Tiger's Early 2026 AI Observations

Executive Summary

AI is broadly recognized as transformative, yet early 2026 remains marked by noisy, polarized narratives that obfuscate near-term decision-making. This memo provides an operator's framework for allocating capital and talent based on what is observable today: step-change speed in well-scoped work without reliable autonomy, a preference for proven workflow techniques over speculative vendor "AI features," and the architectural choices required to convert experimentation into durable advantage.

1) Conflicting ideas: Commentary tends to polarize into "AI is a bubble" versus "AI is an immediate discontinuity," but the practical question should be anchored in observable signals: unit economics, emergent practices, and early evidence of labor substitution.
2) AI capabilities: AI feels like magic but stalls at 80–90% completion, which is insufficient for most use cases. AI labs are focused primarily on inside-the-model improvements, which have not solved the remaining 10%. The immediate solution is outside-the-model workflow and logic, which are bigger initial investments but offer the predictable results needed.
3) Growth ideas: The growth thesis is pragmatic—apply proven models and workflow techniques to redeploy existing product and go-to-market capacity toward high-ROI execution, rather than investing in speculative "AI features" that are difficult to sell, maintain, or defend.
4) Cost reduction ideas: The cost-reduction thesis is similarly pragmatic—standardize workflows and apply measurable automation with today's reliable techniques, rather than purchasing unproven vendor "AI features" or relying on end-to-end autonomous agents that are not yet dependable.
5) Risks and considerations: Setting the conceptual risks aside, the operational failures are self-inflicted through poor platform selection, weak portability, and inadequate data foundations. This combination turns AI transformations into brittle demonstrations and expensive vendor lock-in rather than durable capability.

Conflicting Ideas

The AI narrative is polarized between two extremes. One camp argues AI is largely a supply-side bubble: capital is flooding in, but adoption will be too slow and too incremental to

justify today's investment levels. The other claims the opposite: that AI is an immediate discontinuity that will upend the SaaS business model and rapidly obsolete large portions of white-collar work. Boards and markets oscillate between these views, shifting capital allocation and trying to "AI-proof" strategy. Both claims cannot be true at the same time, yet most coverage treats one extreme as inevitable without serious engagement with opposing facts.

To move past the headline-level extremes, it helps to anchor the discussion in observable signals—where capital is being deployed, what the unit economics imply, and where productivity gains are already showing up in labor and software markets.

1) Total hyperscaler capital expenditures are trending toward $2 trillion. This planned spending is discretionary and can be delayed or canceled.
2) The capital expenditure has been a boon for suppliers, driving up market values of previously commoditized, pricing-taking businesses, such as Western Digital (500%) and Lumentum (900%).
3) The new data centers need to produce $500 billion revenue (using Equinix asset turnover as a benchmark) to generate sufficient returns for investors.
4) AI-native software engineers are spending $300/day on Claude Code, running 5 to 10 coding agents in parallel. At 250 working days, that is $75,000/year, comparable to the fully loaded cost of a single high-quality offshore FTE.
5) For context, 6 million professionals in India and the Philippines directly serve the U.S. market in IT and BPO roles.
6) SMBs are beginning to do DIY internal software development for $20/month instead of paying contractors $200K per project. If strong internal talent costs $200K/year and a typical project takes 3 months, the internal cost is $50K or 75% savings.
7) "SaaS-pocalypse" erased $1 trillion of market value in 2026, and likely $1.5–2 trillion if measured from mid-2025.
8) White-collar information businesses (e.g., Morningstar) saw declines comparable to the "SaaS-pocalypse" drawdown.
9) Broadly, the latest earnings season did not feature surprise revenue declines attributable to AI-driven substitution. Instead, commentary centered on new AI features, increased capex, and hardware revenue growth.
10) In contrast, Anthropic and OpenAI have repeatedly absorbed ideas from single-feature AI startups that showed traction, effectively commoditizing the feature and undermining the standalone startup.

Taken together, these facts suggest the near-term outcome is less about a sudden "AI end state" and more about second-order effects: capex cycles that can whipsaw suppliers,

rapid labor substitution in well-scoped digital work, a repricing of software moats, and accelerating feature commoditization as frontier labs pull capabilities into the platform.

1) If AI capex ultimately proves unsustainable, the downside is asymmetric. Hyperscalers may lose two years of discretionary spending and still exit with their core businesses intact. Suppliers are far more exposed: a vanished buildout becomes a sudden revenue cliff, and price-taking vendors that funded their own capex may face distress or bankruptcy.
2) Even at large headline numbers, $500B of incremental revenue is plausible if AI enables "digital re-shoring"—substituting software for meaningful portions of offshore labor across IT and BPO.
3) The labor impact will extend beyond offshore delivery. U.S. white-collar roles that resemble "offshorable" work will be pressured, and businesses without durable economic moats, effectively in the same substitution bucket, will be at risk of disappearing.
4) Raw code is no longer IP or provides any real barriers to entry. However, the prompt chain, framework, and broader environment that allows coding assistants to generate consistent outcomes are the new IP.
5) Companies with real barriers to entry have a path to win: adapt their products and go-to-market to help customers execute the "great digital re-shoring" and accelerate medium-sized business digitization.
6) AI startups built around a single compelling feature without defensible distribution, data, or workflow lock-in will be vulnerable to platform absorption—and will have their differentiation competed away by the large AI players.

AI Capabilities:

A practical way to cut through the polarized narrative is to focus on what AI systems can reliably do in 2026. AI feels like magic but stalls at 80–90% completion out-of-the-box, which is insufficient for most use cases. The last 10–20% of quality, determinism, and integration remain elusive to the most advanced reasoning models. The immediate solution is outside-the-model workflow and logic provided by extensive human guidance during deployment.

1) In my experience, even the best AI agents (e.g., Claude Code) still require heavy guidance. The productivity gains can be 10–100x and feel like magic, but without clear direction the output reliably stalls at 80–90% especially on more challenging problems.

2) A useful framing is “inside the model” vs. “outside the model.” Inside-the-model means changing model weights—where the frontier labs have concentrated effort, resulting in newer reasoning models. Outside-the-model means augmenting behavior with prompt chains, logic, and checks; this is often described as building wrappers.
3) Inside-the-model approaches are generally more robust and generalizable across contexts and edge cases. The tradeoff is that outputs can be more generic and less predictable—and materially more expensive (top reasoning models can cost 10–20x more per API call).
4) In addition, inside-the-model improvements will drive models to be bigger, better, and slower. This means no edge device deployments for large models.
5) Outside-the-model approaches are usually workflow-specific. With standardized prompt chains plus heuristic checks, they can achieve near-deterministic results—but they require meaningful upfront engineering, and they become brittle as the workflow changes.
6) In theory, combining the two sounds attractive, but in practice it often produces the worst tradeoffs: strict prompt chains and checks can negate any model-level gains, while cost and rigidity compound rather than improve.
7) My view is that smaller models are “good enough,” and most companies will prioritize near-deterministic outcomes, low inference cost, speed, edge devices, and scalable workflows. This creates substantial opportunities across industries and functions to build outside-the-model solutions.

Five Growth Ideas:

The growth opportunities below assume a pragmatic posture. The focus is to use today’s proven models and workflow techniques to ship value faster, rather than burning budget on speculative “AI features” that are hard to sell, hard to maintain, and easy for platforms to commoditize. The goal is to redeploy existing products, engineering, and go-to-market capacity toward high-ROI changes that compound, without betting the roadmap on uncertain breakthroughs.

1) Re-prioritize revenue features: The product roadmap should be re-scored because the customer value vs. build cost frontier has shifted, and formerly “too expensive” revenue features may now be cheap to ship and accretive to ARR.
2) Accelerate feature delivery: The organization should be redesigned to ship revenue features ~10x faster by removing non-engineering bottlenecks in discovery, design, QA, launch, and enablement.

3) Multiply top sales capacity: Top conversion salespeople can be paired with AI agents to expand account coverage and increase effective selling time, which will lift team-wide average conversion.
4) Compress the sales cycle: AI agents can streamline contracting, pricing, and sales governance so deals move without manual handoffs, and routine approvals and exception reviews can run asynchronously instead of being batched into weekly meetings.
5) Expand into mid-market: The gains from items 1–4 can make it economical to move down-market, but the business will need a mid-market-ready package with modified product surfaces, tighter packaging, new pricing, and a reworked sales funnel.

Five Cost Reduction Ideas:

The cost-reduction levers below assume a similarly pragmatic posture. The focus is to standardize workflows and apply proven automation with today's models, rather than paying vendors for speculative "AI features" or assuming fully autonomous agents will be reliable enough to replace teams. The aim is to take cost out of existing processes (engineering, cloud spend, ops, and back office) with controls, measurement, and change management that make the savings immediate and durable.

1) Automate commodity software development: The engineering org should replace low-value build work with coding assistants, backed by clear standards, code review discipline, and targeted re-skilling.
2) Streamline documentation and QA: The product team can use AI tooling to draft docs and generate test cases, then keep humans focused on edge cases, acceptance criteria, and release gating.
3) Automate internal support and analytics: The business can standardize recurring ops work (e.g., product ops, engineering ops, routine dashboards) and automate it to reduce headcount load or consolidate fragmented roles.
4) Insource commoditized vendor work: The best targets are standardized services with measurable output quality—such as translation, design production, and transcription—where automation can deliver predictable savings.
5) Refactor legacy product code: Legacy code creates persistent "shadow costs" in support, reliability, and delivery speed, and refactoring improves the user experience. In addition, refactoring towards better AI interpretability will drive future productivity and savings.

Risks and Considerations:

This section is intentionally practical and not a comprehensive treatment of security, compliance, safety, or AI governance. It focuses on the execution decisions for a results-oriented operator. The most common self-inflicted failure modes are platform lock-in, weak portability on a rapidly shifting stack, and underinvestment in the data foundation—each of which can turn AI spend into brittle demos rather than durable capability.

1) Choose the right platform: The short answer is Google, because most alternatives are incomplete, expensive, harder to learn, less portable, and designed to capture IP and lock customers into a vendor-led “AI platform” narrative. If the choice is contested, three teams should run a short hackathon across three platforms to validate cost, developer velocity, and lock-in risk.
2) Over-index on portability: The AI stack is changing too quickly to predict, but it will consolidate around Python and the hyperscalers, so the engineering team should enforce portability standards to preserve relevance for the future.
3) Over-index on data infrastructure: The data model, pipelines, and governance layer underpin every AI workflow and agent, so the company should treat them as core infrastructure and invest in expert refactoring and ongoing operational rigor.