR/NLP preprocessing, indicator computation, multi-level assessment, and adaptive scenario adjustment. The architecture satisfies EASA transparency requirements through explicit fuzzy rules, DST conflict quantification, BBN causal traceability, and SHAP explanations.

Key words: *air traffic controllers, team competence, teamwork indicators, adaptive simulator training, ANFIS, Dempster-Shafer Theory, Bayesian networks, SHAP analysis, automated assessment.*

Formulation of the problem. The air traffic management (ATM) system is a complex socio-technical system in which technological means, regulatory procedures, and the human factor form an integrated risk environment. The modern regulatory paradigm for training and maintaining the competence of aviation personnel is defined by the transition to a competency-based training and assessment (CBTA) approach, enshrined in ICAO Doc 9868 PANS-TRG [1] and relevant EASA and EUROCONTROL materials, which emphasize the critical role of non-technical skills, including teamwork, situational awareness, and error management, in ensuring flight safety. However, the practical implementation of CBTA, despite the availability of such tools as the Non-Technical Skills (NOTECHS), Team Resource Management (TRM), The Normal Operations Safety Survey (NOSS), and Day-to-Day Safety Survey (D2D), faces challenges related to ensuring objectivity and reliability in assessing ATCO' non-technical skills, particularly their teamwork [2-4]. Adaptive learning requires objective real-time data on the current level of competence for dynamic scenario adjustment [5].

TRM programs and approaches to observations in normal operations (NOSS, D2D) were developed precisely as tools that not so much record errors but make visible what «works well» in everyday work and allow systematic strengthening of these practices. However, the absence of formalized criteria/indicators and mathematical models for automated assessment of team competence is a significant obstacle. The conditions of such a regulated industry as civil aviation, which requires high interpretability of AI system decisions, actualize the search for hybrid

© Palonyi A.S., Zienov D.O., 2025
Стаття поширюється на умовах ліцензії CC BY 4.0

models capable of combining expert knowledge from TRM/CBTA, formalized through explicit rules, with adaptation and learning capabilities on limited empirical and synthetic data. In this context, the development of a system of objective indicators of teamwork that would simultaneously reflect both problematic aspects of team interaction and positive practices based on TRM materials, with their subsequent integration into a hybrid model of automated assessment of team competence, becomes particularly important.

Analysis of recent research and publications. The Ukrainian scientific school in the field of socio-technical systems, particularly the works of Shmelova T.F. and Sikirda Yu.V., developed deterministic and stochastic models of operator decision-making in the air navigation service system, including for emergency situations and joint decision-making «crew-ATC» [6, 7]. These studies created a methodological foundation for building models oriented toward team activity.

Recent scientific research is aimed at finding objective assessment methods. Considerable attention is paid to analyzing communicative interaction. The development of automatic speech recognition (ASR) and natural language processing (NLP) technologies allows automating this process. Research shows that integrating ASR into ATM systems can improve communication accuracy and reduce workload [8]. The ATCO2 project demonstrated the possibility of collecting, transcribing, and processing more than 5,000 hours of ATC audio using robust ASR models, language identification, and contextual enhancement with radar data [9]. More recent works propose specialized ASR/NLP architectures for ATC communications, including multilingual models [10], deep integration of semantic analysis in communication quality assessment [11]. NLP models such as transformers are applied to understand the contextual content of communications and assess risks based on analysis of radio exchange transcripts [12].

Another direction is the use of psychophysiological indicators and eye-tracking data to assess cognitive load and situational awareness using machine learning [13, 14].

In the context of developing adaptive learning systems, the need arises for mathematical modeling of the competence assessment process. The application of classical deep learning methods is complicated by the absence of large labeled datasets and requirements for transparency (interpretability) of AI systems in aviation [15]. This necessitates the relevance of using hybrid approaches that combine expert knowledge and AI algorithms. Fuzzy logic allows formalizing fuzzy assessment criteria and operating with linguistic variables [16]. Bayesian Belief Networks are used to model risks and cause-and-effect relationships under uncertainty [17]. For ATCOs, the work of Du et al. is illustrative, in which an HFACS-BN model was built based on analysis of 142 ATC-related incidents, allowing quantitative assessment of the impact of various groups of human error factors on the risk of undesired events [18]. Adaptive neuro-fuzzy inference systems (ANFIS) combine the learning ability of neural networks with the interpretability of fuzzy logic, which is promising for competence assessment tasks [19]. The possibilities of generating synthetic data through simulation modeling for pre-training AI models are also being explored [20].

Task statement. This article addresses both the development a formalized system of indicators of ATCO' teamwork and a comprehensive hybrid mathematical model for automated assessment of team competence, suitable for use in adaptive simulator training systems in the absence of large arrays of real training data.

Outline of the main material of the study. The development of an automated assessment system requires a transition from abstract categories of teamwork to specific, measurable indicators that can be obtained from objective data sources in the simulator environment: simulator logs, data from controller working positions (CWP), and communication recordings. The proposed system of indicators for assessing ATCOs' team interaction skills is based on conceptual foundations that consider safety as the system's ability to work successfully under changing conditions, and not just as the absence of incidents and aviation events (unlike the classical Safety-I concept). It predominantly focuses on ensuring the maximum number of acceptable daily results, viewing people as a resource for ensuring system stability rather than as a source of errors. Based on Eurocontrol recommendations in TRM and the modern Safety-II approach in aviation, our model of team competence is built on the principles of:

– assessing the team's ability to avoid and mitigate threats (according to the ICAO TEM approach);

– analyzing the correct application by the team of established patterns and procedures for effective interaction under normal conditions during the daily work of the controller shift [4; 23];

– identifying cause-and-effect relationships between threats, intermediate failures of teamwork, and undesired events (HFACS-/BBN-oriented approaches [13]).

The team competence of ATCOs is considered as a multidimensional vector:

$$C_j=(C_1, C_2, C_3, C_4, C_5) \quad (1)$$

where each component C_j characterizes a certain aspect of the ATCO's team activity as part of the duty controller shift.

For each component, two subspaces of indicators are defined:

− **Safety-Performance Indicators (SPI) –** quantitative indicators characterizing risky or problematic aspects of ATCO behavior (delays in detecting conflicts between aircraft, untimely information transfer, lack of coordination, etc.);

− **Good-Practice Indicators (GPI)** – positive indicators that record manifestations recommended by TRM practice (short briefings during work, cross-checking, proactive workload redistribution, etc.), which are not antonyms of SPI but describe additional dimensions of team effectiveness. GPI indicators are not reduced to «absence of errors» (SPI) but record active forms of effective communication.

Analysis of recent research allowed identifying several important blocks of behavioral manifestations that found reflection in our proposed system of team indicators: 1) closed standardized communication; 2) mutual monitoring and support; 3) shared situational awareness and mental models; 4) joint decision-making and risk management; 5) joint workload management and team structure adaptation; 6) emergency situation management and team resilience.

Thus, for each component C_j, a vector is formed:

$$z_j=(S_j, g_j) \quad (2)$$

where S_j is the SPI subvector, g_j is the GPI subvector.

Below is presented a system of teamwork indicators (Tables 1-5) with emphasis on communication between the executive controller (EC) and the planning controller (PC) (an incomplete list of parameters is provided). The alignment of SPI and GPI indicators within C1–C5 provides dual coverage of both errors and violations in ATCOs' team interaction and positive practices of their teamwork.

Our proposed hybrid model for assessing ATCOs' team competence consists of three interrelated components (levels):

1. **An artificial neural network with a fuzzy inference system (ANFIS)** for integrating indicators, which forms a basic quantitative assessment of team competence TCS based on normalized SPI/GPI.

2. **A module whose functioning is built on the application of Dempster-Shafer Theory (DST)** is used to aggregate automatically obtained assessments from various sources, taking into account their reliability and contradictions. The DST module allows eliminating the influence of mutually contradictory sources, explicitly representing the «mass of ignorance» regarding the competence level, and adaptively changing the weight of channels (ASR/NLP, physiology, expert assessments) according to their reliability [21-23].

3. **A Bayesian network model of teamwork risk (BBN)**, which enables cause-and-effect analysis, linking threats, intermediate reductions in assessments across five criteria, and the probability of undesired scenario states.

Let us consider the principle of operation of the ANFIS module. Let $x=(x_1, ..., x_n)$ be a vector of nor-

Table 1

Component C1 «Coordination and Communication»

Type	Indicator Code	Description
SPI	$C1_{S1}$	Latency of critical intra-team response – average time from the moment one team member initiates a request to another's response in situations designated as critical by the scenario.
	$C1_{S2}$	Proportion of incomplete communication cycles, calculated as the percentage of cases of incomplete coordination act cycles to the total number of critical communications between the executive controller (EC) and planning controller (PC), including intra-sector coordination.
	$C1_{S3}$	Frequency of deviations from standard phraseology in intra-team communication capable of affecting the unambiguous interpretation of message content.
	$C1_{S4}$	Frequency of communication errors with aircraft pilots as the number of detected readback/hearback errors, incorrect callsigns, violations of standard radio exchange phraseology.
GPI	$C1_{G1}$	Structuredness of internal briefings, assessed by the frequency of using predetermined procedures/schemes for short briefings (particularly regarding current air situation, forecast of its development, emergence of threats, and task distribution).
	$C1_{G2}$	Presence of explicit checks of shared situation understanding as the number of cases when a controller shift team member initiates clarifying questions or formulates summarizing statements regarding the action plan.
	$C1_{G3}$	Proactive information support, determined by the frequency of preventive informing of a colleague about future changes (expected conflicts, sector overload) before their actual occurrence, without external request.
	$C1_{G4}$	Promptness of conducting a complete external coordination cycle as the average time between identifying a situation requiring coordination with another sector/ATM body and the actual completion of coordination.

Table 2

Component C2 «Shared Situational Awareness»

Type	Indicator Code	Description
SPI	$C2_{S1}$	Delay in team conflict detection relative to the moment when the conflict becomes formally detectable.
SPI	$C2_{S2}$	Divergence of mental models, assessed by the number of cases when the EC's chosen decision contradicts the PC's current plan.
SPI	$C2_{S3}$	Missed opportunities for cross-monitoring, calculated by the number of cases when one team member does not respond to an obvious discrepancy or error by another within a defined time interval.
GPI	$C2_{G1}$	Consistency in joint forecasting of possible events, determined by the frequency of coordinated predictive statements about situation development from EC and PC or both adjacent sector controllers.
GPI	$C2_{G2}$	Transparent coordinated replanning, assessed by the number of cases when a plan change (e.g., route change or approach sequence) is accompanied by explicit verbalization and confirmation from both controllers.
GPI	$C2_{G3}$	Use of information tools to support shared SA (strip markings, additional screen labels, shared notes, etc.).

Table 3

Component C3 «Team Decision-Making and Threat Management»

Type	Indicator Code	Description
SPI	$C3_{S1}$	Time for team decision from the moment of threat identification to implementation of coordinated action plan.
SPI	$C3_{S2}$	Proportion of decisions violating separation minima (regardless of severity) or procedural requirements (determined by procedural consistency index of decisions – the proportion of actions corresponding to local procedures out of total decisions made in complex/abnormal situations when alternatives existed).
SPI	$C3_{S3}$	Unilateral critical decisions without coordination, when a plan change by EC or PC is not accompanied by communication with the partner within the rule-prescribed interval.
GPI	$C3_{G1}$	Use of parametric decision-making structures (e.g., explicit articulation of options and criteria (safety/efficiency of decisions) in the context of time, fuel, workload, etc.).
GPI	$C3_{G2}$	Initiative balance maintenance, acting as an indicator of initiative decision distribution between EC and PC, indicating the absence of decision-making "monopoly" in the team.
GPI	$C3_{G3}$	Preventive threat management: number of cases when the team takes preemptive measures before threats arise in air traffic.
GPI	$C3_{G4}$	Variety of considered alternatives, assessed by the average number of explicitly discussed or intra-team processed alternative action options for the practiced class of complex situations.

Table 4

Component C4 «Workload and Resource Management»

Type	Indicator Code	Description
SPI	$C4_{S1}$	Backup behavior index, determined by the number of episodes when one team member at EC/PC level takes on routine tasks of another (answers calls, issues clearances, performs coordination) during peak workload periods.
SPI	$C4_{S2}$	Frequency of overload situations for one team member by proxy metrics (event density per time unit, subjective workload assessments).
SPI	$C4_{S3}$	Untimely task redistribution (when overload continues longer than the permissible threshold without correction).
GPI	$C4_{G1}$	Degree of proactivity in task redistribution, determined by the number of initiative function transfers (e.g., part of coordination calls or tactical decisions) before peak workload onset.
GPI	$C4_{G2}$	Level of adaptability to workload changes, involving assessment of frequency and timeliness of appropriate adaptation mechanisms (opening/closing additional workstations, task redistribution, etc.).
GPI	$C4_{G3}$	Use of "micro-breaks" and micro-briefings for synchronization during workload reduction moments.
GPI	$C4_{G4}$	Coordinated work with distributed surveillance data processing systems and automated tools (e.g., ARTAS, MTCD), when the team coordinates filter/alert configuration and responsibility distribution for their monitoring.

malized indicators (SPI and GPI). The Sugeno-type ANFIS model [19] consists of five layers:

1. ***Fuzzification.*** For each input indicator, a set of fuzzy sets is defined (for example, «low», «medium», «high»). The degree of membership is calculated using a Gaussian function:

$$O_{1,i} = \mu_{A_i}(x_j) = \exp\left(-\frac{(x_j - c_i)^2}{2\sigma_i^2}\right) \qquad (3)$$

Table 5

Component C5 «Flight Safety Management» (tactical level)

Type	Indicator Code	Description
SPI	$C5_{S1}$	Number of undetected team errors within the scenario (errors that were never identified by the team).
SPI	$C5_{S2}$	Late error detection (when correction is made after a critical deviation from norms occurs).
SPI	$C5_{S3}$	Unforced violation of ATM coordination procedures.
GPI	$C4_{G1}$	Openness in discussing own errors, assessed by the number of cases of explicit error acknowledgment, their analysis, and conclusion formation during or immediately after the simulator scenario.
GPI	$C4_{G2}$	Use of flight safety management techniques in normal working conditions during daily operations (e.g., systematic application of standard observable techniques).
GPI	$C4_{G3}$	Explicit leadership index, calculated as the proportion of episodes in which, during emergency/abnormal situations, clear verbal or action leadership is observed (forming a common plan, distributing tasks, forming conclusions) relative to the total number of such situations.
GPI	$C4_{G4}$	Adaptability in role reorganization – assessed by the number of cases when, during an abnormal situation, explicit role redistribution occurs (at the level of joint work of EC, PC, and supervisor).
GPI	$C4_{G5}$	Time spent on team recovery, determined by the time interval from the moment of initial identification of a serious problem (equipment failure, sudden weather deterioration, loss of communication, STCA alert at a dangerous stage of conflict development) to the moment of system return to a stable safe state (resolution of serious conflict, workload normalization, etc.).

where c_i and σ_i are the center and width parameters of the corresponding membership function, set based on expert assessments of instructors.

2. ***Rule application.*** Each node corresponds to a rule of the form: «If $x_1 \in A_{i1}$ and … and $x_n \in A_{in}$, Then $f_i=p_i x_1 + ... + r_i$».

Rule activation is performed according to:

$$O_{2,i} = w_i = \prod_{j=1}^{n} \mu_{A_{ij}}(x_j) \quad (4)$$

3. ***Normalization of indicators*** is performed according to the formula:

$$O_{3,i} = \bar{w}_i = \frac{w_i}{\sum_{k=1}^{R} w_k} \quad (5)$$

4. ***Consequences*** are determined according to:

$$O_{4,i} = \bar{w}_i f_i = \bar{w}_i(p_i x_1 + q_i x_2 + \dots + r_i) \quad (6)$$

5. ***At the output*** we get:

$$TCS = O_{5,1} = \sum_{i=1}^{R} \bar{w}_i f_i \quad (7)$$

where x_j is a normalized indicator (SPI/GPI); $\mu_A i(x_j)$ is the membership function; w_i is the unnormalized rule weight; $\bar{w}_i$ is the normalized weight; p_j, q_j, r_i are parameters of the linear part of the consequence; *TCS* is the integral assessment of team competence.

The consequence parameters p_i, q_i, r_i are trained by a hybrid algorithm using the least squares method with gradient descent on a combination of limited real data and simulation-synthetic datasets generated in agent-based models.

Along with ANFIS as the basic hybrid integration mechanism, our system provides for the use of an alternative module based on decision tree ensembles – XGBoost or Explainable Boosting Machine (EBM) [24]. These models provide high predictive accuracy when working with small and medium datasets and can be made transparent through the use of SHAP analysis.

The generalized gradient-boosted model has the form:

$$\widehat{TCS}_{GB}(X) = \sum_{m=1}^{M} \gamma_m h_m(X) \quad (8)$$

where $h_m(X)$ are individual decision trees; y_m are weight coefficients; $\widehat{TCS}_{GB}$ is the predicted team competence assessment.

For each simulator training session, SHAP analysis allows decomposing the prediction $\widehat{TCS}_{GB}$ into contributions of individual indicators:

$$\widehat{TCS}_{GB}(X) = \phi_0 + \sum_{i=1}^{d} \phi_i \quad (9)$$

where ϕ_0 is the base level (expected model value); ϕ_i is the SHAP value of the *i*-th indicator (SPI_k), GPI_k, individual indicators); *d* is the dimensionality of the feature vector.

Thus, the instructor can see which specific aspects of teamwork most influenced the resulting assessment of ATCOs' team competence (taking into account both negative and positive contributions).

Using XGBoost/EBM as an alternative to ANFIS is particularly appropriate at the stage when a sufficient array of data has been accumulated specifically for a particular air navigation service provider or training center, and explainability is guaranteed by SHAP analysis.

The determined quantitative assessments of team competence TCS та $\widehat{TCS}_{GB}$ and diagnostic indicators obtained from various sensors form a set of assessments $\{e_k\}$ regarding hypotheses about the level of team competence. In the DST formalism, each source specifies a mass function (m_k) on the set of hypotheses Θ. Evidence combination is implemented according to Dempster's rule:

$$m_{12}(A) = \frac{1}{1-K}\sum_{B\cap C=A} m_1(B)m_2(C),\ K = \sum_{B\cap C=\emptyset} m_1(B)m_2(C) \quad (10)$$

where m_1, m_2 are mass functions of two sources; K is the measure of conflict between them; ABC $\subseteq$ Θ.

Unlike the ANFIS module, oriented primarily toward numerical integration of indicators, and the DST module, aimed at robust fusion of multi-channel information, the Bayesian network provides a cause-and-effect level of analysis consistent with the Threat and Error Management concept and HFACS-BN approaches [18]. In its simplest form, the network includes the following groups of nodes:

− «threat» nodes (T) – abnormally high level of air traffic, non-standard airspace configuration, failures of surveillance/communication technical means, etc.;

− «state (level of formation) of team competence components» nodes (K) in the format of discretized levels based on SPI/GPI indicators;

− «intermediate undesired teamwork states» nodes (U) – «uncoordinated decision», «breakdown of shared SA», etc.;

− «consequence» node (C) – «high risk of separation minima violation or entry into active flight restriction zone», etc.

The basic Bayesian inference formula has the form:

$$P(C \mid e) = \sum_h P(C \mid h)P(h \mid e) \quad (11)$$

where *e* – is the vector of observable evidence (ANFIS/DST assessments, threats *(T),* states of team competence components), *h* is the configurations of latent nodes *(U)*.

The application of this approach allows:

− quantitatively assessing the probability of undesired states («high risk of team error»);

− identifying the most probable chains «threat – reduction in the level of a specific component of competence formation – intermediate failure – consequence»;

− linking the final indicator TCS to risk scenarios, which increases the interpretability of results for instructors and inspectors.

To describe the long-term evolution of team competence, a combination of multidimensional item-response theory (IRT) and Bayesian Knowledge Tracing (BKT) is applied [25-27]. Each simulator scenario is considered as an «item» in the task bank, which has complexity and discriminativeness parameters relative to individual components of team competence. Let us consider the two-parameter logistic IRT model:

$$P(X_{ps} = 1 \mid \theta_p) = \frac{1}{1+\exp\ (-a_s(\theta_p - b_s))} \quad (12)$$

where X_{ps} is a binary (dichotomous) random variable describing the result of performing scenario *i* by ATCO *p* («1» – «successful completion» of the scenario); θ_p is the latent level of ATCO *p's* team competence in the selected measured plane; α_s is the sensitivity of the scenario to changes in competence; b_s is the complexity of the training scenario.

To account for temporal dynamics, the BKT model is used for each component C_j:

$$P(L_t \mid Y_{1:t}) = \frac{P(Y_t|L_t)\ P(L_t|Y_{1:t-1})}{P(Y_t|Y_{1:t-1})} \quad (13)$$

where L_t is the event «team competence component is formed» at step t; Y_t is the observation (success/failure of training scenario execution, level of indicators); T is the «learning probability» parameter between steps, reflecting the learning speed; « $P(Y_t\#L_t)$ » is the probability of success given the «learned» component of competence.

IRT/BKT parameters are initialized expertly and refined based on accumulated data, and the obtained assessments θ_p and $P(L_t)$ are integrated with TCS/ $\widehat{TCS}_{GB}$, forming a complete dynamic individual profile of ATC's team competence.

The algorithm of the system's operation within a simulator session is presented in Figure 1.

Data collection involves automated accumulation of simulator logs, data from controller working positions (CWP), and audio recordings of EC-PC communications and EC-pilot communications. During preprocessing, the ASR module converts audio to text; the natural language processing (NLP) module performs markup of structural elements of phrases, identifies deviations from phraseology, and marks coordination acts. Synchronization of timestamps with trajectory data is performed. Then, indicators are calculated across the five criteria. Based on synchronized data, all positive and risk indicators are calculated, including through DST fusion of multi-channel information (in particular, combining NLP assessments and read-back error detection rules). Assessments for individual indicators are aggregated into component and global SPI, GPI indices. Competence assessment involves performing the following operations: the ANFIS module calculates TCS based on SPI and GPI, taking into account fuzzy rules; an additional XGBoost/EBM module calculates $\widehat{TCS}_{GB}$ and SHAP explanations; the BBN module calculates risk probabilities.

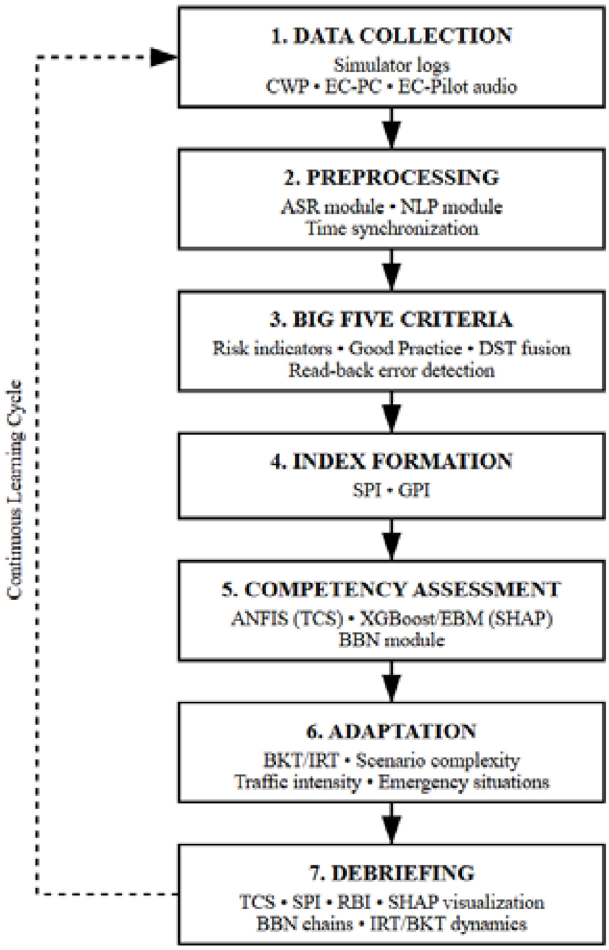

Fig. 1. Training session system algorithm

The adaptation module, relying on TCS, $\widehat{TCS}_{GB}$ and current BKT/IRT assessments, adjusts the complexity of subsequent scenarios: changes air traffic intensity, conflict configuration, introduces abnormal and emergency situations with an appropriate threat profile. The instructor receives a report that combines: numerical assessments TCS, $\widehat{TCS}_{GB}$, SPI, RBI; visualization of SHAP assessments of indicator influence; analysis of BBN risk chains; dynamics of competence according to IRT/BKT between sessions. Thus, the algorithm provides a complete cycle «data – individual team competence indicators – assessments – adaptation – model training», consistent with EASA requirements for transparency and manageability of ML solutions in critical applications [15].

Conclusions. The developed hybrid assessment architecture combines an ANFIS module for fuzzy integration of SPI/GPI indicators, a DST module for robust fusion of multi-channel data sources (ASR/NLP, physiological sensors, expert judgments), and a Bayesian network for cause-and-effect analysis of teamwork risk. Complementing this structure with XGBoost/EBM-type models with SHAP interpretation creates conditions for transparent, explainable analytics oriented toward instructor needs. The proposed combination with IRT/BKT models makes it possible to track long-term competence dynamics and form individual learning trajectories, and the adaptive scenario complexity change module ensures implementation of adaptive learning principles in the simulator environment. The obtained results create a basis for building automated instructor support systems capable of forming targeted diagnostic reports in «online/near real-time», identifying «bottlenecks» in teamwork, justifying the choice of adaptive corrective influences, and flexible revision of the structure of simulator programs.

Among the directions for further work, validation research on extended samples of simulator sessions and, if possible, on data from real operations for various air traffic service units (ATSUs) is planned, taking into account the specifics of their procedures. In addition, further research will be aimed at expanding the model toward inter-shift and inter-sectoral interaction, particularly through construction of multi-level BBN structures describing coordination between several sectors and ATSUs, as well as inclusion in the analysis of non-functional aspects (organizational culture, leadership style, features of airspace configuration changes).

Bibliography:

1. ICAO. Doc 9868. Procedures for Air Navigation Services – Training (PANS-TRG), 3d Ed. 2020. 218 p.
2. ICAO. A42-WP/179. Supporting enhancements in pilot competence through modification to ICAO competency-based training and assessment (CBTA). Working paper, 42nd session Assembly ICAO, 2025. 4 p.
3. EUROCONTROL Guidelines for TRM Good Practices. Edition 1.1. Brussels: EUROCONTROL, 2015. 38 p.
4. EUROCONTROL, SKYbrary. Ensuring Safe Performance in ATC Operations: Introducing Observational Safety Surveys (NOSS, D2D). White Paper. Brussels, 2013. 20 p.
5. Palonyi A. S., Zienov D. O. Issues in the development of an adaptive learning environment for mastering teamwork skills of air traffic controllers. *Scientific notes of Taurida National V.I. Vernadsky University. Series: Technical Sciences*. 2025. Vol. 2, no. 1. P. 175–183. https://doi.org/10.32782/2663-5941/2025.1.2/26 (date of access: 10.11.2025).

6. Шмельова Т. Ф., Сікірда Ю. В., Касаткін М. В. Детерміновані та недетерміновані моделі сумісного прийняття рішень операторами аеронавігаційної системи. *Наука і Техніка Повітряних Сил Збройних Сил України.* 2020. №1(38). С. 91–99. https://doi.org/10.30748/NITPS.2020.38.11 (date of access: 10.11.2025).

7. Shmelova T., Sikirda Y. Models of Decision-Making Operators of Socio-Technical System. *Research Anthology on Reliability and Safety in Aviation Systems, Spacecraft, and Air Transport*. 2021. P. 287–319. https://doi.org/10.4018/978-1-7998-5357-2.ch011 (date of access: 10.11.2025).

8. Wang Z, Jiang P, Wang Z, Han B, Liang H, Ai Y, Pan W. Enhancing Air Traffic Control Communication Systems with Integrated Automatic Speech Recognition: Models, Applications and Performance Evaluation. *Sensors*. 2024. Vol. 24, no. 14. https://doi.org/10.3390/s24144715 (date of access: 10.11.2025).

9. Zuluaga-Gomez J., et al. Lessons Learned in Transcribing 5000 h of Air Traffic Control Communications for Robust Automatic Speech Understanding. Aerospace, 2023, Vol. 10, No. 10. https://doi.org/10.3390/aerospace10100898 (date of access: 10.11.2025).

10. Sarhan A., Fathy R., Ali H. Intelligent air traffic control using NLP-enhanced speech recognition and natural language generation. *Journal of Electrical Systems and Information Technology.* 2025. Vol. 12, no. 1. https://doi.org/10.1186/s43067-025-00234-9 (date of access: 13.11.2025).

11. NASA. Enhancing Air Traffic Control Planning with Automatic Speech Recognition. NASA Technical Report, 2024. URL: https://ntrs.nasa.gov/citations/20240002389 (date of access: 13.11.2025).

12. Pang Y., Kendall A., Porcayo A., Barsotti M., Jain A., Clarke J-P. From Voice to Safety: Language AI Powered Pilot-ATC Communication Understanding for Airport Surface Movement Collision Risk Assessment. 2025. https://doi.org/10.48550/arXiv.2503.04974 (date of access: 15.11.2025).

13. Hauland G. Measuring Individual and Team Situation Awareness During Planning Tasks in Training of En Route Air Traffic Control. *The International Journal of Aviation Psychology.* 2008. Vol. 18, no. 3. P. 290–304. https://doi.org/10.1080/10508410802168333 (date of access: 16.11.2025).

14. Li Q, Molloy O, El-Fiqi H, Eves G. Applications of Machine Learning in Assessing Cognitive Load of Uncrewed Aerial System Operators and in Enhancing Training: A Systematic Review. *Drones*. 2025. Vol. 9, no. 11. https://doi.org/10.3390/drones9110760 (date of access: 16.11.2025).

15. EASA Concept Paper: First usable guidance for Level 1 machine learning applications – A deliverable of the EASA AI Roadmap. EASA, 2021. 145 p.

16. Jevšček M. Competencies assessment using fuzzy logic. *Journal of Universal Excellence.* 2016. Vol. 5, no. 2. P. 187–202.

17. Zhou Z., Yu X., Zhu Z., Zhou D., Qi H. Development and application of a Bayesian network-based model for systematically reducing safety risks in the commercial air transportation system. *Safety Science.* 2023. No. 157. https://doi.org/10.1016/j.ssci.2022.105942 (date of access: 17.11.2025).

18. Lyu T., Song W., Du K. Human Factors Analysis of Air Traffic Safety Based on HFACS-BN Model. Applied Sciences. 2019. Vol. 9, no. 23. https://doi.org/10.3390/app9235049 (date of access: 17.11.2025).

19. Jang J.S.R. ANFIS: Adaptive-Network-Based Fuzzy Inference System. *IEEE Transactions on Systems Man & Cybernetics*. 1993. Vol. 23. P. 665–685. https://doi.org/10.1109/21.256541 (date of access: 17.11.2025).

20. Synthetic data generation – AnyLogic Simulation Software : web-site. URL: https://www.anylogic.com/features/artificial-intelligence/synthetic-data/ (дата звернення: 17.11.2025).

21. Hamda N.E.I., Hadjali A., Lagha M. Multisensor Data Fusion in IoT Environments in Dempster–Shafer Theory Setting: An Improved Evidence Distance-Based Approach. *Sensors*. 2023. Vol. 23, no. 11. 24 p. https://doi.org/10.3390/s23115141 (date of access: 17.11.2025).

22. Xiao F. Multi-sensor data fusion based on the belief divergence measure of evidences and the belief entropy. *Information Fusion*. 2019. Vol. 46. P. 23–32. https://doi.org/10.1016/j.inffus.2018.04.003 (date of access: 17.11.2025).

23. Angeline F. K., Johnsi S. I. Multi Sensor Data Fusion and Smart Decision Making Using Dempster-Shafer Theory: A Case Study. *International Journal of Scientific Engineering and Research*. 2019. Vol. 7, no. 5. P. 27–31. https://doi.org/10.70729/17051902 (date of access: 17.11.2025).

24. Lundberg S., Lee S.-I. A Unified Approach to Interpreting Model Predictions. *In: NIPS'17*: Proceedings of the 31st International Conference on Neural Information Processing Syste. 2017. P. 4768–4777. https://doi.org/10.48550/arXiv.1705.07874 (date of access: 17.11.2025).

25. Bergner Y., Halpin P., Vie J.-J. Multidimensional Item Response Theory in the Style of Collaborative Filtering. *Psychometrika,* Vol. 87. 2022. P. 266–288. https://doi.org/10.48550/arXiv.2301.00909 (date of access: 17.11.2025).

26. Sarsa S., Leinonen J., Hellas A. Empirical Evaluation of Deep Learning Models for Knowledge Tracing: Of Hyperparameters and Metrics on Performance and Replicability. 2022. 70 p. https://doi.org/10.48550/arXiv.2112.15072 (date of access: 17.11.2025).

27. Šarić-Grgić I., Grubišić A., Gaspar A. Twenty-five years of Bayesian knowledge tracing: a systematic review. *User Modeling and User-Adapted Interaction.* 2024. No. 34. P. 1127–1173. https://doi.org/10.1007/s11257-023-09389-4 (date of access: 17.11.2025).

Пальоний А.С., Зєнов Д.О. AUTOMATED ASSESSMENT OF AIR TRAFFIC CONTROLLER TEAM COMPETENCE: AN ADAPTIVE TRAINING APPROACH

Предметом статті є розробка формалізованої системи показників командної роботи авіадиспетчерів та комплексної математичної моделі автоматизованої оцінки командної компетентності в адаптивній тренажерній підготовці. Метою статті є розробка об'єктивні показники командної роботи авіадиспетчерів та модель автоматизованої оцінки, придатну для систем адаптивної тренажерної підготовки в умовах обмежених реальних навчальних даних. Завдання статті: формалізувати вимірювані показники, що отримуються з логів тренажера, робочих місць диспетчерів та комунікацій; розробити подвійну систему індикаторів, що охоплює показники, які характеризують ризиковані або проблемні аспекти поведінки авіадиспетчера (SPI) та індикатори гарних практик (GPI); спроєктувати трирівневу архітектуру оцінювання, що інтегрує ANFIS, теорію Демпстера-Шейфера та Баєсівські мережі; включити градієнтно-підсилені моделі з інтерпретованістю SHAP; інтегрувати відстеження компетентності через моделі IRT/BKT. Використані методи: нечітка логіка та ANFIS, теорія Демпстера-Шейфера, Баєсівські мережі, градієнтно-підсилені алгоритми (XGBoost/EBM) з аналізом SHAP, теорія відгуку на завдання та Баєсівське відстеження знань. Стаття представляє п'ятивимірну модель командної компетентності, що охоплює координацію/комунікацію, спільну ситуаційну обізнаність, спільне прийняття рішень, управління навантаженням та тактичне управління безпекою. Кожен компонент включає SPI-метрики для проблемних поведінок та GPI-метрики для позитивних практик. Архітектура оцінювання використовує ANFIS для нечіткої інтеграції, DST для злиття багатоджерельних даних та Баєсівські мережі для причинного аналізу ризику. Моделі XGBoost/EBM з SHAP забезпечують інтерпретовані альтернативи. Система функціонує через замкнений цикл: збір даних, попередня обробка ASR/NLP, обчислення індикаторів, багаторівнева оцінка та адаптивне коригування сценаріїв. Архітектура задовольняє вимоги EASA щодо прозорості через явні нечіткі правила, кількісну оцінку конфліктів DST, причинну простежуваність BBN та SHAP-пояснення.

Ключові слова: *авіадиспетчери, командна компетентність, показники командної роботи, адаптивна тренажерна підготовка, ANFIS, теорія Демпстера-Шейфера, Баєсівські мережі, SHAP-аналіз, автоматизована оцінка.*

Дата надходження статті: 24.11.2025
Дата прийняття статті: 08.12.2025
Опубліковано: 30.12.2025