Presentation Prep

Art of the Pitch

Jan Jannink, John Whaley

CS 224G - 1/28/2025

One Unexpected DeepSeek Side Effect

10 Biggest Billionaire Losers Monday

1. Oracle chairman Larry Ellison (net worth down $27.6 billion)
2. Nvidia CEO Jensen Huang ($20.8 billion)
3. Dell CEO Michael Dell ($12.4 billion)
4. Google cofounder Larry Page ($6.3 billion)
5. Google cofounder Sergey Brin ($5.9 billion)
6. Early Google investor Andreas von Bechtolsheim ($5.4 billion)
7. Tesla CEO Elon Musk ($5.3 billion)
8. Interactive Brokers chairman Thomas Peterffy ($4.1 billion)
9. Broadcom chairman Henry Samueli ($3.7 billion)
10. Broadcom cofounder Henry Nicholas III ($2.8 billion)

Commitments Based on Your Feedback

- Good mix of deeper technical content
- Workshop / demo tools and techniques
- Discussion of scale up methods and infrastructure
- Post more resources in LLM news on slack
- Creating connections with the community
- Expectations for each sprint

Administrative Details

- Attendance secret code! Reinforcement Shorting
  - And I thought crypto was volatile…
- Please invite our github account cs224g to your repo
  - Invite CAs too
- Assessment Methodology
  - Clone repo, study README, review commits (each team member contributes)
  - Follow code set up instructions, Run code

Presentations

- 2 slides
- 3 minutes per team
- Be sure to cover:
  - Problem you are solving and for whom - 30 seconds
  - Your solution - 30 seconds
  - Show a quick demo (screenshots, video) - 60 seconds
  - Your plan vs actual - 30 seconds
  - Something unique or interesting you learned - 30 seconds
- Share a link in your Slack channel, we'll add them to the deck for Thursday
  - Make sure permissions are world-readable!

Sample - Revelio

- Diverse team in skill and backgrounds
    - our own Akshay was on the team
- Intelligence gathering from public data

REVELIO

OSINT Revolution Realized

Olayinka Adekola, Akshay Gupta, Lawrence Liu, James Park

data-pipeline: summarizing

- Why so few RPM?
- LangChain
- LDA algorithm
- Topic summaries in React
- Next: maybe RAG, improve summaries, work with database, start draft report

Web processing, Neo4j/Cypher

- Two hopes for integration of graphs

1. Larger logical jumps for graph + RAG implementation
2. Easy visualizations for users: mapping relationships + interactions b/w entities

Scraped data -> Few-shot GPT (many diff prompts) -> Cypher Generation Alg -> Neo4J

(Hopefully) Neo4J -> Graph + RAG -> Q+A Interaction

data-pipeline: web scraping

- SERP API / Google Dorking
- Reddit API → summarization
- Next: revamp news API, full article output, SERP API, dorking

MongoDB/Atlas, FastAPI

- Settled on MongoDB Atlas (AWS) connected to FastAPI for robustness and speed
- Integrated MongoDB Atlas database with FASTApi backend server
- Developed API calls for frontend and LLM to facilitate data retrieval, storage, and interaction

Discussion

- Good
  - division of labor & presentation
  - technology
- Bad
  - missing narrative
  - is this a solution
- Ugly
  - busy slide
  - information overload

Takeaways

- Engage the audience
- Smooth transitions (practice to make it effortless)
- Make it memorable/memeable (fun, unexpected, to the point)
- Tell your story (now, the plan, the outcome)
- Make us look forward to the next presentation

Building an Effective Data Strategy via UX

Jan Jannink, John Whaley

CS 224G - 1/28/2025

Introduction to Data Strategy and UX

Data as a Strategic Asset

Product Engagement: The Core of Success

Enhancing Feedback Loops in Product Usage

Leveraging LLMs for Better Feedback

The Role of Internal Tools

The Importance of Small Details

Data Quality: Not All Users Are Equal

Addressing Bias and Diversity in Data

Challenges and Call to Action

Data strategy is the true competitive edge in the AI era.

The winners in this era of AI will be the organizations with ongoing access to highly-relevant, differentiated, quality data. In a world where all public data will be ingested and used, unique data is the most important strategic asset.

Jan Jannink, John Whaley

CS 224G -
1/28/2025

Data's Strategic Value

AI Strategy

The winners in AI will prioritize building effective data feedback loops. A static data set is not nearly as useful as ongoing access to data.

Data Is Not Oil: Differentiation Is Important

Data's worth lies in differentiation and targeting, not as a commodity. Effective and sustainable AI requires unique, high-quality data.

Understand Where To Use Data

Good data is not only for machine learning training. It enables you to offer personalized experiences and make data-driven decisions.

Strategy

A strong data strategy starts with product and user engagement.

Product Engagement

Implement strategies to capture high-quality data from engaged users. Great engagement drives retention and data insights.

User Engagement

Emphasize the importance of user engagement in building a strong data strategy. Longitudinal data reflects sustained user engagement.

Data Insights

Highlight the correlation between user engagement, retention, and effective data strategies. Encourage actions that promote user engagement.

Feedback

Making feedback effortless is essential.

Make It Inherent

Capture feedback seamlessly during product usage to enhance data collection without disrupting user experience.

Integration of Tools

Integrate editing and collaboration tools directly into the product for both user convenience as well as data feedback.

Instant User Gratification

Provide instant gratification for user corrections to enhance engagement.

User Feedback Techniques

Label Feedback

The user provides labels (A is good, B is bad, C is neither) to assess products or features.

Qualitative Feedback

Encourage users to provide detailed qualitative feedback to understand their experiences and pain points.

Discreet Feedback

Ask for discreet feedback to simplify data collection and gauge overall satisfaction (good/bad).

Passive Learning

Observe user behavior passively to gather data on how they interact with the product or platform.

Feedback Explanations

Prompt users to explain the reasons behind their ratings or labels to uncover insights.

AI Learning Overrides

Allow users to override AI-generated feedback, and learn from these changes to improve personalization.

Using LLMs to Enhance Feedback Quality

- LLMs allow for natural follow-up questions and to capture feedback subtleties.
- Improve understanding of not just the what, but the why.
- Integrate natural language for an enhanced user feedback loop.
- Efficiently summarize and extract key points from large unstructured textual data.

Tools

“Product” Includes Internal Tools

Internal tools, like those for data labeling and cleaning, are essential for maintaining high data quality in AI.

It is usually worth the investment, as empowering your employees, data labelers, and data reviewers has a multiplying effect.

For instance, at Redcoat AI, we devoted equal effort to our internal data labeling tool as our external product.

The Impact of Attention to Detail

Small Details Matter

Examples like spell-check and Grammarly demonstrate how small details can make a big difference in user experience.

Gamification Elements

Incorporating gamification elements can make the feedback process more enjoyable for users and encourage their participation.

Immediate Feedback

Offering immediate feedback satisfaction, like allowing users to correct data in real-time, enhances engagement and data accuracy.

Data Quality: Not All Users Are Equal

Factors Affecting Data Quality

- Data from experts is more valuable than data from novices.
- Data from consistent users is more valuable than data from inconsistent ones.
- Longitudinal data is valuable, as it is hard to get and essential to understand long-term value and retention.

Prioritize Data Value Through Validation And Diversification

- Use metadata and cross-validation to identify the most valuable users.
- Prioritize contributions from experts and consistent users.
- Prioritize data quality and diversity over sheer data quantity.

Managing Bias and Diversity in Data

Dangers from Lack of Diversity and Representation in Data

- Example at UnifyID: 95% of beta users were male techies from the Bay Area. So it didn't work for middle-aged women in Korea.
- Know your data and users to detect and address potential biases.
- Understand the inherent bias in generative AI tools and models.

Importance of Diverse Data Sets

- Creating diverse data sets is vital for quality and representation.
- Diversity in data boosts innovation, minimizes biases, and fosters inclusivity.
- Be creative about how to get data from a more diverse, representative set of users.
- Building a diverse data set is hard! But doing hard things creates differentiation and value.

Calls To Action

We can do better than conversational chatbots.

Encourage user engagement with proactive interfaces. Foster a sense of interactivity to enhance the user experience.

Let’s not recreate Clippy. Embrace creativity!

Embrace creativity and a user-centric design approach to deliver unique and impactful user experiences.

AI agents can be more interactive.

The best humans know when and what questions to ask. AI agents should do the same.

Consider data strategy in your products.

Think about how you can incorporate data strategy into products. Evaluate UX ideas based on the value of the data they will generate.