



LANGCHAIN ZOOMCAMP SESSION 1

Harpreet Sahota

Yes – these sessions are recorded.

They're made available to Substack subscribers.

If you're registered for this course, you're automatically added to my Substack!

ABOUT ME

- MSc in mathematics and statistics
- Former actuary, biostatistician, and data science lead
- Currently building DevRel at Deci AI and managing the Deep Learning Daily community
- Previously, I was developer advocate at Comet ML and Pachyderm
- Sacramento native
- Not a cringey LinkedIn “influencer”



WHO THIS SERIES IS FOR

- Data Scientists who want to learn how to leverage language models to drive business value
 - 1+ years of Python experience
 - 2+ years of ML experience
 - 1+ years of DL experience
- Engineers who want to learn prompting techniques
- Product managers who want to understand the complexities of building LLM-powered applications
- Tech folks who want to know what this LangChain buzz is all about



WHAT THIS
SERIES WILL
NOT COVER

- Math/Theory behind Transformers
- Architectures of different LLMs
- Self-supervised learning
- Deep learning in general
- Fine-tuning LLMs
- LlamaIndex

WHAT THIS SERIES WILL COVER

- Using LangChain with open-source and closed-source models
- Prompt engineering
- Retrieval Augmented Generation
- Agents
- Building applications with LangChain

WHY AM I DOING THIS COURSE?

I'm contracted to create a course for LinkedIn Learning, tentatively titled *Prompt Engineering with LangChain* and I want to test my material and get your feedback

TENTATIVE SCHEDULE

September 15 – Introduction and Get Inspired!

September 22 - Components of LangChain

September 29 – Chains and Prompt templates deep dive

October 6 – An overview of prompting strategies

October 13 – Retrieval Augmented Generation

October 20 – Agents and tools

October 27 and after – Building stuff!

INTRODUCTION

- World shaped by technology
- Daily tech rooted in language
- Language models' pivotal role
- Drive search engines, voice assistants
- Machines understand, generate language

DEFINITION OF LANGUAGE MODELS

- Language models: understand, generate
- Predicts next word/sequence
- Validates word sequence probability
- "Validity" mirrors human-like patterns
- Two primary language model types: probabilistic and neural

PROBABILISTIC LANGUAGE MODELS

- Predicts using n-gram probabilities
- N-gram: sequence of 'n' words
- Example: "I love dogs" 2-grams
- Limitation: lacks deep context

NEURAL NETWORK-BASED LANGUAGE MODELS

- Use "Attention" for context
- "Large Language Models" = Transformers
- Understands with self-Attention
- Categories: BERT, GPT, T5
- GPT-4: advanced generative model

FEATURES OF LARGE LANGUAGE MODELS

- Emergent abilities in LLMs
- Interface: prompting (e.g., GPT-4 API)
- GPT-3's in-context learning
- Understand LLM function and format
- Blends research and engineering

WHY YOU NEED TO UNDERSTAND LLMS

- Reliance on digital platforms
- Grasp LMs for digital appreciation
- Utilize, evaluate tech tools
- LLMs: leap in AI capabilities
- Redefining tech interaction potential

HOW DO LLMS GENERATE TEXT?

- LLMs: black box text machines
- Intuition on text generation
- Process: start, generate, append
- Token: chunk of text
- Steps: Encoding, Prediction, Decoding

ENCODING

- Encoding: text to vectors
- Embedding layer maps words
- Vectors capture semantic meaning
- Semantically close words = close vectors
- "Cat", "kitten": similar vectors

PREDICTION

- Prediction based on encoded input
- Uses learned language patterns
- Self-attention for context-awareness
- Softmax: vector to probability
- Selects most probable word/sequence

DECODING

- Translates vectors to text
- Generates word by word
- Depends on model size
- Quality influenced by training data
- Importance of seen tokens during training

RANDOMNESS, VARIATION, AND HUMAN INFLUENCE IN LLM OUTPUTS

- LLMs: randomness and varied responses
- Balance via temperature, top-k, top-p
- Temperature influences "creativity"
- Top-k: select from top probabilities
- Top-p: cumulative probability threshold
- Tools refine LLM outputs
- Ensures alignment with human intent

APPLICATIONS OF LLMS

CUSTOMER EXPERIENCE AND SUPPORT

- Chatbots: natural conversations
- Virtual Assistants: real-time support
- Sentiment Analysis: gauge customer sentiment

SOCIAL MEDIA

- Automated article assistance
- Blog, social media content
- Generate product descriptions
- LLMs: risk of similarity
- Potential unintentional plagiarism

E-COMMERCE AND RETAIL

- Real-time translation tools
- Document translation for businesses
- Localization of software, websites
- LLMs offer efficiency
- Ensure authenticity, human touch

FINANCE

- Fraud detection and prevention
- Financial news and trading analysis
- Loan underwriting, credit assessment
- Ensure data security
- Accurate financial data interpretation

MARKETING AND ADVERTISING

- Content categorization for recommendations
- Targeted advertising, marketing
- Enhance search engine results
- Risk: misrepresent brand voice
- Potential misleading content

CYBER LAW

- Privacy policy analysis, compliance
- Legal research, case analysis
- Cybersecurity incident response
- Interpret intricate legal language
- Maintain data confidentiality

HEALTHCARE

- Medical diagnosis assistance
- Literature review in biomedicine
- Patient data for personalized treatment
- Risk of misdiagnosis, incorrect suggestions
- Ensure thorough testing, validation

LET'S GET INSPIRED!