

A Finished Tool Is Not Proof It Works

Deconstructing the AI Development Illusion in Public Services & Product Strategy

Wondering how Gemini would analyze the recent PSA post regarding Steve Blank learning experience based on ChatGPT analysis. See the following:

Source Article: Public Services Alliance (PSA) | Author: Kris (Sep 23, 2026)

Referenced Work: Steve Blank, “The Year AI Came For Us: Teaching Entrepreneurship Will Never Be The Same”

Executive Thesis

A finished, functional software product created using Generative AI does not constitute proof that the underlying problem has been solved.

Speed of development and surface-level polish are dangerous proxies for real-world efficacy. A working tool is merely the beginning of a hypothesis test, not the result of one.

1. Executive Overview & Context

In recent years, the democratization of artificial intelligence—specifically Large Language Models (LLMs) and automated coding agents—has drastically reduced the time required to build fully functional software prototypes. What once took engineering teams months to develop can now be built in days or hours.

However, as highlighted by entrepreneurship educator Steve Blank and analyzed by the Public Services Alliance (PSA), this rapid prototyping capability introduces a profound psychological and operational hazard:

“When teams build polished tools rapidly, they often spend less time learning what users actually need. When presented with evidence challenging their ideas, their investment in the finished product makes them slow to change course.”

2. Breakdown of Key Arguments

The ‘Illusion of Progress’ in AI-Accelerated Development

Traditional software development imposed a natural tax on building, forcing founders and teams to validate ideas extensively before committing heavy engineering resources. AI removes this friction, creating an ‘illusion of progress’ where a clean interface and functional code are mistaken for market validation or problem solved.

Confirmation Bias & Sunk-Cost Fallacy

When a team arrives with a working prototype, visual polish generates a false sense of certainty. This creates a strong sunk-cost effect: even when initial user feedback strongly contradicts the core assumptions of the tool, developers are reluctant to pivot because the tool ‘already works.’

The Public Sector Vulnerability

The article emphasizes that while this phenomenon damages startups, it is particularly dangerous in public services and governance:

- **Healthcare Applications:** A patient-facing application can feature a slick UI and accurate medical text without actually helping patients access appropriate care or improving clinical outcomes.
- **Educational Technology:** An AI tutoring platform can generate neat lesson plans and answers without demonstrating measurable improvements in student learning or engagement.

3. Three Pillars of Genuine AI Tool Validation

To prevent mistaking software completion for real-world efficacy, public sector leaders, investors, and product managers must evaluate solutions against three rigorous criteria:

Validation Pillar	Key Question to Ask	Red Flag / Anti-Pattern
1. Co-Design & Problem Definition	Who helped define the problem? Were end-users, patients, or frontline workers involved from Day 1?	The tool was built entirely based on developer assumptions without end-user input.
2. Post-Deployment Feedback	What specific outcomes and feedback did real users report after engaging with the tool in production?	Measuring success purely by app downloads, active sessions, or visual polish.
3. Pivot Agility	What structural changes were made to the tool when user experience contradicted the original plan?	Refusing to alter workflow logic despite clear evidence of user frustration or inefficiency.

4. Strategic Trade-Offs & Critical Evaluation

While the article provides a necessary reality check, an analytical evaluation reveals both critical strengths and operational risks in applying this philosophy:

Strengths of the Analysis

- **Exposes Goodhart’s Law in Tech Deployment:** When ‘speed of software delivery’ becomes the metric, actual end-user impact ceases to be prioritized.
- **Highly Timely for the Generative AI Era:** Directly addresses the 2024–2026 paradigm where LLMs make full-stack prototyping trivial, shifting the core bottleneck from technical execution to human validation.

Nuances & Potential Risks

- **Risk of Bureaucratic Paralysis:** Public sector procurement and deployment are already notoriously slow. Demanding exhaustive validation cycles before piloting can stifle helpful innovations.
- **Operational Cost Overhead:** Continuous co-design and iterative field testing require ongoing staff, budget, and coordination—resources that resource-strapped public agencies often lack.

5. Conclusion & Actionable Takeaways

The central takeaway for organizational leaders, technology buyers, and public policy experts is clear:

AI accelerates product creation, but it does not accelerate human learning or social validation.

True efficacy is proven through measurable real-world outcomes and user empowerment, never by working code alone.