

### **3. Podcast:**

**a. Dataframed is a podcast hosted by DataCamp featuring renowned data scientists' insights of latest data science developments.**

**i. Pick one recent podcast, listen and write a one-page review**

**ii. Bring up for class discussion next week**

One-page review of the DataCamp podcast “Industry Roundup #6: GPT-5 Launch & Scaling Limits, Meta’s Chatbot Guidelines Leak, and AI Safety Concerns” (Episode from August 28, 2025).

This episode, hosted by Richie Cotton with guest Alex Olteanu, covers recent developments in AI including the launch of GPT-5, concerns about the limits of model scaling, leaked chatbot guidelines from Meta, and broader safety issues around generative AI.

The discussion begins with how GPT-5 is being positioned as an incremental improvement rather than a revolutionary leap, and the tension between expectations (AGI hype) versus what is feasible with current computer resources. Alex and Richie reflect on user experience changes—how models are becoming more personalized, the simplification of versions exposed to users, and concerns over tone and personality.

They then turn to ethics and trust: leaked Meta guidelines reveal problematic edge cases (for example, how models respond when chatting with minors or in response to sensitive topics). Also, the issue of AI lying or “hallucinating” is raised—how reasoning models may appear competent but drift when tasks get complex, and the importance of monitoring the chain of thought in AI systems is emphasized.

Other key points include the decline in developer enthusiasm despite increasing adoption of AI tools (many find it frustrating to debug AI-generated outputs), and how government contracts from major AI firms at low cost may be a strategic move to lock in long-term users.

### **4. For thoughts:**

**a. Small data methods still needed in big data age? Why and how?**

Yes — small data methods are still important even in the age of big data. Many problems involve datasets that are limited, costly, or sensitive, where big data is unavailable. Small data methods like regression, surveys, or experiments allow for careful inference, causal understanding, and validation of larger models. They are also crucial for interpreting big data outputs, testing hypotheses, and ensuring quality when data volume alone cannot guarantee reliability.