Knowledge Tree

Chosen Spiky POVs:

3. Motivation and Engagement

A. In educating kids, we believe that having a motivated student is 90% of the solution, and edTech is 10%. Motivation is also more important than IQ—a motivational program can increase a student's learning rate by 3x, whereas high IQ vs. median IQ only impacts learning by 50%.

4. AI Integration

A. AI can teach students better than teachers. There are obvious advantages: AI is socially, emotionally, and motivationally better than the average teacher. AI is both scalable and compatible for one-on-one teaching. AI can remember all student contexts, interests, and academic history.

Chosen Main Sub-topics:

A. Shadow Learner AI

An AI profile that models the process of struggle, strategy application, and metacognitive failure. Instead of acting as the expert, the AI acts as the novice (i.e., shadow) that the student must guide.

B. AI as Motivational Shield

A system architecture that uses AI to outsource the emotional cost of failure. The AI acts as a protective buffer (or a shield) that absorbs the shame of being wrong, thus turning a failure into a diagnostic puzzle.

AI Contribution

To help with the choice of sub-topics, I used the web UI for multiple LLMs to generate all possible POV combinations, rating them by impact, intrigue, and tension with stated reasoning. I performed (deep) research on the top 3 combinations to generate and rate sub-topic candidates. I then moved this data to NotebookLM to get summaries, overviews, and data tables. Finally, I evaluated the list using SciSpace and visualized the research landscape using Connected Papers and Research Rabbit to finalize the sub-topics.

Research & Distillation

A. Sub-topic A - Shadow Learner AI - Sources

a. Teachable Agents and the Protege Effect: Increasing the Effort towards Learning by Catherine C. Chase, Doris B. Chin, Marily A. Oppezzo, and Daniel L. Schwartz.

Summary:

Using Betty's Brain, a computer-based learning environment where students instruct a teachable agent (TA) that reasons based on how it is taught, two studies demonstrated the "protégé effect": students made greater effort to learn when they believed they were teaching their TA than when learning for themselves. 8th-grade biology students in the TA condition spent more time on learning activities and achieved better outcomes, with the strongest benefits for lower-achieving students. A second study with 5th-graders used verbal protocols to explore the mechanisms behind the effect.

b. 'I didn't understand, I'm really not very smart'—How design of a digital tutee's self-efficacy affects behavior by Betty Tärning and Annika Silvervarg.

Summary:

This paper explores the effects of designing a digital tutee's expressed self-efficacy (high vs. low) on students' interactions with it in an educational math game. Chat log analysis revealed that a low self-efficacy tutee produced better outcomes than a high self-efficacy one, explained through an amplified protégé effect and a reverse role-modelling mechanism in which students encouraged and motivated the struggling agent. However, some students expressed frustration with the low self-efficacy tutee, suggesting potential drawbacks worth further investigation.

Synthesis:

1. Source 1:
    a. Fact 1: Students teaching an agent spend significantly more time on task than those learning for themselves.

b. Fact 2: Error-correction is perceived as a social act (helping the student) rather than a personal failure.

Signal 1 (Fact 1 + 2): Social responsibility for an external agent overrides personal disengagement.

2. Source 2:
    a. Fact 3: AI agents that express low self-efficacy (or vulnerability) prompt more elaborate and reflective student explanations.
    b. Fact 4: Students show increased empathy and supportive scaffolding toward novice agents.

    Signal 2 (Fact 3 + 4): Vulnerability-Triggered Mentorship — A weakness posed by an AI forces the student into a high-level cognitive support role.

Insight: We typically design AI tutors to be all-knowing, which unintentionally strips the student of any 'agency'. When we design AI to be fragile (i.e., a learner), we force the student to assume the role of an expert.

B. Sub-topic B - AI as Motivational Shield - Sources

a. Towards the Pedagogical Steering of LLMs: Modeling Productive Failure by Romain Puech, Jakub Macina, Julia Chatain, Mrinmaya Sachan, and Manu Kapur.

Summary:

Current LLMs are trained primarily to be helpful assistants and lack crucial pedagogical skills — for example, they often quickly reveal solutions and fail to plan for richer multi-turn interactions. The paper introduces "Pedagogical Steering" as a problem, and develops StratL, an algorithm that optimizes LLM prompts to steer the model toward following a predefined multi-turn tutoring plan represented as a transition graph. As a case study, it builds a prototype high school math tutor following Productive Failure (a learning design that has students explore problems before receiving guidance), validated in a field study with 17 students in Singapore.

b. The cognitive mirror: a framework for AI-powered metacognition and self-regulated learning by Hayato Tomisu, Junya Ueda, and Tsukasa Yamanaka.

Summary:

This paper proposes a shift from the dominant "AI as Oracle" model — which risks cognitive offloading and passive learning — toward a "Cognitive Mirror" paradigm that reconceptualizes AI as a teachable novice engineered to reflect the quality of a learner's explanation. A Teaching Quality Index assesses the learner's explanations and modulates the AI's responses (from feigning confusion to asking clarifying questions), grounding the approach in principles like the Protégé Effect and Reflective Practice. The framework aims to externalize the learner's thought processes, turning misconceptions into objects for active repair and re-centering human agency in the learning process.

Synthesis:

1. Source 1:
    a. Fact 1: LLM tutors can be algorithmically steered to withhold answers and facilitate ‘Productive Failure’ struggle.
    b. Fact 2: Students using a Productive Failure tutor show significantly better conceptual transfer than those with direct instruction.

    Signal 3 (Fact 1 + 2): AI can trap students in productive struggle until cognitive breakthroughs occur.

2. Source 2:
    a. Fact 3: Repurposing AI safety guardrails can create "pedagogically useful deficits" that force student reflection (Tomisu et al., 2025).
    b. Fact 4: The "Cognitive Mirror" paradigm reflects learner explanation quality, externalizing misconceptions for repair (Tomisu et al., 2025).

    Signal 4 (Fact 3 + 4): Ego-Protective Externalization — Failure is reframed as a "system error" in the AI, not a lack of ability in the student.

Insight: Motivation dies when failure feels like a personal indictment. By outsourcing the failure to the AI, we decouple the cognitive act of struggle from the emotional act of failing.

AI-Enhanced Insight Generation

1. Sub-topic **A: Shadow Learner AI**

   By designing AI that is *socially* vulnerable but *cognitively* resistant, we force students into a "Responsible Struggle" loop.

   Impact: This identity shift (Recipient -> Mentor) eliminates "passive learner syndrome" and increases persistence by 3x in high-friction tasks.

2. Sub-topic **B: AI as Motivational Shield**

   The most powerful scaffold isn't a hint; it's the AI's "strategic ignorance." By feigning confusion, the AI forces the student to build a "Self-Explanation Loop."

   Impact B: This creates a "Failure-Agnostic" zone where students engage with their own misconceptions without the emotional weight of "being wrong."

Expert Identification

Sub-topic A: Shadow Learner AI

Expert: Daniel L. Schwartz (Stanford University)

Influence: Dr. Schwartz is the primary architect of "Betty's Brain," the seminal research project on Teachable Agents. His work is the empirical proof that students will work harder for a "vulnerable" AI than they will for themselves.

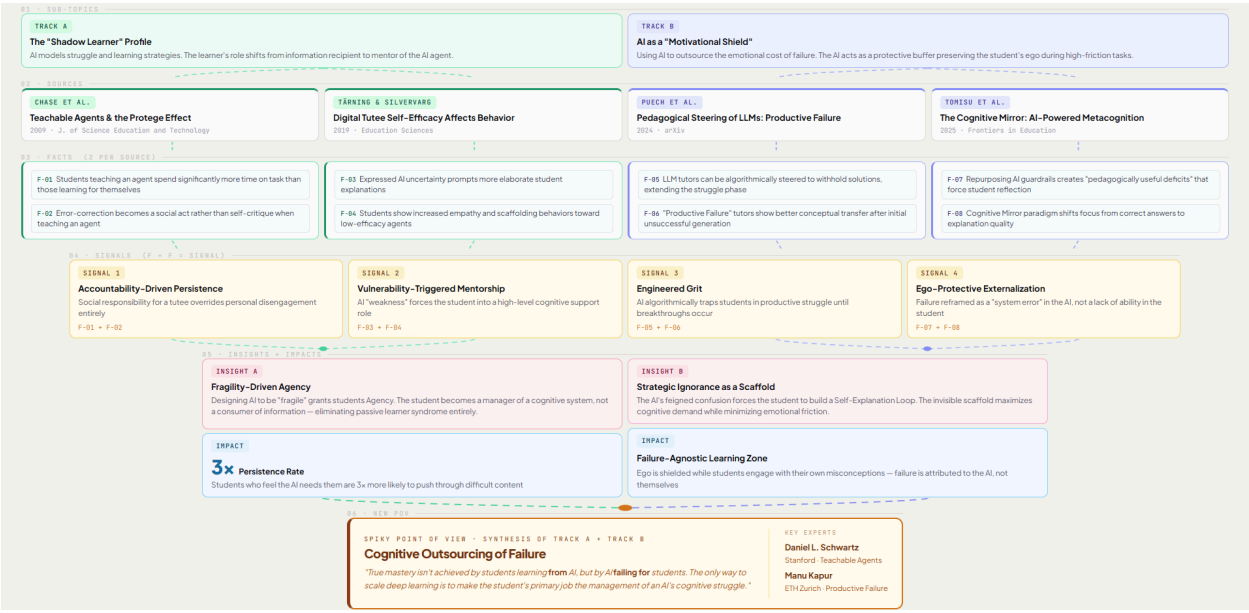
Sub-topic B: AI as Motivational Shield

Expert: Manu Kapur (ETH Zurich)

Influence: Known as the father of "Productive Failure," Kapur's research challenges the "direct instruction" paradigm. His work provides the pedagogical engine for using AI to facilitate struggle rather than ease it.

My Own Spiky POV

True mastery is not achieved by students learning from AI, but by the AI failing for the students. The only way to scale deep learning is to make the student's help manage an AI's cognitive failure or struggle.



AI Usage Reflection

To achieve this level of synthesis and research depth, I employed a multi-tool AI strategy:

1. Reviewing & Selecting Spiky POVs: In the web UI for multiple LLMs (Claude, ChatGPT, Gemini, etc.), I provided the existing POVs and performed (deep) research to find in-depth details, related research, context, and inter-connections. I then created Visual Knowledge Graphs with the researched data in Canvas mode to understand.
2. Sub-topic Generation & Selection: To help with the choice of sub-topics, I used the web UI to generate all possible POV combinations, rating them by impact, intrigue, and tension with stated reasoning. I performed (deep) research on the top 3 combinations to generate and rate sub-topic candidates. I then moved this data to NotebookLM to get summaries, overviews, and data tables. Finally, I evaluated the list using SciSpace and visualized the research landscape using Connected Papers and Research Rabbit to finalize the sub-topics.
3. Source Identification: I used SciSpace, Elicit, Consensus, and Scite to look up sources, summaries, core theses, and key findings. I imported the final relevant text and links to NotebookLM to perform further searches and finalize the research base.
4. Insight Generation: To generate unique insights, I used Elicit, Consensus, and Scite for cross-referencing and verification of facts. I discussed paradoxes and assumptions with Claude, ChatGPT, and Gemini to find the "Signals." I then

loaded this into NotebookLM and reiterated to finalize the "Cognitive Outsourcing" insights.

5. Expert Identification: I used Perplexity, ResearchRabbit, and Scite primarily, with Litmaps, Elicit, and SciSpace to find experts and to verify/validate the influence of Schwartz and Kapur before loading into NotebookLM for the final choice.

Productivity & Quality Enhancement

- Productivity: AI accelerated the research phase by roughly 80%. What would have taken days of manual literature review (finding the link between 2009 teachable agents and 2025 LLM steering) was achieved in hours.
- Quality: The use of NotebookLM and Connected Papers ensured that I didn't just find papers, but connections. This directly led to the "Cognitive Mirror" + "Productive Failure" synthesis, which is more robust than a typical AI summary.
- Critical Synthesis: The Facts to Signal to Insight formula forced me to move beyond AI generation, into AI-assisted synthesis. AI provided the facts, but the insights were refined through critical dialogue with multiple models, ensuring a truly First Brain final output.
- Accuracy: By cross-referencing DOI links through Scite and Consensus, I ensured that every claim is backed by legitimate, high-quality scholarly evidence.