Why Build AI Platforms

Novel applications of AI are best for this

- When you're building something that doesn't exist yet, you need architecture designed specifically for that problem
- Complex AI applications require precise control over how prompts are structured, how state is managed across multiple AI calls, and how data flows between different parts of the system
- Building from scratch means you can optimize every layer for your specific use case—from how you call the AI to how you handle edge cases to how you manage costs at scale
- The architecture choices you make define what's possible. Emstrata's multi-layer narrative system works because the entire platform is built around maintaining continuity and tracking story threads

The Cost Reality of AI Apps

How much an AI application will cost to run, test, and maintain

- AI API usage is the basis for how much you will be charged. Usage is measured in tokens (MToks). Different models and APIs will charge you different amounts. Take into consideration that different models also have different strengths that are often not reflected in industry benchmarks

- Input and Output costs are often different prices, with output tokens typically costing more, meaning that an application that wants to output large AI responses consistently will pay more than applications with larger context windows

- The size of your context window will not just run up prices but could effect performance according to many studies on the matter, this is why managing the amount of tokens we're sending to the API matters and why standardized methods of controlling and organizing tokens into easily readable inputs is preferable