

Week 1–2: Task Preparation & Evaluation Design

Thesis topic: VR + RAG + LLM tutor for DCS F/A-18C cold start

Focus of Weeks 1–2:

- Fix the **training task**
- Fix **how we measure performance and learning**

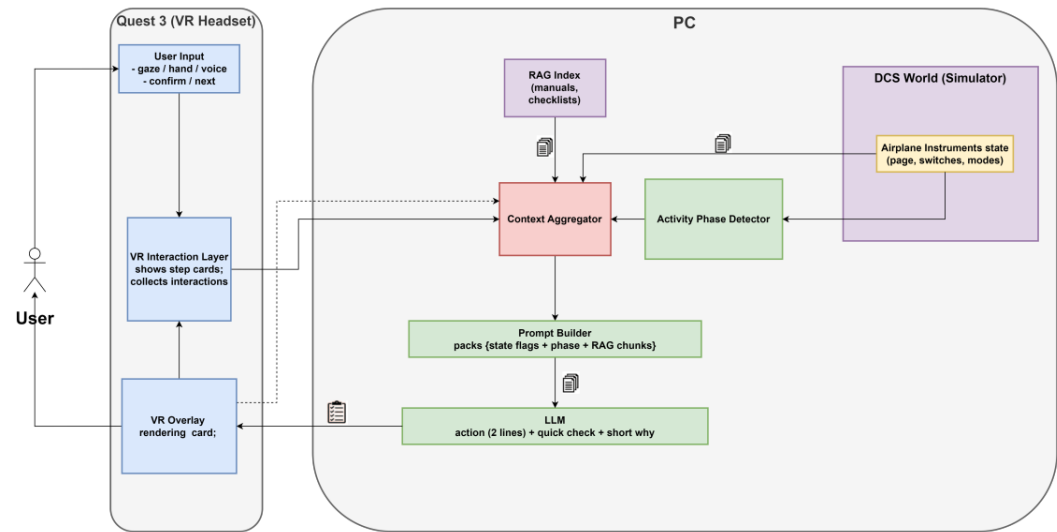


Table 1 24-week thesis work plan

Time period	Milestone	Key tasks
Weeks 1–2	Task preparation & evaluation design	Finalize task syllabus (e.g., aircraft startup); define error taxonomy and scoring sheets; collect DCS manuals and checklists; draft quizzes and NASA-TLX; run initial DCS+VR smoke tests and capture sample logs before travel
Weeks 3–4	RAG+LLM backend	Implement document ingestion and chunking for manuals/checklists; build retrieval pipeline; design prompt templates for “Action, Check, Why, Source”; test step generation outside VR (console or simple UI)
Weeks 5–6	VR display MVP & integration	Build in-VR step-card MVP on Quest 3; implement communication between Quest and PC backend; connect to RAG+LLM service; pilot the video+notes baseline and basic in-VR guidance
Weeks 7–8	State grounding	Add read-only DCS state flags (e.g., via <code>Export.lua</code>); integrate state into prompts; enforce citations; refine step-card layout and interaction details
Weeks 9–10	V1 freeze	Freeze grounded tutor V1; optimize usability/latency; lock measures and study materials
Weeks 11–12	Rehearsal	Small-cohort rehearsal; ethics/data-protection checklist; finalize protocol
Weeks 13–18	Main study	Run A/B/C study (n ≈ 5–10, beginner–intermediate); data quality checks; ablation (ungrounded vs. grounded)
Weeks 19–20	Analysis	Statistics and figures; replication package (prompts, rubrics, scripts)
Weeks 21–24	Write-up	Draft thesis; appendices; defense slides

Status & Next Steps

Weeks 1–2 deliverables (documents) are prepared in local Git repo:

- Canonical task specification (master steps + syllabus)
- Error coding & scoring framework
- Workload and knowledge instruments (NASA-TLX + quiz)

Goals for Weeks 1–2

- Finalize **task syllabus** (aircraft startup scenario)
- Define **error taxonomy** and **scoring sheets**
- Draft **knowledge quizzes** and **NASA-TLX** for workload
- Prepare for initial **DCS+VR smoke tests** and sample logging



TU Clausthal

Task file Definition

Selected **F/A-18C cold start to taxi-ready** as the core task

fa18c_startup_master.md:

- 25 startup steps (S01–S25) grouped into 6 phases (P1–P6)
- For each step: official wording, learner-friendly explanation, cockpit area

*Defines the ‘**right operation**’ and **state machine** in system*

Appendix - Training Task Syllabus.md:

- Scenario, start/end conditions, phase goals
- Success / failure criteria for a “taxi-ready” aircraft

*What exactly do we have the subjects do?
How many segments are there in this task,
and what is the objective of each segment?*

F/A-18C Cold Start – Master Step List

This file defines the canonical F/A-18C cold-start sequence used in the study. It is derived from the DCS F/A-18C manual / training mission description and is organized into phases for analysis and training.

Each row is an *instructional* step (what we track in the experiment), not necessarily a one-to-one mapping to all real-world checklist items.

Columns:

- **StepID**: Stable identifier used in scoring sheets and logs.
- **Phase**: Coarse-grained phase of the startup procedure.
- **OfficialStep**: Wording close to the DCS manual / in-game tutorial.
- **ShortExplanation**: Simplified phrasing for learners and the tutor UI.
- **CockpitArea**: Primary cockpit region where the action happens.

Master Step Table

StepID	Phase	OfficialStep (DCS-style wording)	ShortExplanation (for learners/tutor)	CockpitArea
S01	P1 – Power-up & safety	Set the BATTERY switch to ON and confirm both Left and Right Generators are ON.	Turn on aircraft battery and verify both generators are on to power the jet.	Right console
S02	P1 – Power-up & safety	Move and hold the fire detection switch to FIRE TEST A and wait for all the audio caution messages to play. Once complete, wait 10 seconds and then do the same for FIRE TEST B. Between running FIRE TEST A and FIRE TEST B, you can reset the battery switch to rewind the fire test tape.	Run fire detection tests A and B and listen for all caution tones to confirm the fire warning system works.	Left console
S03	P1 – Power-up & safety	APU switch to ON and wait for green APU READY light.	Start the APU and wait until the APU READY light comes on.	Left console

Task Syllabus: F/A-18C Cold Start to Taxi-Ready

This syllabus specifies the procedural training task used in the study: a cold start of the F/A-18C from a “cold and dark” parking position to a taxi-ready configuration in DCS World (VR).

The canonical step sequence and detailed step wording are defined in `fa18c_startup_master.md`. This syllabus focuses on task scope, phases, and success criteria for experimental evaluation.

1. Scenario Definition

- **Platform**: DCS World, F/A-18C Lot 20, VR mode (Meta Quest 3).
- **Starting condition**: Aircraft parked at an airfield ramp or carrier deck in a cold-and-dark state (engines off, no electrical power, parking brake set).
- **End condition (“taxi-ready”)**:
 - Both engines running and stabilized in the green range.
 - Electrical and bleed air systems configured for normal operation.
 - Core avionics and INS alignment initialized for taxi and departure.
 - Flight control system (FCS) checks completed, takeoff trim set.
 - Key instruments configured (altimeters, attitude references, BINGO fuel).
 - Parking brake released only when the instructor/system permits taxi.

Learners are instructed to follow the standard F/A-18C cold-start procedure as defined in the DCS manual, with minor simplifications to ensure consistent measurement across participants.

2. Phase Structure

For analysis and guidance, the 25 steps in `fa18c_startup_master.md` are grouped into six phases. Each phase has a clear functional goal and observable completion conditions (DCS state + cockpit configuration).

Phase P1 – Power-Up & Safety (S01–S03)

fa18c_error_coding_guide.md

Defined a procedural error taxonomy:

- Omission, Commission, Order error, Parameter error, State violation
- Dead-end and Help-request at trial level

Detailed Explanation of Error Classification and Scoring Rules

fa18c_scoring_sheet_template.md:

- Trial-level summary (TotalErrorScore, Count_OM/OR/SV, time, etc.)
- Step-level coding (StepID, Phase, Error_*, StepErrorScore)

2. Error Taxonomy

Each observed deviation from the canonical procedure is coded into one or more of the following categories. Errors are always coded with respect to a **target step** (StepID) and the corresponding **phase**.

For each category we give a short definition and one or more examples from the F/A-18C cold-start task.

2.1 Omission (OM)

Definition: A required step in the canonical procedure is not performed at all by the end of the trial.

- **Code:** OM
- **Example (S02 – Fire test):**
 - The learner never runs FIRE TEST A or B before proceeding to engine start.
- **Example (S14 – OBOGS ON):**
 - The learner never turns on OBOGS and FLOW, even though both engines are running and they taxi away.

Notes:

- If a step is performed only **partially** or in an obviously wrong way, code additional categories (e.g., PA Parameter error) on the same step.

1. Trial-Level Summary

ParticipantID	Condition	TrialID	Completed	DeadEnd_DE	HelpRequest_HR	TaskTime_sec
P01	grounded_rag_llm	T01	yes	no	no	1234
P02	video_notes	T01	no	yes	yes	1500
...

Field hints

- ParticipantID: e.g., P01, P02, ...
- Condition: e.g., video_notes, llm_ungrounded, grounded_rag_llm
- TrialID: e.g., T01 (if each participant runs multiple trials)
- Completed: yes / no (taxi-ready criteria met?)
- DeadEnd_DE: yes / no (coded from step-level + observation)
- HelpRequest_HR: yes / no
- TaskTime_sec: time from first cockpit interaction to end condition
- TotalErrorScore: sum of all StepErrorScore for the trial
- CriticalErrorScore: sum of StepErrorScore on steps with Critical = yes
- NonCriticalErrorScore: sum of StepErrorScore on steps with Critical = no
- Count_OM, Count_CO, Count_OR, Count_PA, Count_SV: total counts per error category
- Comments: free-text notes for anomalies, interruptions, etc.

Subjective Workload & Knowledge Measures

fa18c_nasatlx_vr.md:

- **Raw NASA-TLX**
adapted to “F/A-18C cold start in DCS VR”
- 6 scales (Mental, Physical, Temporal, Performance, Effort, Frustration, 0–20)

fa18c_coldstart_quiz.md:

- 12-item quiz bank for **pre / post / retention**
- Focus on **sequence understanding** and **state preconditions**
- Examples: when to advance throttle, when to set INS, what FCS checks include

NASA Raw-TLX – Subjective Workload

Task: F/A-18C Cold Start in DCS World (VR)

Please answer the following questions about **the cold start procedure you just performed** in DCS World (F/A-18C, VR). There are no right or wrong answers – we are interested in your **personal experience**.

For each scale, mark a value from **0 to 20**, where:

- 0 = Very low / very good
- 10 = Medium
- 20 = Very high / very poor

You can also use numbers in between (e.g., 7, 13, ...).

1. Mental Demand

How mentally demanding was the task?

Consider how much thinking, remembering, deciding, and concentrating was required to complete the cold start procedure in VR.

- 0 = Very low mental demand (almost automatic, no thinking needed)
- 10 = Moderate mental demand
- 20 = Very high mental demand (very complex, required intense concentration)

Rating (0–20): _____

Status & Next Steps

Next steps (Weeks 3–4):

- Implement **RAG+LLM backend** using the master step list
- Use the scoring and quiz templates as the basis for the main study protocol