

Development roadmap

This page is the compact strategic view of thesis development. It is omitted from submission output. Scientific claims remain owned by the active thesis; executable behavior and measurements remain owned by source, tests, configuration, and immutable evidence artifacts.

Author review and professor feedback

Author commitments agreed on 16 October 2026:

Target	Required outcome
26 Oct 2026	Aim to deliver Chapters 1-3 progressively by 26 October; the minimum packet is a newly human-reviewed Introduction and complete Foundations chapter.
30 Nov 2026	All manuscript content, including figures, captions, appendices and declarations, has been human reviewed and revised; the exact candidate is suitable for submission. Existing experimental and release gates still require their own evidence.
12 Dec 2026	Request final professor feedback and review by 12 December; this is the author's planning target, not a confirmed supervisor commitment.

Table 1: Human review and feedback targets; dates do not certify completed review or empirical evidence.

The experimental schedule below was last reviewed on 2 October and is not revalidated by this author-review plan. Data-generation and training-run scheduling is deferred until LRZ throughput is established.

Implementation: partial

Evidence: pending

Reviewed 2 Oct 2026 refresh due 16 Oct 2026. Target: reproducible full-draft freeze by 30 November 2026, followed by a safety buffer until submission on 7 January 2027.

Promotion gate: Close P1 before confirmatory policy claims.

Development source: docs/typst/thesis/development/roadmap.toml; active thesis, package, test, configuration, and evidence owners

TL;DR

Done

Pilot infrastructure

Candidate-benchmark smoke evidence, a bounded two-scene orbit MVP, and a one-scene, five-step pilot rollout report are committed. The October review retains P0 as complete for pilot infrastructure only; these artifacts do not establish population-scale support or policy gains.

Now

Evidence substrate

Freeze the multi-scene store, scene-level split, camera/depth and target-RRI fidelity, candidate-support admission, and throughput receipts while the thesis narrative and conceptual figures remain under review. The separate research publication PR #524 records a larger development collection; completion reconciliation, portable evidence verification and deep cohort audit remain pending before explicit admission. P1 remains in progress; subsequent milestone dates are planning targets conditional on its closure.

Blocked

Confirmatory claims

Lookahead-headroom, learned-value, and population-coverage claims remain blocked until paired multi-scene evidence and an immutable experiment registry exist. The results chapter still gates confirmatory claims, and its default report is a development fixture; neither that fixture nor the separate pilot rollout report discharges P1-P3.

Next

Paired evidence to freeze

Run equal-budget baselines, train and evaluate the finite-horizon value model under actor-visible inputs, synthesize results and limitations, and reach the reproducible full-draft freeze recorded below.

Critical path

The strategic dependency is evidence-first: close the substrate, establish paired baseline outcomes, evaluate the learned value model, synthesize the claims, and freeze the full draft. The December-to-January interval is a safety buffer, not planned scientific work. This is the scheduled order, not a scientific prerequisite chain: held-out Q1 prediction and matched representation comparisons can proceed without paired policy endpoints. Endpoint evidence is required for a policy-improvement claim; oracle headroom remains a separate optional diagnostic.

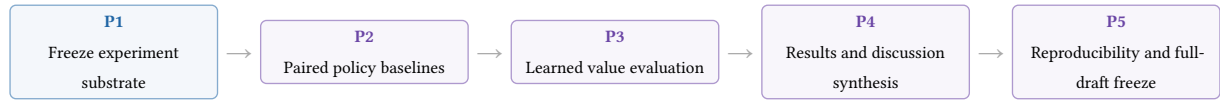


Figure 1: Scheduled development gates through the 30 November 2026 full-draft freeze. Arrows show planned order; policy endpoints and oracle headroom are not prerequisites for Q1 prediction.

Milestones

ID	Phase	Dates	State	Exit gate
P0	Pilot infrastructure	7 Jul–30 Aug 2026	done	Committed smoke and MVP artifacts compile and remain explicitly bounded as pilot evidence.
P1	Freeze experiment substrate	31 Aug–2 Oct 2026	in progress	Multi-scene immutable store, scene-level split, camera/depth and target-RRI fidelity, candidate-support admission, and throughput receipts.
P2	Paired policy baselines	3 Oct–20 Oct 2026	planned	Equal-budget random-valid, greedy, and bounded-lookahead comparisons report uncertainty, support, and oracle receipts.
P3	Learned value evaluation	21 Oct–10 Nov 2026	planned	Held-out ranking, calibration, and paired policy value are reported under actor-visible inputs.
P4	Results and discussion synthesis	11 Nov–22 Nov 2026	planned	Final tables, failure cases, coverage limits, and claim dispositions resolve to the immutable evidence registry.
P5	Reproducibility and full-draft freeze	23 Nov–30 Nov 2026	planned	Full draft, reproducible configurations, final figures, and release checks pass with no unresolved placeholders.
P6	Safety buffer and submission	1 Dec–7 Jan 2027	safety buffer	Advisor feedback, formal checks, final upload, and the authenticated submission receipt are closed without reopening the scientific core.

Table 2: Evidence-gated schedule from the completed pilot infrastructure to submission. Dependencies are checked against the adjacent roadmap data.

Blockers

- **B1 — Multi-scene experiment substrate** (affects P1): The committed artifacts do not yet establish a fresh multi-scene store, scene-level split, camera/depth fidelity, target-RRI robustness, and representative throughput in one immutable receipt chain.
- **B2 — Candidate-support scale admission** (affects P1, P2): The positive two-scene orbit MVP is explicitly not scale-up admission; family survival, framing, diversity, and proposal regret still require broader evidence.
- **B3 — Confirmatory experiment registry** (affects P2, P3, P4): The default report bundle is a development fixture; the separate pilot report covers one scene. Neither establishes confirmatory comparisons, which need immutable run/config identities, coverage accounting, uncertainty, and claim-to-artifact resolution.

The primary evidence pointers remain beside each record in `roadmap.toml`; the roadmap contract checks their existence, while scientific review decides whether they are sufficient for promotion. Bounded null and negative findings are reportable under the same matched-evidence conditions as positive findings; positive oracle headroom is not a prerequisite for learned-policy comparison.

Promotion queue

Promotion queue — candidate: Promote bounded-lookahead and learned-value results only after paired held-out evidence is immutable.

Source: P2-P3 experiment registry and rollout reports target: Results and Discussion gate: matched held-out comparisons, actor-visible evaluation, uncertainty, and support coverage; oracle headroom is a separate diagnostic and is not required to report a null, negative, or positive policy difference

Promotion queue — blocked: Resolve candidate-support and substrate limits before claiming population coverage.

Source: P1 evidence receipts and candidate benchmark manifests target: Experimental Design and Results gate: multi-scene support, family survival, scene-level split, and representative throughput

Promotion queue — deferred: Keep online discrete and continuous target-then-pose extensions outside the core until the offline finite-candidate gates close.

Source: Discussion and roadmap P1-P3 target: Discussion gate: stable offline learned-value evidence and an explicit post-core scope decision

Maintenance contract

`roadmap.toml` is the single owner for the snapshot, dates, states, dependencies, blockers, release checks, and evidence pointers rendered above. Update it at least every 14 days and whenever a gate changes state. `make thesis-roadmap-contract` rejects stale review dates, broken dependencies, missing local evidence pointers, divergence from the thesis submission date, or references to the retired M1 snapshot. Internal trackers and hosted issue state may inform a refresh but are not imported as public thesis truth.

The native table, cards, and critical-path strip are deliberately sufficient: they show exact milestones and the consequential dependency chain without a second diagram data model. A date-dense Gantt can be reconsidered only if this compact projection no longer answers the planning question.

HM/FK07 release gate

These checks are human-owned and apply to the exact final candidate and the author's authenticated records; the repository cannot satisfy them by itself.

- **R1 — Applicable rules:** Confirm the individually assigned programme and SPO version in PRIMUSS; public reading versions establish only the current baseline.
- **R2 — Registration record:** Retain the PRIMUSS confirmation, official start date, individual deadline, assigned examiners, supervisor-specific form requirements, and requested attachments.
- **R3 — Exact upload set:** Before the individual deadline, verify the exact final PDF and every PRIMUSS-marked mandatory document or attachment. A repository build, digest, or upload alone is not submission evidence.
- **R4 — Conditional paper copy:** Provide an identical paper copy only if the examiner or authenticated PRIMUSS record requires one under the current FK07 process.
- **R5 — Final-PDF review:** Check the declaration and AI/tool disclosure in the exact PDF prepared for upload; confirm any mandatory signature or attachment against the authenticated record.
- **R6 — Submission closure:** Finalize with the PRIMUSS action "Abschlussarbeit abgeben", retain the authenticated timestamp, and archive the exact submitted PDF; only that receipt closes the gate.

Official baseline checked 31 August 2026: HM examination regulations, FK07 thesis process, PRIMUSS registration guide, PRIMUSS submission guide. At the full-draft freeze, recheck the official versions and the individually assigned instructions in authenticated PRIMUSS records.

Freeze and submission

By 30 November 2026, the full draft, experiment registry, configurations, figures, and release checks must be reproducible and free of unresolved placeholders. The interval from 1 December to 7 January 2027 is reserved for advisor feedback, formal HM/FK07 checks, final upload, and recovery from unexpected defects. Only the authenticated submission receipt closes the final gate.