CEAS Aeronautical Journal
https://doi.org/10.1007/s13272-025-00870-x

ORIGINAL PAPER

CPS prototype development for AI-based scenario adaptation in flight simulator training

H. Lilla[1] · O. Niggemann[2] · T. Netzel[1]

Received: 27 February 2025 / Revised: 4 June 2025 / Accepted: 13 June 2025

Abstract
Evidence-based training as part of competency-based training and assessment confronts pilots with unexpected events in realistic scenarios in order to promote problem-solving and adaptability. Linking theory and practice is essential to promote these competencies. To achieve this, a cyber-physical system is presented that enables this through the innovative approach of "deep-linking keywords." A heuristic scoring function determines a fulfillment score for each keyword. Based on the assessment, scenario-based training is adapted, enabling necessary individualization. Compared to existing systems, the prototype generates a coherent dataset that bridges knowledge work and scenario-based training, allowing for comprehensive scenario adaptation. The cyber-physical system consists of a computer-based training system built on the Django framework, a Basic Instrument Training Device, and flight simulator software, integrated via an application programming interface. After each evidence-based training session, performance data are processed through structured analysis pipelines to extract and evaluate scenario-linked feature vectors. This enables iterative parameter optimization for adaptive scenario control. Building on the prototype and the proven effectiveness of the heuristic scoring function, a large dataset will be compiled, and the rule-based method will be replaced by machine learning to enhance safety, effectiveness, and efficiency in aviation through highly individualized training enabled by an AI-based cyber-physical system.

Keywords AI-based CPS · Flight simulation · Machine learning · Scenario-based training · CBTA · EBT

1 Introduction

Data-driven approaches are playing an increasingly important role in the analysis and diagnostics of cyber-physical systems (CPS). They enable continuous adaptation to dynamic conditions and complement classical expert systems when models are incomplete or difficult to maintain [1]. Similar needs arise in competency modeling, where the relationship between knowledge, skills, and professional actions must be captured and adapted to individual learning progress [2–4]. In pilot training, this is particularly relevant for adapting realistic scenarios to learner needs [5–7].

✉ T. Netzel
thomas.netzel@haw-hamburg.de

1 Department Automotive and Aeronautical Engineering, Hamburg University of Applied Sciences, Berliner Tor 9, 20099 Hamburg, Germany

2 Institute of Automation Technology, Helmut-Schmidt-University, Holstenhofweg 85, 22043 Hamburg, Germany

Evidence-based training (EBT), within the framework of competency-based training and assessment (CBTA), promotes adaptive competencies through scenario-based challenges [8–10]. This requires alignment of theoretical knowledge and practical execution, as enabled by work-based learning (WBL) concepts [11, 12]. Simulator-based training is central to this process [7, 13, 14], but the assessment of competencies demands a holistic view that includes declarative, procedural, strategic, and adaptive knowledge [15].

Recent AI-based systems support this through feedback models, intelligent tutoring assistants, or learning dashboards [16–18]. However, these solutions typically focus on either theoretical assessments or simulator performance. The discrepancy between cognitive and behavioral data remains unresolved. While WBL-oriented concepts exist in adjacent domains such as smart manufacturing and Industry 5.0, they lack a methodological integration of knowledge diagnostics and real-time performance data that would support reflective and adaptive learning in aviation [19].

Published online: 29 August 2025

A major obstacle is the incoherence of datasets across training phases, as highlighted by Gehr and Dunagan [20]. Without unified data structures, CPS-based adaptation remains limited. New standards are emerging to support such integration [21, 22], but their application to flight training remains underexplored.

This paper addresses these challenges by developing a CPS for pilot training that links theoretical knowledge work with scenario-based training through semantic keywords and structured performance data. The following research questions guide the development:

RQ1: How can metrics from knowledge work and simulator training be integrated into a data-driven model to capture individual learning progress?

RQ2: How can these metrics serve as a valid basis for the algorithmic adaptation of realistic simulator scenarios?

RQ3: How can an integrated model support adaptive competency diagnostics in CBTA-based training environments?

RQ4: How can a CPS be designed to enable adaptive coupling of theory and practice phases through an AI-based system?

To address these questions, we developed a CPS prototype combining a web-based training environment (Django), the X-Plane flight simulator with control hardware, and diagnostic modules for theory and practice. Competency levels are assessed through multiple-choice, calculation, and free-text tasks, and linked to scenario performance. Learner actions are mapped to domain-specific keywords via expert profiles. These are aggregated into feature vectors and used to adapt simulator parameters dynamically. Scenario evaluation, structured feedback, and threat-and-error management [5] support problem-solving and adaptability. The resulting data structure enables both formative and summative diagnostics and forms the basis for individualized, competence-driven training across all pilot development phases. To underpin the research questions and the proposed system architecture, the next section provides a structured overview of existing concepts and technologies in the domains of competency-based pilot training, artificial intelligence in assessment, and cyber-physical systems.

2 State of the art

2.1 Competency-based pilot training

2.1.1 Competency-based training and assessment

Traditional aviation training primarily focuses on obtaining and maintaining qualifications by accumulating theoretical knowledge, logging flight hours, and passing exams [9]. This qualification-based approach stands in contrast to competency-based training, which focuses on the development of skills and competencies aligned with defined performance standards [23]. This enables pilots to successfully manage situations and solve problems [8]. The paradigm shift is driven by the increasing complexity of aviation, aiming to prepare pilots for unpredictable situations [5, 6, 24] and is aligned, for example, with the "ATP(A) Integrated Course Manual" [25] and the EASA Part-ORA guidance [26]. Scenario-based training (SBT), as pat of work-based learning, embeds each learning objective within authentic flight tasks and supports contextualized competence development. WBL emphasizes the integration of theoretical and practical knowledge in real-world environments, thus providing a didactic foundation for SBT in aviation. This approach promotes the application of theoretical knowledge in complex operational situations and strengthens professional judgment, decision-making, and problem-solving skills. In aviation training, the structure of WBL aligns well with the CBTA/EBT framework and supports adaptive, task-specific learning.

2.1.2 Evidence-based training

The EBT concept adopts a data-driven approach. Based on a validated data foundation and a data-driven model [5], a systematic and competency-based training and assessment framework for pilots is implemented [8]. During training and the associated competency acquisition, pilots go through two sessions. The key phases include the evaluation (EVAL) and maneuvers training phase (MT), as well as the tailored training phase, which serves as SBT.

2.1.3 Computer-based training

Computer-based training (CBT) facilitates the computer-assisted delivery of theoretical content in a modular structure. The learning objectives (LO) conveyed within this modular framework are defined in Part-FCL [10]. The integration of LOs into CBT enables computer-based learning, allowing students to acquire content across different knowledge levels [2]. Embedding a competency model into professional actions enables the assessment of a pilot's demonstrated performance, enabling measurable competency diagnostics [8, 10, 27].

2.2 AI-based assessment

Machine learning (ML) enables data-driven analysis of learning progress and the assessment of learning trajectories [2, 17, 28–30], as shown in Fig. 1.

Within the EBT-based CPS structure, learning progression follows defined phases from knowledge diagnostics (Phase 0) to performance-based evaluation (Phase 7). As shown in Fig. 1, the effectiveness of these phases depends on timely data interpretation. Descriptive and predictive analytics in early phases (Phase 0–3) allow targeted adaptation before degradation occurs, whereas diagnostic analytics in later phases (Phase 6–7) only explain deficiencies retrospectively. This underscores the need for timely scenario adaptation between Session 1 and Session 2.

The CPS distinguishes between gain and loss of learning efficiency. Scenario control must occur before performance drops (i.e., before the event), as delayed intervention reduces training impact. This justifies integrating CBT-based knowledge diagnostics (Phase 0) prior to simulator sessions to enable prescriptive adaptation in Phase 4. By linking theoretical insights with practical performance through keyword-based KSA vectors, the CPS fosters early feedback and prevents consolidation of incorrect routines.

To meet CBTA [31] and Area 100 KSA requirements [26, 32], the prototype follows a structured cycle of formative and summative assessment. In the prototype, this is operationalized as follows: Phase 0 provides computer-based diagnostics, Phase 3 enables debriefing with root cause analysis, and Phase 7 supports scenario-based performance assessment.

Summative assessment, conducted at the end of the training cycle, evaluates final competency levels and can be analyzed using ML as part of a data-driven approach [33]. Results from all phases are aggregated into a weighted feature vector ($\vec{\alpha}_{\text{total}}$) and visualized in a session-wise tensor to track progression across scenarios.

The EBT framework ensures evidence-based scenario adjustments and alignment with ICAO and EASA guidelines

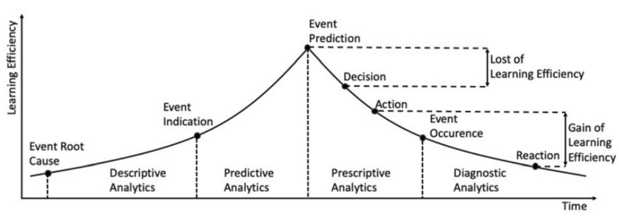

Fig. 1 Data analytics in CPS (adapted from [2]) Inaccurate adaptation of the scenario-based training session leads to a loss of learning efficiency. This can be mitigated by incorporating CBT-based assessment results into scenario design. Reflective debriefing contributes to a gain of learning efficiency by improving alignment in subsequent training sessions and by enabling scenario tuning through ML model refinement

[21, 22]. AI-based applications in aviation must comply with these standards to ensure safety and traceability.

2.3 Synthetic flight training devices

Synthetic flight training devices play a central role in aviation training and education, allowing pilots to develop competencies under safe, controllable, and reproducible conditions [5, 6, 34]. The simplest form of flight training devices (FTDs) is the basic instrument training devices (BITDs). These devices enable familiarization with instruments and cost-effective procedure training [7, 35–37]. Studies comparing the effectiveness of BITDs with real flight training [7, 13, 38] have found no significant differences in performance assessments for tasks such as the instrument proficiency check (IPC) when conducted in BITD training compared to real flight training. The combination of scenario-based simulator training and real flight training enhances the efficiency and cost-effectiveness of pilot education. The combination of scenario-based simulator training and real flight training enhances the efficiency and cost-effectiveness of pilot education and can additionally be used to generate synthetic training data for AI-based assistance systems [39]. Thus, BITDs appear to be a cost-effective and suitable means for imparting and reinforcing competencies, while also accounting for sustainability aspects [14].

2.4 Cyber-physical system

CPS for use in training and education enable innovative and practical training concepts [40, 41]. As a didactic framework, WBL allows the application of knowledge in realistic training scenarios [14], facilitating a close integration of theory and practice [11, 12]. By incorporating synthetic flight training devices, CPS provides a highly realistic flight environment, optimizing pilot training [7, 13, 14]. The collected data enable data-driven competency development to support individualized learning [2–4]. Diagnosed knowledge is directly transferred into realistic scenarios to facilitate adaptive learning [11, 12, 42] in a holistic learning experience [43]. The didactic component of the CPS, structured as CBT, utilizes linear heuristics [19], data analytics [44], AI-based systems [45], and pretrained language models such as KeyBERT for semantic keyword extraction [46], to monitor learning progress and enable individualized task adaptation [47]. A major challenge lies in data integration, as measurement values from theoretical learning and simulator training differ significantly [20, 48].

3 Prototype

The prototype, as shown in Fig. 2, is designed around the structured integration of keyword-based features into both theoretical knowledge work and flight simulator training.

Rather than aiming for full automation, the current implementation demonstrates the proof of concept of "deep-linking keywords" through a heuristic scoring function. In addition, an AI-based component is already integrated in the form of a pretrained language model (KeyBERT) to extract domain-specific keywords from free-text responses. More advanced machine learning techniques will be integrated at a later stage to support individualized scenario adaptation through keyword-based feature vectors and to enhance explainability. This architecture enables a highly individualized learning process within a work-based learning (WBL) framework. Scenario-based training (SBT) is adapted using a heuristic scoring function based on feature vectors representing knowledge, skill, and attitude (KSA). Each component is derived from task-specific keywords associated with a fulfillment score reflecting the learner's demonstrated competence. By linking theory and practice through these keyword-based features, the system promotes contextual application of knowledge, adaptive decision-making, and reflective acquisition—key elements for improving safety, effectiveness, and efficiency in pilot training.

The CPS is implemented using Django as a web framework and follows the Model-View-Template pattern to ensure modularity. A modular web interface connects the system to the X-Plane flight simulator via X-Plane Connect (XPC), enabling seamless scenario execution in a BITD.

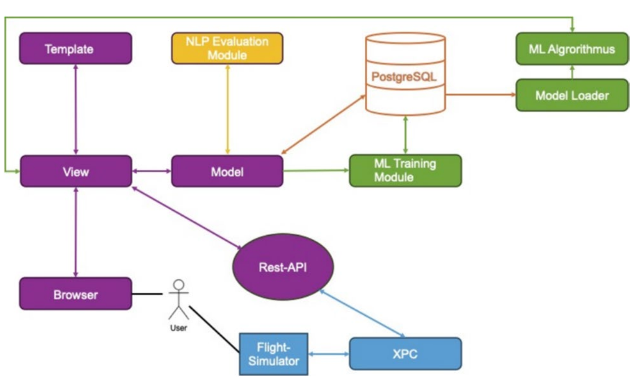

Fig. 2 System architecture of the CPS The View coordinates data exchange between the Django-based application (violet), the database (orange), ML modules (green), and the simulator interface (blue). The adapted output tensor, generated by the ML algorithm, is stored in PostgreSQL via the Django Model and View and transmitted to the simulator interface through a modular web interface (REST-API) to enable adaptive scenario control

Training data from both theoretical and simulator-based activities are stored in a PostgreSQL database and evaluated using the heuristic scoring function. At this stage, the scenario adaptation relies on expert-weighted feature combinations. Once a sufficiently large dataset has been gathered, this rule-based method will be replaced by a decision tree algorithm to enable more nuanced adaptation. The two approaches—heuristic and ML-based—will then be systematically compared, with particular attention to competence classification, feature attribution, and explainability via SHapley Additive exPlanations (SHAP).

Sections 3.2 and 3.3 detail the conceptual framework for knowledge and simulator-based training, while Sect. 3.4 outlines the current implementation status.

3.1 Storyboard: stall scenario

Using the "Stall Scenario" as a representative use case, this section illustrates how theoretical knowledge from Phase 0 (CBT) and practical experience from Phase 3 are integrated through a deep-linked feature vector to foster individualized learning and adaptive decision-making. This use case is structured in seven distinct phases according to the EBT baseline model, combining knowledge acquisition, simulator-based tasks, and reflective assessment to enhance learner competence in stall recovery—one of the most safety-critical tasks in general aviation.

Aerodynamically induced loss-of-control events caused by stall conditions remain one of the leading causes of fatal accidents in general aviation. Within the CBTA framework, the scenario is designed to foster both theoretical understanding and practical competence within a realistic training environment. The objective is to enable learners to detect a stall condition early, respond appropriately, and stabilize the aircraft.

The scenario is structured into seven distinct training phases (see Fig. 3), following the baseline session model of evidence-based training, and integrates knowledge acquisition, simulator-based practice, and reflective evaluation within a task-specific learning context.

Phase 0—knowledge acquisition (CBT) In this initial phase, learners complete a variety of theory-based tasks that target declarative, procedural, and reflective aspects of relevant knowledge. These tasks are aligned with predefined keywords (e.g., "Stall Speed") and are used to diagnose the learner's theoretical entry level. Although no feedback is provided at this stage, the results are recorded for subsequent integration into the debriefing process.

Phase 1—briefing Learners are presented with a standardized flight simulator scenario. The briefing includes aircraft

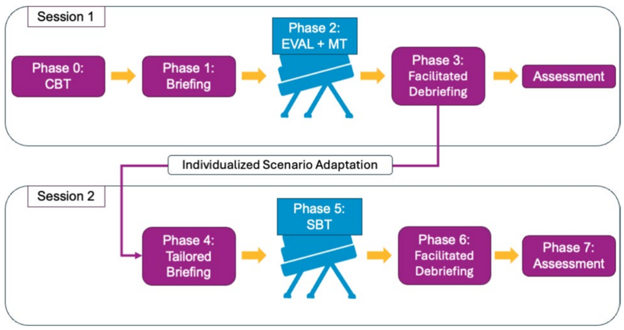

Fig. 3 **Baseline EBT FSTD session flow** Systematic structure of the EBT-based learning cycle in the CPS prototype (adapted from the IATA Evidence-Based Training Implementation Guide [6]). The diagram extends the baseline EBT model by incorporating a CBT-based knowledge phase (Phase 0). The complete sequence from Phase 0 to Phase 7, as discussed in Chapter 3, is visualized. In the CPS prototype, the assessment for Session 1 is computed algorithmically during the structured debriefing in Phase 3 by aggregating evaluation results from theoretical tasks (CBT) and simulator performance into a keyword-specific feature vector. This vector directly informs the individualized scenario adaptation in Phase 4 and the tailored SBT in Phase 5. Phases 0–1, 3–4, and 6–7 represent knowledge-related activities (violet), while Phases 2 and 5 correspond to simulator-based training (blue).

configuration, weather conditions, initial position, and learning objectives. No adaptation is applied at this stage, enabling a neutral baseline assessment of practical performance. This ensures comparability across learners.

Phase 2—evaluation and maneuvers training During the evaluation and maneuvers training (EVAL + MT), learners execute the baseline scenario to provoke a controlled stall. The system logs flight and behavioral data (e.g., airspeed, altitude loss, and heading deviation). This complements the theoretical input from Phase 0.

Phase 3—facilitated debriefing After completing the scenario, learners receive structured, keyword-based feedback that integrates their theoretical and practical performance. Each dimension—knowledge, skill, attitude—is evaluated and contextualized. The feedback highlights what was done well, where performance gaps remain, and how theory and practice were aligned (Table 1).

The learner's competence level is derived from normalized scores in the range [0–1]. The mapping to the evidence-based training standard is defined in Table 6. The knowledge fulfillment score α_k is computed using the heuristic scoring function defined in Sect. 3.2. This enables targeted feedback on theoretical strengths and weaknesses.

Phase 4—tailored briefing Based on the keyword-specific performance from the heuristic scoring function, the system generates a personalized scenario. The adapted briefing outlines the modifications in the scenario setup (e.g., reduced altitude, increased turbulence) and specifies the targeted learning objectives. Learners are encouraged to address previously identified weaknesses while consolidating their demonstrated strengths.

Phase 5—scenario-based training Learners perform the tailored scenario in the simulator. Compared to the baseline scenario, the difficulty level is adjusted to provide an appropriate challenge. Adjustments may include environmental stressors, earlier failure triggers, or modified aircraft configurations. The goal is to foster progression without overloading the learner. All performance metrics are recorded again.

Phase 6—facilitated debriefing A second structured debriefing follows the adapted simulator session. The learner's progress is assessed by comparing performance before and after adaptation. Improvements are acknowledged, persistent gaps are analyzed, and tailored recommendations are provided. This reflective phase promotes self-assessment and prepares the learner for summative evaluation or further iteration (Table 2).

The overall score α_{total} is computed as the weighted average of the normalized KSA values and serves as a summative indicator for progression assessment.

Phase 7—assessment and reflection The final phase aggregates all available performance data into a comprehensive assessment. The scores across Session 1 and Session 2 are visualized and summarized for each keyword. The resulting classification enables instructors and learners to determine whether minimum standards have been achieved or if remedial instruction is necessary. It also facilitates long-term tracking of individual learning progress (Table 3).

The overall score per keyword is derived using the heuristic scoring function described in Sect. 3.2, integrating knowledge, skill, and attitude components.

Table 1 EVAL + MT debriefing—keyword: Stall Speed

Dimension	Learning object	CBT	Session 1	Path to competence
Knowledge (K)	Stall threshold	0.89	–	Effective (4)
Skill (S)	Stall recovery	–	0.70	Minimum acceptable (2)
Attitude (A)	Recovery behavior	–	0.635	Minimum acceptable (2)

Table 2 Performance for keyword: Stall Speed

Dimension	Learning object	Target	Session 2	Path to competence
Knowledge (K)	Stall threshold	0.90	0.95	Exemplary (5)
Skill (S)	Stall recovery	0.75	0.80	Effective (4)
Attitude (A)	Recovery decision-making	0.75	0.82	Effective (4)
Overall score(α_{total})		**0.800**	**0.851**	**Effective (4)**

3.2 Knowledge work

The theoretical part of the training system is implemented as a CBT module within the CPS platform. It serves to assess the learner's knowledge in three cognitive dimensions: declarative (DK), procedural (PK), and reflective (RK) knowledge. All tasks are aligned with domain-specific keywords (e.g., "Stall Speed"), and each keyword is mapped to one or more task types.

Phase 0—CBT

Declarative knowledge (DK) These tasks use multiple-choice formats to evaluate the learner's recall of technical definitions, standard values, or regulatory requirements. An example would be: *"State the IAS at which the stall warning occurs as specified in the Pilot Operating Handbook (POH)."* The scoring is based on the ratio of correct responses to total responses:

$$\text{DK-Score} = \frac{\text{Number of correct answers}}{\text{Number of total questions}} \in [0, 1]$$

Procedural knowledge (PK) Learners solve aviation-related problems, typically requiring calculations, interpretation of diagrams, or application of operational rules. An example is: *"Calculate the stall speed of the Cessna 172 in level flight under standard atmospheric conditions at MTOW and clean configuration."* Results are scored in a binary fashion:

$$\text{PK-Score} = \begin{cases} 1 & \text{if the result is correct} \\ 0 & \text{otherwise} \end{cases}$$

Reflective knowledge (RK) Open-ended questions assess the learner's ability to transfer knowledge to new situations or explain causal relationships. An example is: *"Explain why forward stick input is essential during stall recovery."* These responses are evaluated using Natural Language Processing (NLP) techniques, particularly semantic similarity to expert-provided reference answers:

$$\text{RK-Score} = \text{sim}(a_{\text{student}}, a_{\text{expert}}) \in [0, 1]$$

The similarity function is implemented using vector-based keyword representations via a pretrained KeyBERT model, comparing the semantic proximity of learner (a_{student}) and expert (a_{expert}) answers.

Example: RK score for "Stall Speed" Prompt: Explain why forward stick input is essential during stall recovery.

Learner answer: "It helps to reduce angle of attack."

Expert answer: "Forward stick reduces angle of attack, breaking the stall and restoring lift."

NLP Similarity Score: sim = 0.78 (based on cosine similarity of KeyBERT keyword vectors).

Knowledge feature vector Each keyword k is associated with a three-dimensional vector based on the learner's scores in the cognitive dimensions:

$$\mathbf{v}_{\text{Knowledge}} = \begin{bmatrix} \text{DK} \\ \text{PK} \\ \text{RK} \end{bmatrix}$$

This vector captures the learner's theoretical competence in a structured and comparable format and forms the basis for subsequent aggregation and scenario adaptation.

Fulfillment score per Keyword The aggregated knowledge score α_k for a keyword k is calculated using a weighted linear combination of the three dimensions:

$$\alpha_{\text{Knowledge}} = w_{\text{DK}} \cdot \text{DK} + w_{\text{PK}} \cdot \text{PK} + w_{\text{RK}} \cdot \text{RK} \quad \text{with} \quad \sum w_i = 1 \tag{1}$$

Table 3 Fulfillment scores, learning gains and final assessment per scenario

Keyword	**Session 1**			**Session 2**			**Delta**			α_{total}	**Level**	**Remedial**
	α_K	α_S	α_A	α_K	α_S	α_A	Δ_K	Δ_S	Δ_A			
Stall speed	0.89	0.70	0.635	0.95	0.80	0.82	+0.06	+0.10	+0.185	0.851	3	–
Load factor	0.80	0.80	0.73	0.85	0.87	0.80	+0.05	+0.07	+0.07	0.842	3	–
Bank angle	0.86	0.76	0.70	0.90	0.82	0.78	+0.04	+0.06	+0.08	0.833	3	–

This heuristic scoring function α_k computes the overall knowledge fulfillment per keyword and enables scenario adaptation based on individual learning needs.

Weightings may be defined per keyword or scenario to emphasize specific dimensions (e.g., reflection-heavy topics). For the keyword *"Stall Speed"*, for example, higher weight may be placed on the reflective explanation of stall recovery procedures.

Example: keyword "Stall Speed" Assume the learner provides the following responses:

$$\mathbf{v}_{\text{Knowledge}} = \begin{bmatrix} \text{DK} \\ \text{PK} \\ \text{RK} \end{bmatrix} = \begin{bmatrix} 1.0 \\ 1.0 \\ 0.78 \end{bmatrix}$$

With the following weights:

$$w_{\text{DK}} = 0.2, \quad w_{\text{PK}} = 0.3, \quad w_{\text{RK}} = 0.5$$

The resulting knowledge fulfillment score is

$$\alpha_{\text{Knowledge}} = 0.2 \cdot 1.0 + 0.3 \cdot 1.0 + 0.5 \cdot 0.78 = 0.89$$

This value is used during debriefing (Phase 3) to explain how well the learner understood the topic in theory and serves as a basis for scenario adaptation in Phase 5. A low score in reflective knowledge, for instance, would trigger increased focus on the attitude dimension—since insufficient theoretical understanding of stall recovery suggests that the learner may not respond appropriately in a real-time scenario, potentially compromising safety and decision-making under stress. The CBT environment is built using Django and presented via a responsive web frontend. Each question type is mapped to a model entry, storing user answers, timestamps, and scores. The NLP pipeline runs asynchronously after submission to allow immediate retrieval of similarity-based scores for RK tasks. Final keyword vectors are stored in the PostgreSQL database and synchronized with the simulator interface via JSON during scenario preparation.

3.3 Simulator-based training

The simulator-based training process comprises Phases 1 through 6 and is structured into two iterative loops: the initial baseline scenario and the personalized, adapted scenario. In both iterations, real-time performance data from the simulator are continuously recorded and evaluated along the Skill and Attitude dimensions. These data serve as inputs to the heuristic scoring functions, support scenario adaptation, and contribute to both formative and summative assessment within the CPS training cycle.

Phase 1—briefing In this phase, the CPS initializes a standardized baseline scenario without any adaptive elements. The aircraft is positioned in a predefined flight state (e.g., 3500 ft MSL, calm weather (CAVOK), trimmed configuration), and the corresponding training task and competency objective (e.g., stall recognition and recovery) are presented to the learner. All scenario parameters remain constant and establish a reference for subsequent adaptation. No performance data are recorded during this phase; however, the simulator environment is configured via XPC commands according to the briefing parameters.

Phase 2—evaluation and maneuvers training Once the briefing is completed, the learner performs the assigned task in the simulator. During this session, key flight parameters are monitored in real time at 200 ms intervals. In this example, evaluation focuses on four core flight metrics derived from the predefined scenario:

- **IAS:** Indicated airspeed,
- **ALT:** Altitude deviation,
- **HDG:** Heading deviation,
- **TRST:** Thrust setting (0 = idle, 1 = full power).

Each parameter is continuously compared against predefined reference values. The deviation from the expected trajectory is used to assess how accurately the learner maintained the flight path. These four parameters collectively form the skill feature vector:

$$\mathbf{v}_{\text{Skill}} = \begin{bmatrix} \text{IAS} \\ \text{ALT} \\ \text{HDG} \\ \text{TRST} \end{bmatrix}$$

where each entry reflects the degree of conformity to the expected flight path on a scale from 0 (complete deviation) to 1 (perfect match), computed over the duration of the maneuver.

In addition to this flight path conformity, three continuous parameters are extracted from flight data to assess the learner's behavioral response:

Situational awareness (SA) Situational awareness is assessed based on the time delay between the stall cue and the learner's first control input (e.g., pitch or power adjustment). A delay of less than 2 s is considered ideal and yields a high SA score. Reaction times above 5 s result in a significantly reduced score. The final value is normalized between 0 and 1, reflecting the timeliness of the initial response to an unexpected event.

Action quality (AQ) AQ evaluates the learner's control response to deviations in the aircraft's flight path. Four core parameters are continuously monitored: heading (HDG), altitude (ALT), airspeed (IAS), and thrust (TRST). If a deviation exceeds a defined threshold for one of these parameters, the system checks whether a corrective control input follows within a predefined time window (e.g., 2–3 s). A reaction is counted as correct if it occurs in the expected direction:

- HDG deviation: lateral control via aileron or rudder,
- ALT deviation: thrust increase or decrease,
- IAS deviation: pitch control via elevator,
- excessive or insufficient TRST: pitch or throttle correction.

For each of these four regulation pairs, one point is assigned if both a significant deviation and an appropriate corrective action are detected. The AQ score is calculated as the ratio of successful control responses to the total number of observed deviations (maximum of four), resulting in a value between 0 and 1.

Flight stabilization (FS) Flight stabilization measures how quickly the aircraft returns to a stable flight state after the initial control input. A short stabilization time (e.g., less than 5 s) indicates good control and results in a high FS score. Longer durations reduce the score accordingly. The value is normalized to a scale from 0 (instability) to 1 (fast and stable recovery).

These parameters form the attitude feature vector:

$$\mathbf{v}_{\text{Attitude}} = \begin{bmatrix} \text{SA} \\ \text{AQ} \\ \text{FS} \end{bmatrix}$$

Phase 3—facilitated debriefing After the simulator session, a structured debriefing is conducted. From a technical perspective, the CPS merges the knowledge vector from Phase 0 (cf. 3.2), the normalized skill vector from Phase 2 (cf. 3.3), and the continuous attitude vector derived from simulator data.

To support formative assessment and instructional feedback, individual KSA fulfillment scores are computed using weighted scoring functions. These scalar values allow comparison against threshold values and support the selection of remedial or advanced scenarios in subsequent phases.

Skill fulfillment score

The aggregated skill score α_{Skill} is computed using the weighted sum of the normalized simulator parameters:

$$\alpha_{\text{Skill}} = w_{\text{IAS}} \cdot \text{IAS} + w_{\text{ALT}} \cdot \text{ALT} + w_{\text{HDG}} \cdot \text{HDG} + w_{\text{TRST}} \cdot \text{TRST} \tag{2}$$

$$\text{with} \quad \sum w_i = 1$$

Attitude fulfillment score

Likewise, the attitude score α_{Attitude} is derived as follows:

$$\alpha_{\text{Attitude}} = w_{\text{SA}} \cdot \text{SA} + w_{\text{AQ}} \cdot \text{AQ} + w_{\text{FS}} \cdot \text{FS} \tag{3}$$

$$\text{with} \quad \sum w_i = 1$$

Total fulfillment score

Finally, the total fulfillment score α_{total} is computed as the weighted sum of the knowledge, skill, and attitude scores:

$$\alpha_{\text{total}} = w_K \cdot \alpha_{\text{Knowledge}} + w_S \cdot \alpha_{\text{Skill}} + w_A \cdot \alpha_{\text{Attitude}} \tag{4}$$

$$\text{with} \quad w_K + w_S + w_A = 1$$

Example: keyword "Stall Speed" As derived in Phase 0 (cf. Sect. 3.2), the aggregated knowledge score for the keyword *Stall Speed* is

$$\alpha_{\text{Knowledge}} = 0.89$$

Assume the normalized simulator scores for the skill vector are Then:

$$\alpha_{\text{Skill}} = 0.25 \cdot (0.8 + 0.5 + 0.9 + 0.6) = 0.7$$

The simulator-derived attitude scores are

$$\text{SA} = 0.85, \quad \text{AQ} = 0.4, \quad \text{FS} = 0.9$$
$$w_{\text{SA}} = 0.3, \quad w_{\text{AQ}} = 0.5, \quad w_{\text{FS}} = 0.2$$

Then,

$$\begin{aligned} \alpha_{\text{Attitude}} &= 0.3 \cdot 0.85 + 0.5 \cdot 0.4 + 0.2 \cdot 0.9 \\ &= 0.255 + 0.20 + 0.18 = 0.635 \end{aligned}$$

Using default weights $\mathbf{w} = [0.30, \ 0.40, \ 0.30]$, the total fulfillment score at this stage is

$$\alpha_{\text{total}}^{(1)} = 0.3 \cdot 0.89 + 0.4 \cdot 0.70 + 0.3 \cdot 0.635 = 0.784$$

This result reflects that while the learner maintains good situational awareness and regains flight stability effectively, the delayed or inappropriate control response to flight path deviations (low AQ) indicates incomplete consolidation of stall recovery routines. The result corresponds with a low RK score in Phase 0 and suggests targeted reinforcement in the next tailored scenario (cf. 3.1).

This analytic structure aligns with Area 100 KSA learning objectives LO 100 02 and LO 100 03, which emphasize decision-making, threat-and-error management, upset recovery, and resilience. By quantifying knowledge, skill, and attitude fulfillment, the CPS enables fine-grained diagnostic analysis and informed instructional adaptation.

All results are stored keyword-wise in the CPS database and visualized in the user interface. These metrics form the basis for adaptive scenario configuration in Phase 4 and are later aggregated for longitudinal evaluation in Phase 7 (cf. 3.4).

Phase 4—tailored briefing In this phase, the CPS constructs a personalized scenario configuration based on the aggregated knowledge, skill, and attitude scores per keyword (cf. Phase 0-–3). This configuration is encoded in a weighted scenario feature vector (SFV):

$$\mathbf{V}_{\mathrm{SFV}} = \begin{bmatrix} w_{\mathrm{DK}} \cdot \mathrm{DK} & w_{\mathrm{PK}} \cdot \mathrm{PK} & w_{\mathrm{RK}} \cdot \mathrm{RK} \\ w_{\mathrm{IAS}} \cdot \mathrm{IAS} & w_{\mathrm{ALT}} \cdot \mathrm{ALT} & w_{\mathrm{HDG}} \cdot \mathrm{HDG} \\ w_{\mathrm{TRST}} \cdot \mathrm{TRST} & \mathrm{null} & \mathrm{null} \\ w_{\mathrm{SA}} \cdot \mathrm{SA} & w_{\mathrm{AQ}} \cdot \mathrm{AQ} & w_{\mathrm{FS}} \cdot \mathrm{FS} \end{bmatrix}$$

Each cell is computed from prior scores (cf. Phase 0-–3) and scenario-specific weightings. This matrix serves to parameterize simulator settings such as turbulence intensity, trigger altitude, scenario timing, and potential distractions.

Example: SFV for keyword "Stall Speed" Given the previously derived values (cf. 3.2 and Phase 2–3):

$$\mathrm{DK} = 1.0, \quad \mathrm{PK} = 1.0, \quad \mathrm{RK} = 0.78$$
$$\mathrm{IAS} = 0.8, \quad \mathrm{ALT} = 0.5, \quad \mathrm{HDG} = 0.9, \quad \mathrm{TRST} = 0.6$$
$$\mathrm{SA} = 0.85, \quad \mathrm{AQ} = 0.4, \quad \mathrm{FS} = 0.9$$

with

$$w_{\mathrm{DK}} = 0.2, \quad w_{\mathrm{PK}} = 0.3, \quad w_{\mathrm{RK}} = 0.5$$
$$w_{\mathrm{IAS}} = 0.25, \quad w_{\mathrm{ALT}} = 0.25$$
$$w_{\mathrm{HDG}} = 0.25, \quad w_{\mathrm{TRST}} = 0.25$$
$$w_{\mathrm{SA}} = 0.3, \quad w_{\mathrm{AQ}} = 0.5, \quad w_{\mathrm{FS}} = 0.2$$

results in

$$\mathbf{V}_{\mathrm{SFV}} = \begin{bmatrix} 0.20 & 0.30 & 0.39 \\ 0.20 & 0.125 & 0.225 \\ 0.15 & \mathrm{null} & \mathrm{null} \\ 0.255 & 0.20 & 0.18 \end{bmatrix}$$

Didactic focus of the briefing The tailored briefing highlights specific learning needs as identified in the aggregated KSA scores: while theoretical knowledge ($\alpha_{\mathrm{Knowledge}} = 0.89$) and aircraft handling ($\alpha_{\mathrm{Skill}} = 0.70$) are generally solid, the low action quality (AQ = 0.4) and suboptimal thrust adaptation suggest difficulties in dynamically managing the flight path.

To close this gap, the briefing reintroduces the aerodynamic principles behind stall recovery, emphasizes the interplay between pitch, power, and attitude correction, and configures a slightly more demanding scenario. Additional complexity (e.g., turbulence or time pressure) is introduced to assess the learner's resilience and decision-making under dynamic conditions.

The goal is to strengthen understanding and execution of stall recovery procedures in line with Area 100 KSA learning objective LO 100 03, focusing on flight path management under manual control.

Phase 5—scenario-based training In this phase, the learner completes a second simulator session under adaptive conditions. Based on the scenario feature vector (SFV) derived in Phase 4, the CPS dynamically adjusts the aircraft configuration, scenario parameters (e.g., altitude, cue timing), and environmental conditions (e.g., turbulence, distractions) to match the learner's current performance profile. The aim is not to penalize performance gaps, but to address them through simplified conditions while increasing complexity in areas of strength. For example, low heading accuracy leads to stabilized flight conditions, while high resilience introduces turbulence to evaluate recovery behavior. All parameters are automatically mapped to simulator datarefs via a plugin-based interface. Note that the skill vector in this phase reflects performance under modified conditions; thus, values may differ from those in Phase 2 even for similar control behavior, due to scenario-specific adaptation thresholds. Performance data are collected using the same telemetry structure as before, enabling direct comparison of KSA fulfillment across phases. This allows the system to evaluate whether the learner has internalized theoretical feedback and improved their performance under the adapted scenario constraints.

Skill fulfillment score

Based on updated telemetry from Phase 5, the normalized skill vector is

$$\mathbf{v}_{\mathrm{Skill}} = \begin{bmatrix} \mathrm{IAS} \\ \mathrm{ALT} \\ \mathrm{HDG} \\ \mathrm{TRST} \end{bmatrix} = \begin{bmatrix} 0.9 \\ 0.8 \\ 0.9 \\ 0.7 \end{bmatrix}$$

Assuming uniform weights:

$$w_{\mathrm{IAS}} = w_{\mathrm{ALT}} = w_{\mathrm{HDG}} = w_{\mathrm{TRST}} = 0.25,$$

we compute:

$$\alpha_{\mathrm{Skill}} = 0.25 \cdot (0.9 + 0.8 + 0.9 + 0.7) = 0.25 \cdot 3.3 = 0.825$$

Attitude fulfillment score

The attitude vector from telemetry is

$$\mathbf{v}_{\mathrm{Attitude}} = \begin{bmatrix} \mathrm{SA} \\ \mathrm{AQ} \\ \mathrm{FS} \end{bmatrix} = \begin{bmatrix} 0.90 \\ 0.85 \\ 0.70 \end{bmatrix}$$

with weightings

$$w_{\text{SA}} = 0.3, \quad w_{\text{AQ}} = 0.5, \quad w_{\text{FS}} = 0.2,$$

resulting in

$$\begin{aligned}\alpha_{\text{Attitude}} &= 0.3 \cdot 0.90 + 0.5 \cdot 0.85 + 0.2 \cdot 0.70 \\ &= 0.27 + 0.425 + 0.14 = 0.835\end{aligned}$$

These updated values reflect the learner's improved control and situational behavior under adaptive scenario conditions and serve as the basis for the final reflection in Phase 7.

Phase 6—facilitated debriefing Following the adapted scenario, a second structured debriefing is conducted. In contrast to Phase 3 (cf. 3.3), this session not only provides descriptive feedback but focuses on quantifying individual learning progress. The CPS compares the learner's performance from Phase 5 to the initial scores from Phase 2, computing delta values:

$$\Delta_{\text{Skill}} = \alpha^{(5)}_{\text{Skill}} - \alpha^{(2)}_{\text{Skill}} \tag{5}$$

$$\Delta_{\text{Attitude}} = \alpha^{(5)}_{\text{Attitude}} - \alpha^{(2)}_{\text{Attitude}} \tag{6}$$

Example: learning progress in the stall scenario Assume the following fulfillment scores were recorded for the keyword *"Stall Speed"* (Table 4):

These improvements indicate a successful transfer of theoretical and procedural knowledge into applied behavior. Specifically, the Skill delta reflects more precise aircraft handling and better coordination of pitch and thrust. The Attitude gain shows faster cue response, appropriate control input, and quicker flight stabilization—hallmarks of improved decision-making and resilience under increased workload.

This development aligns with the EBT Word Picture for *Application of Knowledge* and *Aircraft Flight Path Management, manual control*, confirming progression within the CBTA framework. No remedial action is required, and the learner proceeds to summative assessment in Phase 7.

All results are stored keyword-wise in the CPS database and contribute to the longitudinal training record.

3.4 Deep-linking keywords for CBTA

The CPS prototype enables adaptive and explainable learning by semantically linking theoretical and practical training components through keyword-based data structures. Each simulator session is associated with a domain-specific keyword (e.g., "Stall Speed") and produces standardized performance indicators across the knowledge, skill, and attitude (KSA) dimensions. These values are collected per phase, aligned with the evidence-based training (EBT) session structure, and systematically evaluated.

The resulting data are stored session-wise and keyword-specific in structured vectors and are organized over time into a session-indexed matrix. This enables the CPS to track learning progression across multiple simulator sessions, support tailored scenario adaptation, and compute individual competence levels. In particular, Phase 7 performs a summative assessment by aggregating all available KSA values and comparing them to previously recorded scores.

In addition to computing an overall fulfillment score, Phase 7 introduces reflection prompts based on observed learning deltas. These questions help the learner identify cognitive gaps, understand behavioral patterns, and reinforce the connection between theoretical understanding and practical execution. This final step ensures that each training unit concludes with a complete evaluation of demonstrated performance and self-assessed insight—aligning the CPS with CBTA standards for formative and summative competence development.

Phase 7—summative assessment and reflection Based on the deltas in $\mathbf{T}_{\text{KSA}}$, targeted prompts are generated to support causal reasoning and self-assessment, such as

- *"At what IAS did the stall warning activate in Session 2?"*
- *"Which recovery technique improved most and why?"*

These prompts close the learning loop and align with CBTA principles by supporting both summative evaluation and reflective insight (Table 5).

For the keyword *"Stall Speed"*, assume the updated scores:

$$\alpha^{(2)}_K = 0.95, \quad \alpha^{(2)}_S = 0.80, \quad \alpha^{(2)}_A = 0.82$$

Using default weights $\mathbf{w} = [0.30,\ 0.40,\ 0.30]$, the total score is

$$\alpha^{(2)}_{\text{total}} = 0.3 \cdot 0.95 + 0.4 \cdot 0.80 + 0.3 \cdot 0.82 = 0.851$$

Table 4 KSA fulfillment scores before and after adaptation

Dimension	$\alpha^{(2)}$	$\alpha^{(5)}$	Δ
Knowledge	0.89	–	–
Skill	0.70	0.825	+0.125
Attitude	0.635	0.835	+0.20

Compared to the baseline score $\alpha^{(1)}_{\text{total}} = 0.784$, this results in a learning gain:

$$\Delta_{\text{total}} = +0.067$$

This places the learner at **Competency Level 4 (Effective)** according to Table 6.

Keyword-based result tensor To support reflective and summative assessment across all targeted keywords, the system aggregates phase-specific KSA scores and associated learning deltas into a structured result tensor: (Fig. 4)

Through formative and summative assessment of the dimensions knowledge, skill, and attitude within a learning unit across multiple keywords, learning-relevant progress can be made visible, adaptive training decisions can be supported in a data-driven manner, and individual competency development can be systematically consolidated into a coherent dataset.

4 Verification

To verify the feasibility of competence-based scenario adaptation, a linear-heuristic approach was initially implemented, as outlined in Chapter 3 and Equation (4). The evaluation aims to confirm validity, reliability, and objectivity through systematic unit and integration tests conducted in a controlled development environment.

4.1 Unit testing

Unit tests focused on computational accuracy, scenario scaling (basic, moderate, complex), and robustness against boundary conditions. These were iteratively executed during development, revealing and resolving issues in parameter handling and edge case behavior. Final test runs produced consistent results, with all assertions met as expected. In addition, a dataset from twelve test flights was used to assess real-world functionality, although execution was limited to localhost conditions due to the current non-hosted status of the system. The twelve test flights were conducted by students (non-professional pilots) using example scenarios such as "Constant Descent" and "Stall Speed Recovery".

Table 5 KSA fulfillment scores before and after adaptation

Dimension	$\alpha^{(2)}$	$\alpha^{(5)}$	Δ
Knowledge	0.89	0.95	+0.06
Skill	0.70	0.80	+0.10
Attitude	0.635	0.82	+0.185

The simulator environment was operated locally with the full hardware setup, including rudder pedals, stick, and throttle quadrant.

4.2 Integration testing

Integration testing addressed the stability of communication between Django and the XPC interface, focusing on data extraction, authentication via bearer tokens, and transmission latency for adaptive scenario configuration. All functions operated reliably, with response times consistently below the defined threshold of 150 ms. Observed deviations correlated with system load but did not impair functionality.

4.3 NLP component testing

The natural language processing pipeline was evaluated separately to verify keyword extraction, conceptual mapping to predefined knowledge areas, and performance across diverse free-text responses. Tests confirmed stable detection and alignment, although database query durations increased with the number of extracted keywords. This did not affect usability but highlights potential for optimization via indexing or caching. Keyword extraction is currently implemented using `KeyBERT`, which applies a pretrained neural network to identify relevant terms in learner responses. This method constitutes the core AI-based component of the prototype.

4.4 Latency and responsiveness

Since scenario adaptation occurs prior to flight execution, average latencies of around 140 ms remained within

Table 6 Grading scheme based on α_{total}

Level	α_{total} Range	Descriptor
5	$\alpha_{\text{total}} \geq 0.90$	Exemplary
4	$0.80 \leq \alpha_{\text{total}} < 0.90$	Effective
3	$0.75 \leq \alpha_{\text{total}} < 0.80$	Adequate
2	$0.35 \leq \alpha_{\text{total}} < 0.75$	Minimum acceptable
1	$\alpha_{\text{total}} < 0.35$	Not adequate

$$\mathbf{T}_{\text{KSA}} = \begin{bmatrix} \text{Keyword} & \alpha_K & \alpha_S & \alpha_A & \Delta_K & \Delta_S & \Delta_A & \alpha_{\text{total}} \\ \text{Stall Speed} & 0.95 & 0.80 & 0.82 & +0.06 & +0.10 & +0.185 & 0.851 \\ \text{Load Factor} & 0.85 & 0.87 & 0.80 & 0.05 & 0.07 & 0.07 & 0.842 \\ \text{Bank Angle} & 0.90 & 0.82 & 0.78 & +0.04 & +0.06 & +0.08 & 0.833 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \textbf{Overall} & \mathbf{0.90} & \mathbf{0.835} & \mathbf{0.81} & \mathbf{+0.055} & \mathbf{+0.085} & \mathbf{+0.13} & \mathbf{0.8465} \end{bmatrix}$$

Fig. 4 $\mathbf{T}_{\text{KSA}}$ including domain scores, deltas, and total fulfillment level for formative and summative assessment and reflection.

acceptable bounds and did not affect simulator responsiveness or visual synchronization. Minor sampling delays (e.g., 16.67 ms at 60 Hz) and hardware-induced variations were considered negligible for pre-flight configuration.

5 Discussion

The verification results confirm the feasibility of a CPS-based training concept that systematically links theoretical and practical learning via keyword-based feature vectors. The following section discusses how the prototype addresses research questions RQ1–RQ4 and outlines implications for further development.

The knowledge levels (RQ1) are measured during the CBT phase using multiple-choice, calculation, and free-text tasks. These are aggregated into keyword-specific feature vectors and complemented by performance data from simulator sessions. The resulting fulfillment scores reflect individual competence development and are continuously updated, providing a structured basis for scenario adaptation.

The current rule-based heuristic enables these metrics to serve as a valid foundation for scenario adaptation (RQ2). Based on deviations from sample solutions and defined thresholds, the system adjusts simulator parameters to challenge learners according to their individual progress.

The integration of theoretical and simulator assessments enables a data-driven model supporting adaptive competency diagnostics (RQ3). The multidimensional tensor presented in Sect. 3.4 captures the evolution of KSA features and serves as a basis for reflective feedback and summative evaluation.

The CPS design (RQ4) enables adaptive coupling of theoretical and practical training phases, as illustrated in Fig. 2. All data streams are processed via Django and XPC, allowing diagnostic results to inform scenario configuration. In addition to structured tasks, NLP techniques evaluate free-text responses to meaningfully inform the adaptation logic.

The prototype demonstrates that deep-linking of keywords generates a coherent dataset. Each keyword is represented by a feature vector of depth n, linking knowledge, skill, and attitude (KSA) domains. In the “Stall Speed” example, $n = 9$, including three types of declarative, procedural, and reflective knowledge and six simulator indicators: IAS, ALT, HDG, SA, AQ, and FS. This structure enables individualized scenario-based training and forms a robust basis for adaptive diagnostics.

6 Conclusions and outlook

6.1 Conclusion

A CPS prototype was developed that enables scenario-based training to be adapted based on diagnostic results from both theoretical and simulator-based assessments. The core innovation lies in the deep-linking of keywords across knowledge, skill, and attitude dimensions, generating a coherent data structure that connects knowledge work and simulator performance. The integration of evidence-based training into competency-based training and assessment shows that data-driven approaches can support adaptive learning environments. The implemented system combines rule-based heuristics with feature vectors derived from structured and unstructured assessments to enable individualized training paths. The prototype thereby offers a structured foundation for the evaluation of explainable AI methods in flight training.

6.2 Outlook

Future iterations of the CPS will focus on expanding the current feature set within single-pilot training environments. In addition to the already implemented dimensions—application of knowledge, situation awareness, and decision-making—further refinement will target improved diagnostic accuracy and scenario adaptation using interpretable machine learning techniques. A key objective will be the systematic comparison of rule-based adaptation and machine learning models, such as Decision Trees combined with SHAP. This comparative approach aims to evaluate transparency, adaptability, and instructional effectiveness within the same controlled prototype architecture.

6.3 Limitations

This work validates core functionality in a restricted feature set—application of knowledge, situation awareness, and decision-making—in single-pilot scenarios. Behaviorally rich competencies such as communication, leadership and teamwork, or workload management remain outside the scope, as they require multi-crew interaction and additional sensor or voice data. Scenario adaptation is currently implemented using a heuristic scoring function. Data collection for a sufficiently large, labeled dataset is ongoing, enabling the future integration of interpretable ML models such as Decision Trees with SHAP. The prototype serves as a controlled benchmark to evaluate the potential of explainable AI (XAI) in flight training by comparing it to rule-based adaptation under realistic conditions. Feasibility was demonstrated in a set of $N = 12$ simulator sessions validating keyword-based assessment and dynamic training adaptation.

Funding Open Access funding enabled and organized by Projekt DEAL. No funds, grants, or other support was received. Open Access funding enabled and organized by Projekt DEAL. This study was funded by HAW-Hamburg.

Data availability Selected data sets generated during the current study are available from the corresponding author on reasonable request.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

References

1. Oliver, Niggemann, Biswas, Gautam Kinnebrew, John, S., Nemanja, Hranisavljevic, Andreas, Bunte: Konzeptualisierung als Kernfrage des Maschinellen lernens in der Produktion. In: Handbuch Industrie 40: Band 2: Automatisierung, pp. 631–649. Springer (2024). https://doi.org/10.1007/978-3-662-58528-3_73
2. Lilla, H., Tenberg, R.: KI- und LA-basierte Systeme in der Aus- und Weiterbildung von Pilot/-innen und luftfahrttechnischem Personal. In *Deutsche Luft- Und Raumfahrtkongress 2022*, (2023). https://doi.org/10.25967/570300
3. Farelo, David Goldsztajn, Bracaglia, Leshya, Dailey, Peter, Tamrakar, Shailesh, Palladino, Anthony, Carroll, Meredith, Valenti, Andrew: An adaptive learning model for predicting and analyzing student performance on flight training tasks. In *Signal Processing, Sensor/Information Fusion, and Target Recognition XXXI*, volume 12122, pages 278–286. SPIE, (2022). https://doi.org/10.1117/12.2619068
4. Shi Cao, Ph.D., Kearns, Suzanne K., Ph, D., Ewa Niechwiej-Szwedo, Ph.D., Elizabeth Irving, Ph.D., et al.: Computational cognitive modeling of pilot performance in pre-flight and take-off procedures. J Aviation/Aerospace Educ Res **33**(4), 2 (2024). https://doi.org/10.58940/2329-258X.2026
5. Aviation, Safety European Union.: EBT manual version **2**, 1 November (2023)
6. International Air Transport Association (IATA). *Evidence-based training implementation guide*. Montreal - Geneva, (2024)
7. Possebon, Marcello Tambosi: Aeronavegabilidade Continuada-MP-Safety. Ab-initio pilot training model combinating scenario-based training with simulator and real flight. In *XX SITRAER 2023 - Air transportation symposium*, Joinville, SC, Brazil, (2023). Universidade Federal de Santa Catarina (UFSC)
8. ICAO. *Procedures for air navigation services - training*. Third Edition. International civil aviation organization, (2020)
9. International Civil Aviation Organization: *Manual of evidence-based training*. Number 9995 in Doc / international civil aviation organization. International civil aviation organization, Montreal, 1st ed edition, (2013)
10. Agency European Aviation Safety. *Easy access rules for aircrew*. Regulation (EU) No 1178/2011. Publications Office, European Union, (2020)
11. Nixdorf, Steffen, Ansari, Fazel, Sihn, Wilfried: Work-based learning in smart manufacturing: Current state and future perspectives. In proceedings of the conference on learning factories (CLF) (2021). https://doi.org/10.2139/ssrn.3858379
12. Mubarak, Al.: Muneer: sustainably Developing in a Digital World: harnessing artificial intelligence to meet the imperatives of work-based learning in Industry 5.0. Development and Learning in Organizations:An. Int J **37**(3), 18–20 (2022). https://doi.org/10.1108/DLO-04-2022-0063
13. Taylor, Henry L., Talleur, Donald A., Emanuel Jr, Tom W., Rantanen Esa: Effectiveness of flight training devices used for instrument training. Final Technical Report AHFD-05-9/FAA-05-4, Aviation Human Factors Division, Institute of Aviation, University of Illinois at Urbana-Champaign, Savoy, Illinois, May (2005)
14. Maciejewska, Marta, Kurzawska-Pietrowicz, Paula, Galant-Gołębiewska, Marta, Gołębiewski, Michał, Jasiński, Remigiusz: Ecological and cost advantage from the implementation of flight simulation training devices for pilot training. Appl Sci **14**(18), 8401 (2024). https://doi.org/10.3390/app14188401
15. European Helicopter Safety Team (EHEST). Teaching and testing in flight simulation training devices (FSTD), (2015)
16. Guevarra, Michael, Das, Srijita, Wayllace, Christabel, Epp, Carrie Demmans, Taylor, Matthew, Tay, Alan: Augmenting flight training with AI to efficiently train pilots. Proceedings of the AAAI conference on artificial intelligence **37**, 16437–16439 (2023). https://doi.org/10.1609/aaai.v37i13.27071
17. Yang, Shuiqiao, Yu, Kun, Lammers, Thorsten, Chen, Fang: Artificial intelligence in pilot training and education–towards a machine learning aided instructor assistant for flight simulators. In *International Conference on Human-Computer Interaction*. Springer, pages 581–587 (2021). https://doi.org/10.1007/978-3-030-78642-7_78
18. Vargas, Hector, Heradio, Ruben, Chacon, Jesus, De La Torre, Luis, Farias, Gonzalo, Galan, Daniel, Dormido, Sebastian: Automated assessment and monitoring support for competency-based courses. IEEE Access: Pract Innov, Open Solutions **7**, 41043–41051 (2019). https://doi.org/10.1109/ACCESS.2019.2908160
19. Bohlouli, Mahdi, Mittas, Nikolaos, Kakarontzas, George, Theodosiou, Theodosios, Angelis, Lefteris, Fathi, Madjid: Competence assessment as an expert system for human resource management: a mathematical approach. Expert Syst Appl **70**, 83–102 (2017). https://doi.org/10.1016/j.eswa.2016.10.046
20. Gehr, L., Dunagan, L.: Learning analytics applied in commercial aviation: our use cases, obstacles, and help needed from academia. In *Companion proceedings of the 10th international learning analytics and knowledge conference (LAK'20)*, pages 9–12, Frankfurt, Germany, 2020-03-23/2020-03-27
21. Torens, Christoph, Durak, Umut, Dauer, Johann C.: Guidelines and regulatory framework for machine learning in aviation. In *AIAA Scitech 2022 Forum*, page 1132, (2022). https://doi.org/10.2514/6.2022-1132
22. European Union Aviation Safety Agency (EASA). Artificial intelligence roadmap: a human-centric approach to AI in aviation, (2020)
23. European Union Aviation Safety Agency (EASA). Flight examiners' manual (FEM), (2023)

24. European Aviation Safety Agency. Concept paper for the subtask of evidence-based training of RMT.0599: 'Evidence-based and competency-based training', (2016)
25. European Union Aviation Safety Agency (EASA). Airline transport pilot (ATP) integrated course manual, (2024)
26. European Union Aviation Safety Agency (EASA). Annex II to ED decision 2018/001/R: Acceptable means of compliance (amcs) and guidance mate rial (GM) to part-ORA — amendment 5, (2018)
27. Agency, European Aviation Safety.: Evidence-based and competency-based training. Technical report, European Aviation Safety Agency (EASA) (2024)
28. Remolina, Emilio, Ramachandran, Sowmya, Fu, D., Stottler, R., Howse, W.R.: Intelligent simulation-based tutor for flight training. In *Proceedings of the interservice/industry training, simulation, and education conference (I/ITSEC)*, pages 1–13, (2004)
29. Ziakkas, Dimitrios, Pechlivanis, Konstantinos, Keller, Julius: The implementation of artificial intelligence (AI) in aviation collegiate education: A simple to complex approach. In *Proceedings of the 6th international conference on intelligent human systems integration (IHSI 2023) Integrating people and intelligent systems*, (2023). https://doi.org/10.54941/ahfe1002863
30. Sciarrone, Filippo: Machine learning and learning analytics: integrating data with learning. In *2018 17th International conference on information technology based higher education and training (ITHET)*, pages 1–5. IEEE, (2018). https://doi.org/10.1109/ITHET.2018.8424780
31. International Air Transport Association: Competency-Based Training and Assessment (CBTA) Center Program Manual, Edition 2. International Air Transport Association, Montreal - Geneva (2022)
32. European Union Aviation Safety Agency (EASA). Easy access rules for organisation requirements for aircrew (part-ORA), (2020)
33. Uskov, Vladimir L., Bakken, Jeffrey P., Byerly, Adam, Shah, Ashok: Machine learning-based predictive analytics of student academic performance in STEM education. In *2019 IEEE global engineering education conference (EDUCON)*, pages 1370–1376. IEEE, (2019). https://doi.org/10.1109/EDUCON.2019.8725237
34. European Commission. Consolidated text: Commission Regulation (EU) No 1178/2011 of 3 November 2011 laying down technical requirements and administrative procedures related to civil aviation aircrew pursuant to Regulation (EC) No 216/2008 of the European Parliament and of the Council, (2011)
35. Rodrigues, Cristiano, Silva, Daniel Castro, Rossetti, Rosaldo J.F., Oliveira, Eugénio: Distributed flight simulation environment using flight simulator X. In *2015 10th Iberian conference on information systems and technologies (CISTI)*, pages 1–5. IEEE, (2015). https://doi.org/10.1109/CISTI.2015.7170615
36. Marques, Elvira, Carim, Guido, Campbell, Chris, Lohmann, Gui: Ab initio flight training: a systematic literature review. Int J Aerospace Psychol **33**(2), 99–119 (2023). https://doi.org/10.1080/24721840.2022.2162405
37. Cross, Jamie, Boag-Hodgson, Christine, Ryley, Tim, Mavin, Timothy J., Potter, Leigh Ellen: Using extended reality in flight simulators: a literature review. IEEE Trans Visualization Comput Graphics **29**(9), 3961–3975 (2022). https://doi.org/10.1109/TVCG.2022.3173921
38. Zhang, Tianxin: Comparing training effects of virtual reality flight simulation to conventional PC-based flight simulation. PhD thesis, Embry-Riddle Aeronautical University, Daytona Beach, Florida (2022)
39. Sprockhoff, Jasper, Gupta, Siddhartha, Durak, Umut, Krueger, Thomas: Scenario-based synthetic data generation for an AI-based system using a flight simulator. AIAA SCITECH 2024 forum (2024). https://doi.org/10.2514/6.2024-1462
40. Dimitris Mourtzis, E., Vlachou, George Dimitrakopoulos, Zogopoulos, Vasilios: Cyber-physical systems and education 4.0-the teaching factory 4.0 concept. Procedia Manufacturing **23**, 129–134 (2018). https://doi.org/10.1016/j.promfg.2018.04.005
41. Ghafoorpoor Yazdi, Poorya, Sbaragli, Andrea, Peters, Leon, Pilati, Francesco, Thiede, Sebastian: Cyber physical system for reconfigurable learning factories: Combining 3d simulations, reconfigurable layouts and real-time locating systems. In *Conference on Learning Factories*, pages 28–35. Springer, (2024). https://doi.org/10.1007/978-3-031-65411-4_4
42. Bellotti, Francesco, Berta, Riccardo, De Gloria, Alessandro, Primavera, Ludovica: Supporting authors in the development of task-based learning in serious virtual worlds. British J Educ Technol **41**(1), 86–107 (2010). https://doi.org/10.1111/j.1467-8535.2009.01039.x
43. Barth Baron Jr: The sensemaking model for airline pilot training: building and maintaining expert flight path management (2021)
44. Ang, Kenneth Li-Minn., Ge, Feng Lu, Seng, Kah Phooi: Big educational data & analytics: survey, architecture and challenges. IEEE Access: Pract Innov, Open Solutions **8**, 116392–116414 (2020). https://doi.org/10.1109/ACCESS.2020.2994561
45. Yung-Hsiang, Hu.: Effects and acceptance of precision education in an AI-supported smart learning environment. Edu Inform Technol **27**, 2013–2037 (2022). https://doi.org/10.1007/s10639-021-10664-3
46. Issa, Bayan, Jasser, Muhammed Basheer, Chua, Hui Na, Hamzah, Muzaffar: A comparative study on embedding models for keyword extraction using KeyBERT method. In *2023 IEEE 13th international conference on system engineering and technology (ICSET)*, pages 40–45, October (2023). https://doi.org/10.1109/ICSET59111.2023.10295108
47. Gambo, Yusufu: Shakir, Muhammad, Zeeshan: review on self-regulated learning in smart learning environment. Smart Learning Environ **8**(1), 12 (2021). https://doi.org/10.1186/s40561-021-00157-8
48. Jannik, Rosemeyer, Joachim, Metternich: Human-centered design for digital machine learning assistance systems in work-based learning. In: Thiede, Sebastian, Lutters, Eric (eds.) Learning Factories of the Future, pp. 155–162. Springer Nature Switzerland, Cham (2024). https://doi.org/10.1007/978-3-031-65411-4_19