

# ‘Desire Paths’ in AI-Guided Software Learning

Emran Poh

emran.poh.2025@smu.edu.sg

School of Computing and Information Systems

Singapore Management University

Singapore, Singapore

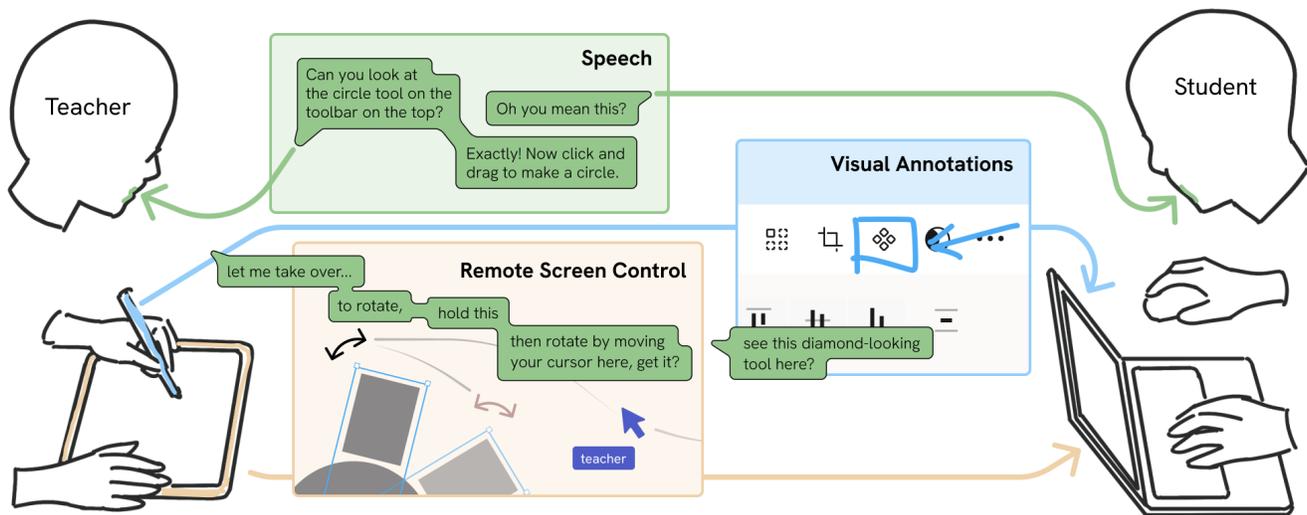


Figure 1: In remote software tutoring, teachers coordinate three modalities: *speech* (conceptual guidance), *visual annotations* (spatial precision, directing attention), and *remote screen control* (direct demonstration). This escalation in response to learner behavior illustrates the precision–agency trade-off and motivates the need for frameworks and study of adaptive multimodal intervention design for tutoring systems.

## Abstract

Learners who deviate from prescribed instructional paths are not necessarily failing—they may be revealing signal about themselves, the curriculum, and the interface. Yet AI tutoring agents for software learning typically treat deviation as error to be corrected. We argue this framing is fundamentally misaligned with how learning actually unfolds. Drawing on the urban planning concept of *desire paths*—informal trails that emerge when pedestrians route around formal walkways—we reframe learner deviations as meaningful signals of unmet needs, prior knowledge, and curriculum gaps. We ground this argument in an empirical study of human software tutors teaching Figma, which reveals that expert tutors already read and respond to desire paths by fluidly shifting across instructional modalities—speech, visual annotation, and remote screen control—in response to learner behavior. We offer a preliminary taxonomy of deviation types, argue that the Human↔Agent↔UI triangle in learning contexts must be redesigned around *negotiated path-finding* rather than path enforcement, and argue that the field needs frameworks and empirical study of adaptive multimodal intervention design for tutoring systems to make this negotiation possible.

## 1 Introduction

When AI tutors guide learners through complex software (e.g., design tools such as Figma), we argue that breakdowns often arise in the learner–agent relationship rather than in the tool itself. Learners skip steps, request demonstrations, seek help, or wander off-path; systems may treat such behavior as deviation from the learning plan or, alternatively, as meaningful signal. We argue for the latter. In urban planning, a *desire path* is the informal trail that forms when pedestrians route around a formal walkway that fails their needs—and effective design responds by paving it. Applied to software learning, learner “deviations” similarly carry signal about prior knowledge, cognitive state, and curriculum fit.

The *assistance dilemma* holds that excessive assistance yields passive learners and poor transfer, while insufficient assistance yields unproductive struggle; the optimal balance varies by learner, moment, and task. Current AI agents typically resolve it statically (fixed curriculum or on-demand help), without sensitivity to deviation signal. We propose a desire path framing instead: the agent reads the learner’s path and responds dynamically rather than pre-specifying support versus autonomy. GUI agents can navigate interfaces on behalf of users, but in learning contexts agent-as-executor may preclude the learner’s cognitive engagement. The educational design challenge is *productive scaffolding*: support that

does not displace the cognitive work of learning. In the standard model, human intent is stable and the agent executes. In learning, we argue, the learner constructs intent through interaction; the agent may therefore need to *interpret and negotiate*, not only execute.

This paper has four aims. We ground the desire path metaphor in our reading of empirical observations of human software tutoring and argue it captures an undertheorized phenomenon. We then connect the precision–agency trade-off in multimodal instruction to desire path interpretation, giving a principled basis for agent response. We offer a preliminary list of deviation types (and possible multimodal responses) that may be worth considering for human–agent–UI collaboration. Finally, we argue for the need for frameworks and for study into adaptive multimodal intervention design for tutoring systems, and outline open research directions.

## 2 ‘Show It, Don’t Just Say It’: The Complementary Effects of Instruction Multimodality for Software Guidance

We conducted an observational study of ten teacher–student pairs in which experienced Figma users taught novices two graphic design features through multi-step procedures [1]. The setup mimicked possible AI tutor modalities by limiting communication to three channels: *speech*, *visual annotations*, and *remote screen control*. Sessions were conducted via videoconferencing with screen sharing, annotation, and remote control; all interactions were recorded and analyzed for modality use, switching, and adaptation. The interventions (tutors’ responsive, modality-shifting moves in the face of learner behavior) struck us as an opportunity for further work. This position paper takes up that thread.

Speech formed the foundation of instruction: effective for conceptual guidance but limited for precise procedure. Annotations complement speech with spatial precision (directing attention, clarifying spatial relations); remote control adds spatial and temporal precision (direct demonstration) but intrudes on learner agency and on students’ sense of ownership of their digital territory—their ability to initiate and control their own actions and shared visual focus. Tutors calibrated modality choice to instructional precision needs and to student receptivity at each moment, consistent with the assistance dilemma. That choice was dynamic rather than curriculum-fixed: hesitation or off-path clicks triggered escalation (speech → annotation → remote control). We treat this graduated response as a form of desire path interpretation.

In the study, human tutors interpreted learner behavior, hypothesized causes, and selected modality in response. We take this interpretive capacity—not modality delivery alone—to be the core competency of effective software tutoring. An agent that can annotate or take control but cannot read *why* a learner deviated will misfire as often as it helps. This paper argues that interpretation is the missing layer, and that building it requires both a conceptual framework for deviation types and empirical study of adaptive multimodal response.

## 3 Desire Paths in Software Learning

In this paper we define *desire path* as learner behavior that routes around the prescribed sequence, reflecting genuine need or preference (purposive rather than random error). We argue that desire paths are neither static nor something to plan for in advance; they are often emergent and transient. The design implication, we propose, is not to anticipate or enumerate them but to build in the *capacity* for learners to take shortcuts when they feel comfortable and to make room for transient learning pathways (paths that may appear and dissolve within a session or across learners), so that low-friction adaptive AI tutoring systems can accommodate users’ desire paths. The definition is intentionally broad; the signal carried depends on deviation type. The following taxonomy is drawn from patterns observed across our ten tutoring sessions [1], organized by the inferrable learner state each deviation type suggests and the response it warrants.

- **Anticipatory skipping.** Bypasses a step (believes they understand). Signal: prior knowledge or over-scaffolding. Response: probe understanding rather than backfilling.
- **Answer-seeking.** Requests demonstration instead of working through. Signal: overload, frustration, or preference for observational learning. Response: shift modality (e.g., verbal → visual) while preserving agency.
- **Exploratory wandering.** Moves to off-task panels or tools. Signal: curiosity, boundary confusion, or mental-model building. Response: avoid immediate redirection (may suppress durable learning).
- **Productive stalling.** Pauses, backtracks, or repeats—often before insight. Signal: consolidation. Response: avoid premature intervention.

The central challenge is distinguishing desire path (signal) from noise. One can ask whether a deviation is *purposive* (intentional rather than mechanical), *recurrent* (recurrence may indicate signal where single instances may be noise), and whether it *correlates with outcomes* (deviations preceding success may be valid routes; those preceding failure may indicate confusion). Measuring deviations would require operationalizing these ideas—action sequences and timing such as skip vs. attempt, latency, and backtracking, and optionally multimodal cues like hesitation or verbal requests—together with logs that support recurrence and outcome correlation. We hypothesize that *recurrence* is the most tractable of these criteria for real-time detection from interaction logs alone—a single off-path click is ambiguous, but repeated skipping of the same step type is not. Purposiveness and outcome-correlation likely require richer signals: multimodal cues such as hesitation, verbal requests, or gaze, and longitudinal learner models that track behavior and inferred knowledge rather than task state alone. Whether recurrence alone suffices for useful real-time classification, or whether richer streams are necessary, is an empirical question we flag as the field’s most immediate.

## 4 Design Implications and Research Agenda

We propose reframing the agent as *co-navigator* rather than *guide*: negotiating path with learner rather than enforcing it. Such agents would maintain a model of learner deviation patterns, interpret deviations against that list, hypothesize the signal each carries, and

respond to that signal rather than to path compliance. We use *negotiated path-finding* for the CHI pattern—tutors shifting modalities to calibrate precision and intrusiveness to demonstrated need [1]. A set of modalities matched to the deviation types above (e.g., anticipatory skipping → low-intrusiveness probe) would support that; the interface must expose behavioral signal and the agent must possess a sufficiently rich set of modalities. As noted in Section 2, current systems do not yet meet these conditions, which motivates the need for frameworks and empirical work on adaptive multimodal intervention design. In the Human↔Agent↔UI triangle, the learner communicates intent partly through deviation (behavior as signal); the agent mediates by interpreting those signals and choosing when to speak, annotate, or take control; the UI must support agent-readable traces (so desire paths are legible) and human oversight (visible path, re-entry, option to intervene). Negotiated path-finding is one concrete instantiation of that triangle when the user is a learner and intent is emergent rather than fixed.

Path-finding could apply not only to *intervention* but to *path structure*: steps that reorder, branch, or reshape from observed desire paths (e.g., repeated skipping may suggest over-scaffolding; recurrent wandering, missing landmarks or useful detours). We use *structurally fluid* for designs that treat the prescribed path as malleable, updated from deviation patterns, learner models, and outcomes so the formal path itself evolves. Because desire paths can be transient (Section 3), the aim is to support *transient pathways* with low-friction re-entry when learners return to the main path, not to fix or pre-plan alternative routes.

## 5 Future Work

The empirical questions follow the framework's three layers. **Detection**: whether recurrence in interaction logs suffices for real-time desire path classification, or whether gaze, voice, and affect are necessary; and how thresholds and features should be operationalized for practical use. **Interpretation**: the extent to which agents can reliably classify the four deviation types above and infer their causes; and what learner-modeling architectures support this. **Response**: which multimodal responses to each deviation type outperform path-enforcement, and how the optimum varies by learner and task—the preliminary mappings in Section 3 are hypotheses awaiting empirical comparison. Cutting across all three: whether graduated response (speech → annotation → control) outperforms fixed modality; interface affordances that expose desire paths to agents without disrupting learner flow; and how structurally fluid path design (Section 4) can evolve the curriculum from deviation patterns while preserving learning objectives.

## 6 Conclusion

Learners who deviate are not necessarily failing—they may be revealing signal about themselves, the curriculum, and the interface. How AI tutoring agents respond to that signal will shape learning outcomes at scale, yet the deviation-as-error framing that dominates current systems discards it entirely.

Human tutors already read desire paths and calibrate response; AI that does likewise would represent a qualitatively different form of tutoring intelligence—one that interprets rather than merely executes. The challenge we invite is developing the frameworks

and empirical methods needed to make that interpretation possible: so that agents and interfaces can detect, legibilize, and respond to the paths learners are already carving.

## References

- [1] Emran Poh, Yueyue Hou, Tianyi Zhang, and Jiannan Li. 2026. Show It, Don't Just Say It: The Complementary Effects of Instruction Multimodality for Software Guidance. In *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems (CHI '26)*, April 13–17, 2026, Barcelona, Spain. ACM, New York, NY, USA, 1–22. doi:10.1145/3772318.3791200