How to build great AI products: Key Excerpts

Context Note:

This summary was developed from a podcast transcript exploring how Vanta builds AI products using large language models (LLMs). If you'd like more details on how this material was compiled — such as the prompt design and iteration methods — please refer to the related blog post on my website. It offers a deeper look into the generation process behind these takeaways and recommendations.

1. **Focus on User Value, Not Just AI Novelty**

   Key Point: Before anything else, ensure the AI solution truly addresses a real user need. A feature that saves five minutes consistently can outweigh dozens of AI "bells and whistles."

   Extended Quote:

   "Building AI features for the sake of building AI features is not exciting... you don't have to build the Next Generation agent... you don't have to do that to have an impact on your customers... you have to get good at these technologies to make them just a simple footnote in your architecture so that you can deliver the right features to your customers."

2. **Treat Evaluation Like Robust Integration Testing**

   Key Point: Vanta emphasizes a rigorous testing approach that goes beyond vague impressions or "vibe checks." Gathering real data on AI performance is essential, especially in high-stakes domains where errors can erode trust.

   Extended Quote:

   "Eval actually just looks a lot like integration testing if you squint but with like a lot of test cases and sort of fuzzy outcomes... you have to go look at the data... you can't just like collect a vibe check like, 'Yeah, it kind of worked.'"

3. **Use Data Specs and Golden Sets for Clarity**

   Key Point: Define clear inputs, outputs, and ideal responses to guide both initial development and iterative refinement. Vanta's "data spec" functions like a PRD for AI, spelling out all relevant scenarios.

   Extended Quote:

   "This is how we communicate to PMs—the idea of like a PRD but in AI land... give me a spreadsheet that's your data set... then we're going to run that data set against this prototype AI feature... that data set early on is the spec for the product."

4. **Engage Subject-Matter Experts in Prompt Design**

   Key Point: Engineers know the technical side, but experts in compliance, security, or any specialized field know what "correct" looks like. Have them define your "ideal outputs" and even write prompts themselves.

   Extended Quote:

   "We have a pretty well ironed out process where our subject matter experts—who are compliance analysts, security experts—will work with product and the engineer to start to think through, 'Okay, here's what an ideal answer looks like, here's why this answer is wrong.'"

5. **Ensure High-Quality Retrieval of Context**

   Key Point: Good retrieval is crucial for any LLM-based product. Garbage in, garbage out applies here—if your AI can't fetch the right context, even the best model will fail.

   Extended Quote:

“The path to productionize… The pieces of productionize for us… one is on retrieval… we operate in a domain that’s really great for large documents… but retrieval is a service that’s almost completely managed within the AI team.”

6. **Perform Systematic Error Analysis and “Quality Hill Climbing”**

   Key Point: Vanta identifies specific ways the model goes wrong (e.g., confusion between multi-factor auth vs. two-factor) and systematically fixes them. Tracking these errors leads to steady, incremental gains.

   Extended Quote:

   “We start to build this notion of all the ways a prompt can go wrong… we might sit there and iterate on a single error case in the prompt repeatedly… it’s this long iterative process of looking at the outputs and describing what ideal outputs look like.”

7. **Maintain Ongoing Oversight and Improvement (Data Drift is Real)**

   Key Point: AI performance can degrade over time as user inputs shift. Set up alerts, monitor usage metrics, and schedule routine reviews to catch regressions.

   Extended Quote:

   “You’re never going to stop swimming in data… it’s not like classic ML where you need a team to constantly monitor, and it’s not like normal software where you can just ignore it… the question is, ‘Do I baby this thing like an ML model, or do I totally ignore it?’ Definitely not the latter.”

8. **Explore Multiple LLM Providers for Best Fit**

   Key Point: Models vary in strengths—for instance, some excel at code generation, others at summarizing. Don’t assume one model will suit every need.

   Extended Quote:

   “We definitely see differences… maybe Claude is really good at code generation whereas it might be less strong at other tasks… we’ll run classification prompts or compliance quizzes and see if they do better or worse.”

9. **Iterative Improvements as the ‘Secret Sauce’**

   Key Point: There’s no magic formula—only the “deeply unsexy” work of relentless testing, prompt tuning, and data analysis. Regularly shipping improvements yields genuine quality gains.

   Extended Quote:

   “It’s just quality hill climbing… we have a lot of ways to improve, but there is no special trick… it’s the deeply unsexy work that gives you capital Q quality, and that’s how you differentiate.”

10. **Dogfood Your Own AI Products Internally**

    Key Point: Use your tools the same way customers will, so you catch flaws early. Vanta’s internal security teams use the same platform to experience it firsthand.

    Extended Quote:

    “Dogfooding is a key feedback mechanism… Vanta uses Vanta pretty extensively… we’re able to roll features out to them, get feedback… we can see if something breaks or if they don’t trust the output.”