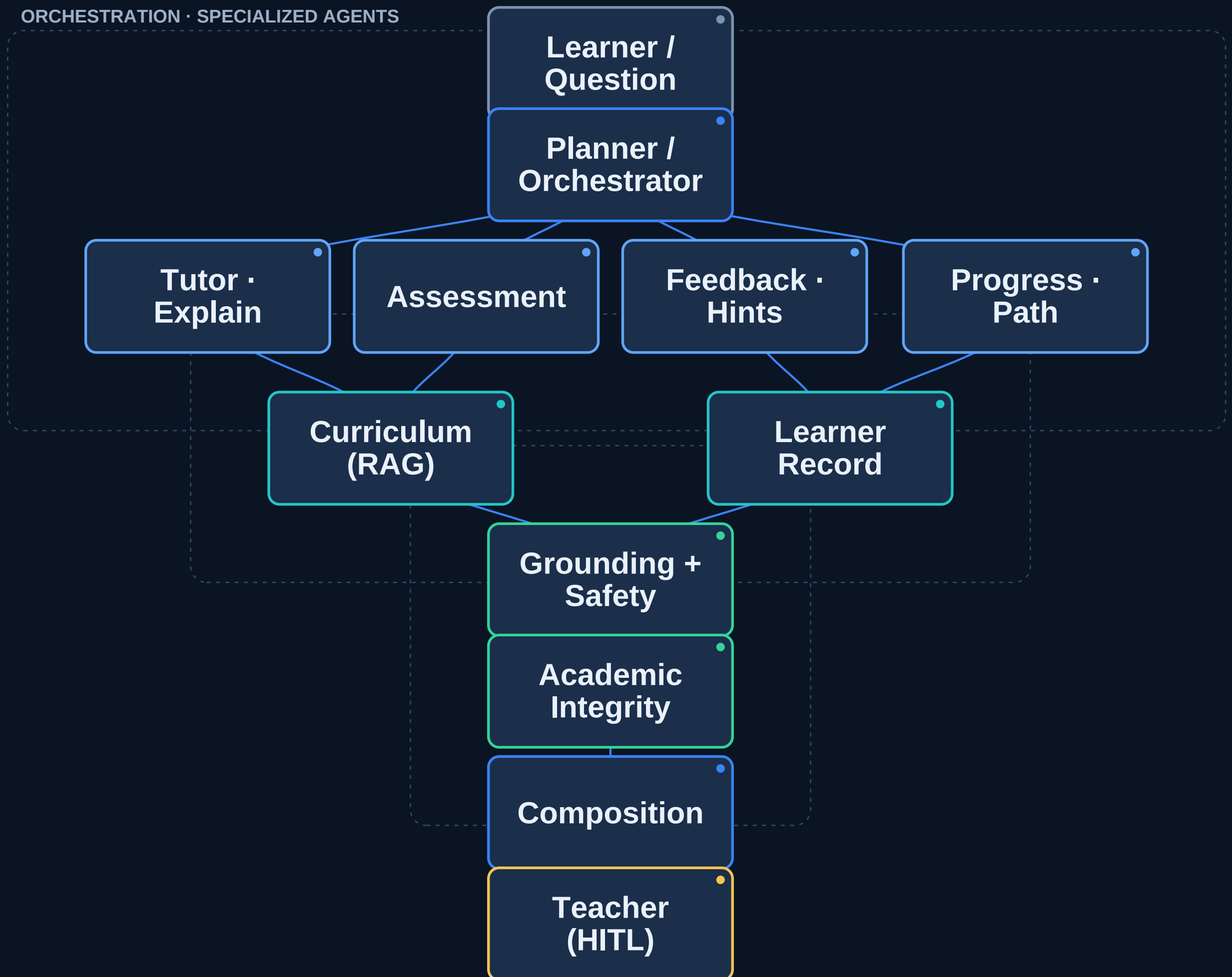


— REFERENCE ARCHITECTURE

Trustworthy Education AI

A reference design for agentic AI tutors — answers grounded in approved curriculum, safe for students, with a teacher in the loop.

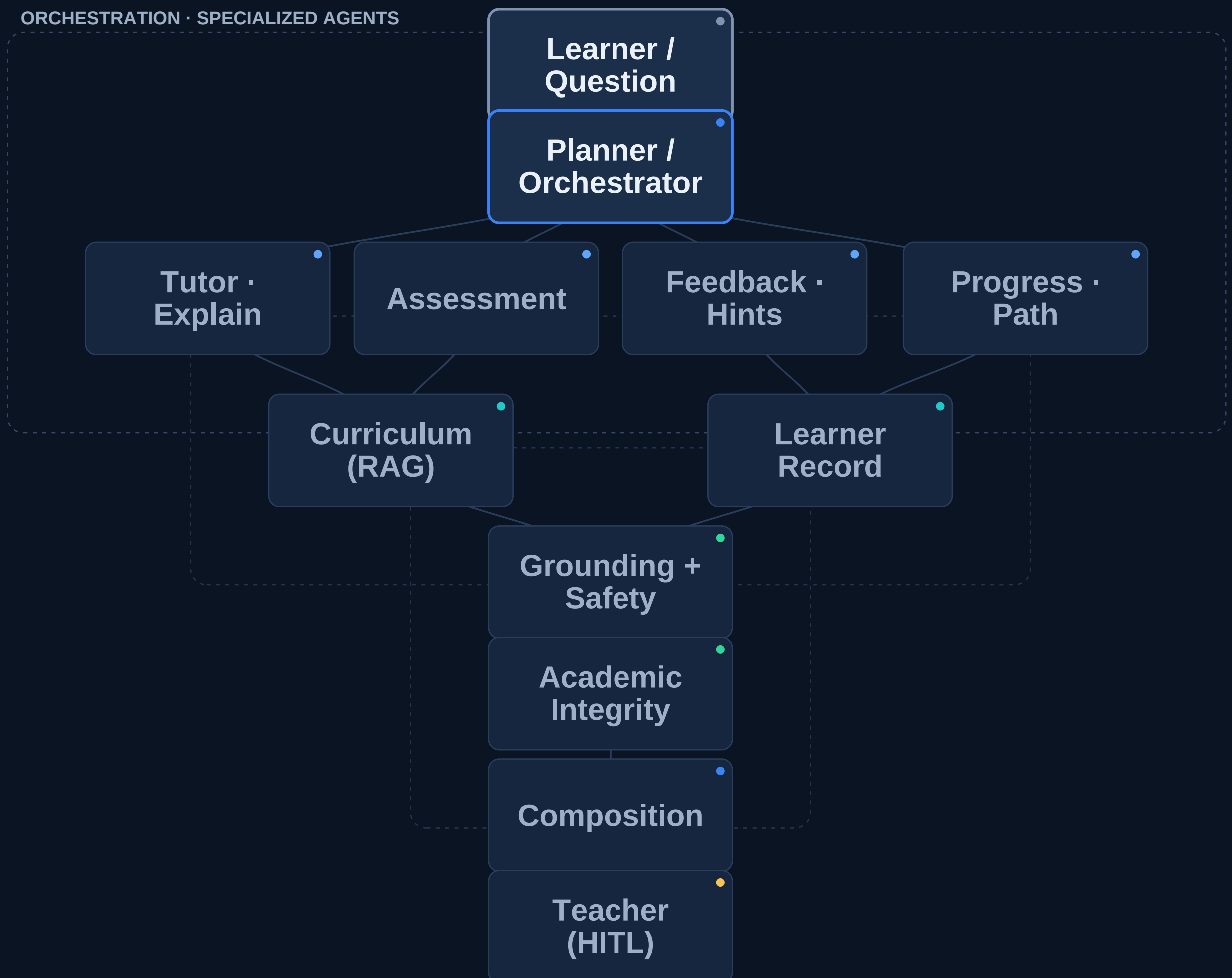
Swipe through the pipeline →



Grounded. Safe. Accountable.

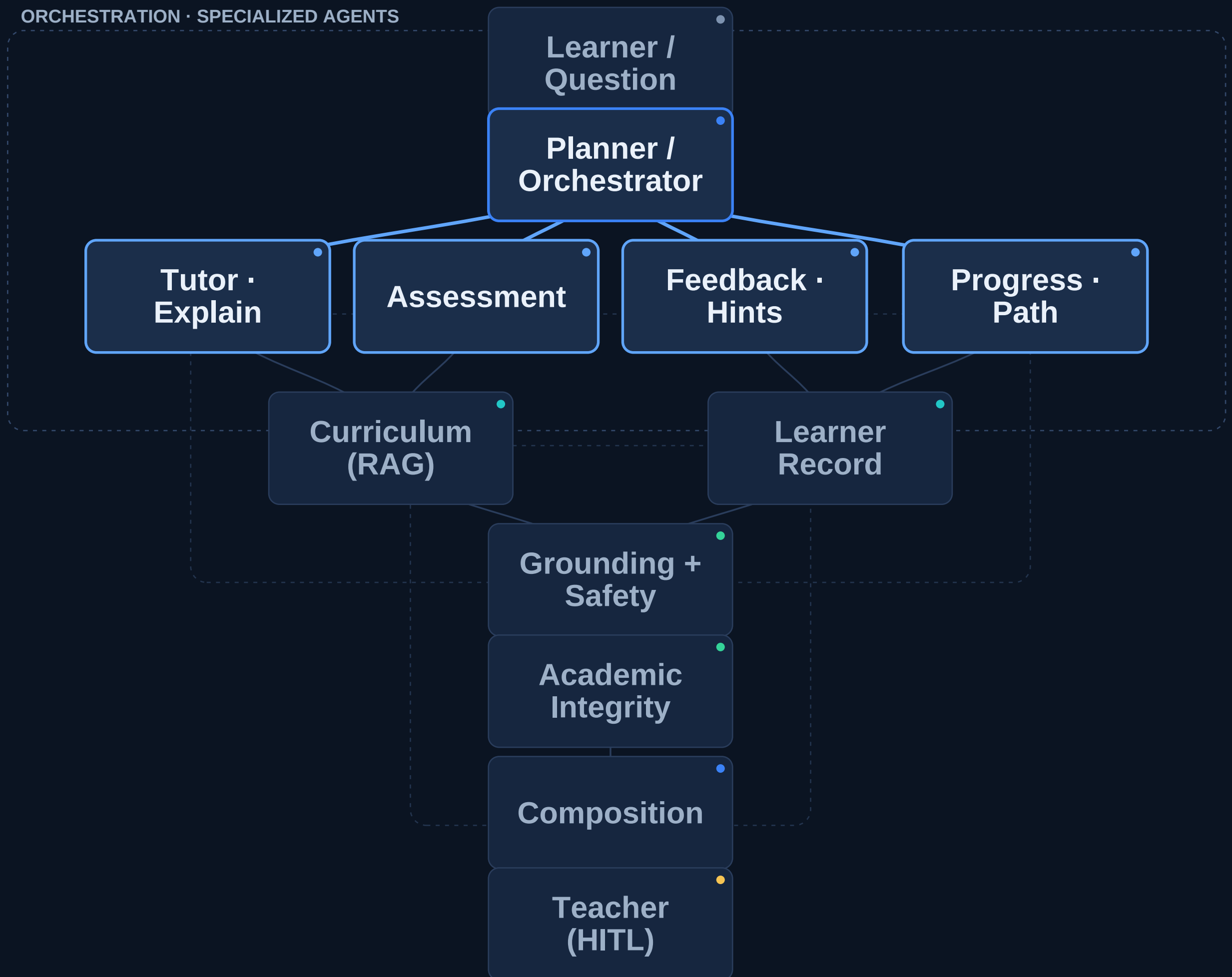
Every question flows through specialized agents, curriculum grounding, safety and integrity checks, and a teacher.

Governance throughout: consent, PII redaction (FERPA/COPPA), content safety, tool RBAC, audit.



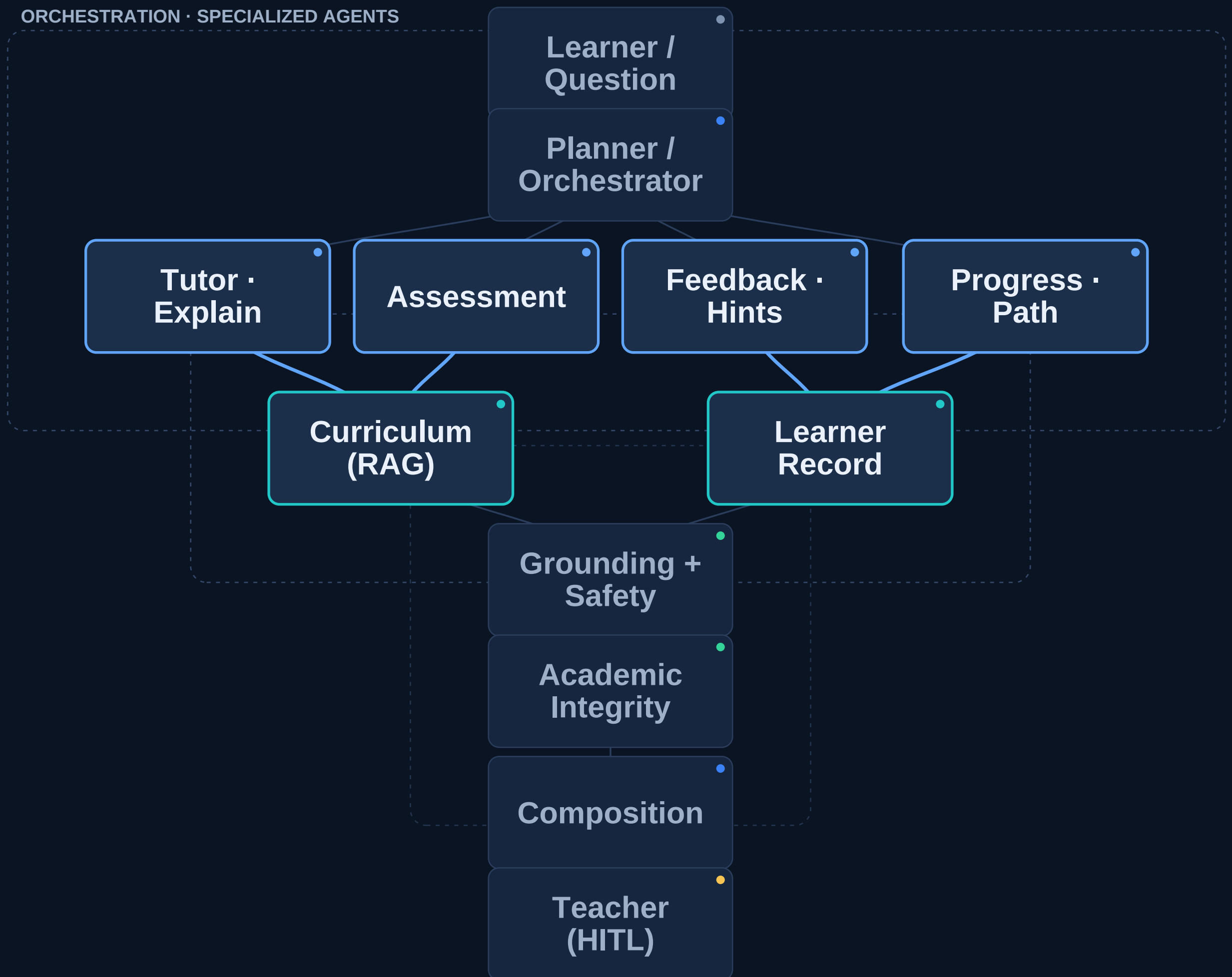
Before any model runs

A learner's question comes in; age-appropriate settings apply and student PII is redacted before a single prompt leaves the boundary.



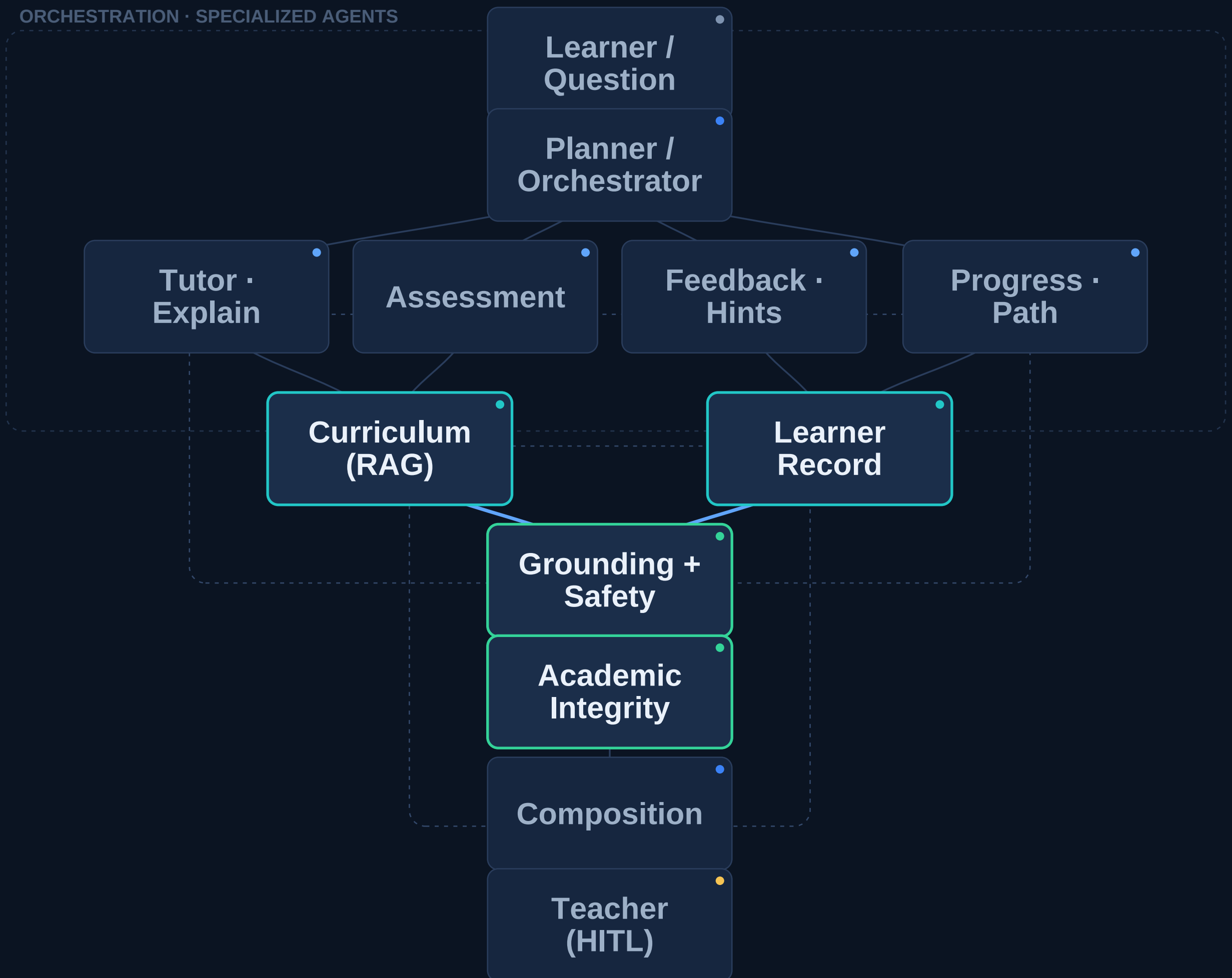
A planner, not one big model

The planner routes each task to a narrow agent — tutoring, assessment, feedback, learning-path — each easier to predict and verify.



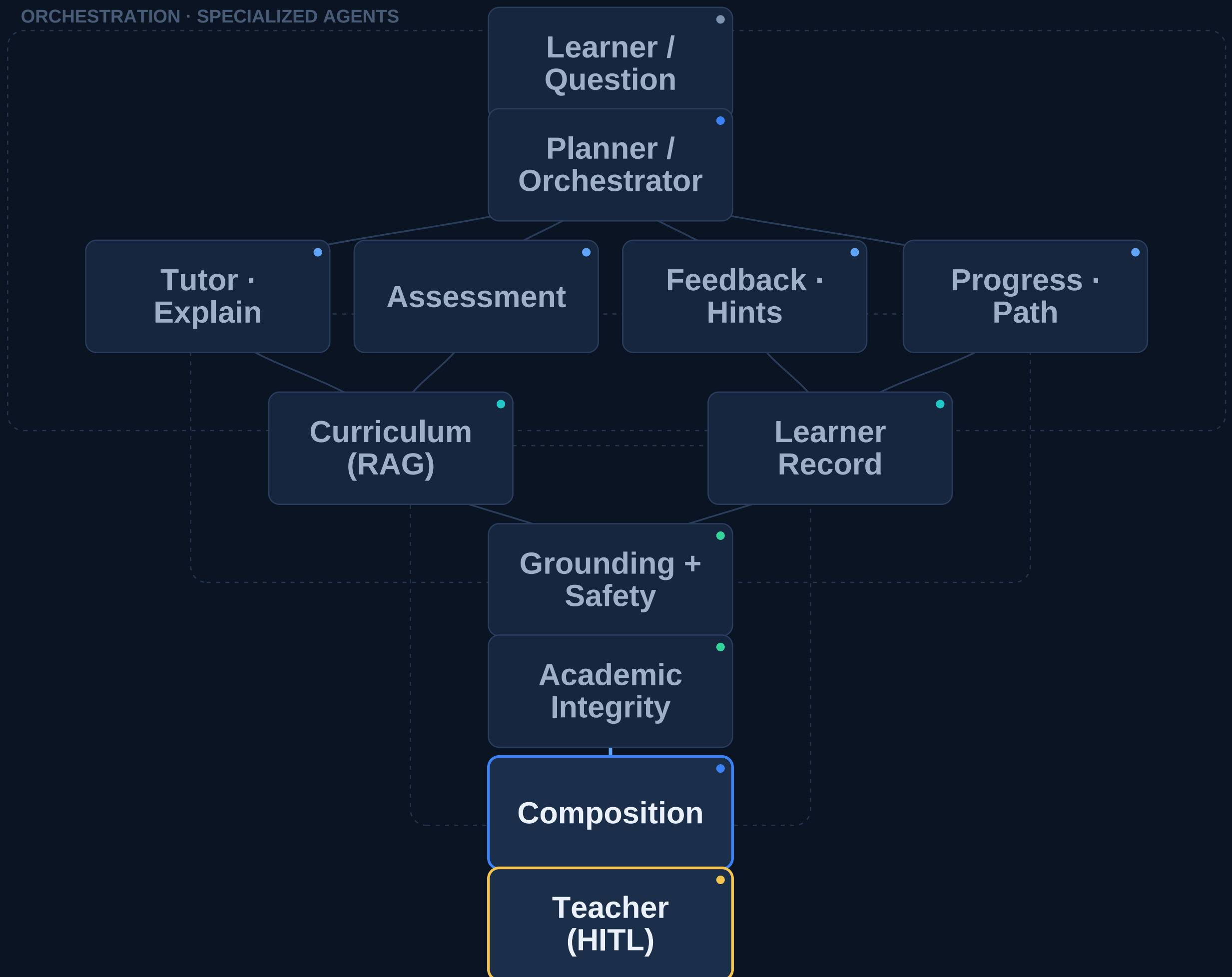
Anchored in approved material

Agents ground answers in the approved curriculum (RAG) and the learner's record. No made-up facts — every answer cites its source.



Grounded, safe, and honest

A grounding + safety guard keeps answers on-curriculum and age-appropriate; an academic-integrity check teaches the concept instead of doing the work for them.



Teacher in the loop

Output is composed with citations and logged for audit — and teachers stay in control, with visibility into every interaction.

Learner / Question & Planner

— Learner / Question

Typed/voice/assignment intake

EX A student types 'I don't get why we flip the inequality sign when dividing by a negative' during algebra homework.

TECH Web/LMS widget, voice-to-text, LTI 1.3 launch; OAuth2/OIDC SSO; per-tenant (school/district) isolation; Request envelope carries course + standard codes

CTRL Age-band tag from SIS for age-appropriate handling; PII minimized at ingress; consent state attached; Rate-limit + abuse filter before any model runs

FIN A cheap classifier triages trivial vs. complex questions so simple ones never reach a premium model.

FALL Unauthenticated or malformed input is rejected at the edge with a friendly retry, never entering the agent loop.

— Planner

Plans + routes sub-tasks

EX It splits the question into 'explain the rule' and 'check the student actually understood', routing each to a different agent.

TECH Orchestrator (Semantic Kernel / LangGraph) on managed compute; Deterministic, versioned routing policy; Per-session step + token budget, circuit breakers

CTRL Tool-RBAC allow-lists per agent; Tenant quotas + isolation; No agent may exceed the learner's grade scope

FIN Step/token budget per session caps spend; the cheap model routes, premium tier only on genuinely hard steps.

FALL Budget or breaker trips -> degrade to a single grounded explanation or hand to the teacher; state preserved to resume.

Tutor & Assessment

Tutor

Explains, step by step

EX *Explains, step by step, that dividing by a negative reverses the inequality - using the textbook's own number-line example.*

TECH Retrieval-grounded generation over curriculum; Worked-example + scaffolding prompt patterns; Citations to source pages / sections

CTRL Cite-or-refuse: no ungrounded claims; Read-only access to curriculum; Reading level matched to grade band

FIN Caches embeddings + frequent explanations; small model drafts, large model only for novel or complex topics.

FALL If the material doesn't cover the question, it says so and routes to the teacher rather than inventing an answer.

Assessment

Formative checks

EX *Generates two quick check questions on negative-coefficient inequalities to confirm the student got it.*

TECH Item generation grounded in standards; Difficulty calibration to the learner model; Answer scoring + misconception detection

CTRL Curriculum-aligned, standards-tagged items only; No high-stakes grading without teacher review; Bias / fairness checks on generated items

FIN Reuses an item bank and only generates new items when coverage gaps appear, cutting model calls.

FALL Low-confidence scoring is flagged for teacher review rather than recorded as a grade.

Feedback & Progress / Path

Feedback

Hints, not answers

EX *Instead of giving the answer, it nudges: 'What happens to the sign when both sides are multiplied by -1?'*

TECH Socratic hint-laddering prompts; Tracks hint level used per problem; Escalates specificity gradually

CTRL Never reveals full solution on graded work; Integrity policy enforced before output; Tone + age-appropriateness checks

FIN Hint templates are mostly deterministic; the model is invoked only when a tailored hint is needed.

FALL If the learner stays stuck after max hints, it offers the worked example and flags for the teacher.

Progress / Path

Mastery + adaptive path

EX *Logs that the learner now handles sign-flipping but still struggles with compound inequalities, adjusting the next lesson.*

TECH Learner model / knowledge-tracing store; LMS/SIS write-back of progress; Spaced-repetition scheduling

CTRL Accommodations (IEP / 504) respected; Progress data scoped to learner + teacher; No profiling beyond educational purpose

FIN Updates are incremental events, not full recomputation, keeping storage + compute cheap.

FALL If the record service is unavailable, it serves the last-known path and queues updates for later.

Curriculum (RAG) & Learner Record

Curriculum (RAG)

The only answer source

EX *Retrieves the exact textbook section and two practice sets on inequalities the answer must draw from.*

TECH Vector store + hybrid search + rerank; Per-course indexing with versioning; Effective-date / edition awareness

CTRL Only approved, signed materials indexed; Access scoped by enrolled course; Stale-edition guard

FIN Index once, query cheaply; rerank only top-k; hot, stable units can be cached to skip retrieval.

FALL No grounded passage found -> the tutor declines and points to the teacher rather than guessing.

Learner Record

Personalize, never leak

EX *Reads that this student has extended-time accommodation and a history of math anxiety, softening pacing.*

TECH LMS/SIS read via LTI / OneRoster; Field-level access scoping; Read-only credentials

CTRL FERPA-aligned access controls; PII masking in prompts; Audit of every record read

FIN Reads a cached, scoped profile rather than hot SIS calls on every turn.

FALL If the record is unavailable, it personalizes conservatively from session context only.

Grounding + Safety & Academic Integrity

— Grounding + Safety

Grounded + age-appropriate

EX *Blocks a draft that drifted into an off-syllabus political tangent, returning only the grounded math explanation.*

TECH Grounding verifier (cite-or-refuse); Safety + age-band classifiers; Toxicity / self-harm / PII filters

CTRL Unsafe output blocked, not shown; Every claim must carry a citation; Escalation to teacher on safety flags

FIN Deterministic checks are near-zero cost and run before delivery - far cheaper than a harmful answer.

FALL Any check fails -> the response is withheld and the learner sees a safe fallback with a teacher hand-off.

— Academic Integrity

Teach, don't shortcut

EX *On a graded essay prompt, it teaches the structure and cites sources instead of writing the essay for the student.*

TECH Assignment-type detection (graded vs. practice); Citation enforcement; Plagiarism-risk heuristics

CTRL No full answers on graded work; Provenance preserved end-to-end; Configurable per teacher / policy

FIN Mostly rule-based gating; model involvement is minimal, so the control adds little cost.

FALL When intent is ambiguous, it defaults to the teaching path and flags for teacher confirmation.

Composition & Teacher (HITL)

— Composition

Assemble verified pieces

EX Assembles the grounded explanation, the two check questions, and citations into one tidy learner reply.

TECH Deterministic templating of verified parts; Inline citation rendering; Readability shaping

CTRL Only verified content is assembled; No new claims introduced at compose; Provenance retained

FIN Pure assembly - no extra model call, so composition is effectively free.

FALL If a piece failed verification, it is dropped and the reply notes the gap rather than fabricating.

— Teacher (HITL)

Human in control

EX The teacher sees a dashboard flag that three students missed sign-flipping and assigns a targeted mini-lesson.

TECH Teacher dashboard + review queue; Override + feedback loop into the learner model; Notifications on safety / integrity flags

CTRL Human sign-off for grades + path changes; Full audit trail of AI actions; Teacher can disable any agent

FIN Risk-tiered routing means only flagged or high-stakes items consume teacher time.

FALL Anything the system is unsure about lands here - the default safe path is always a human.

YOUR TURN

Bringing AI into learning safely?

Here's how I'd architect a trustworthy education AI — grounded in your curriculum, safe by design, FERPA-aware, and teacher-led — on Azure.

DM me “assessment”

Atiqul Islam

Fractional Cloud & AI Architect for SMBs · atiquislam.dev