# Unsupervised Learning

Anote Presented by Ben Setel

# Overview

General Unsupervised Learning

RAG-based Unsupervised Learning for LLMs

Product Demo:

1. Complete answer in prompt
2. Incomplete answer in prompt
3. Relevant info but no answer in prompt

# What is Unsupervised Learning

No labels/answers provided -> can't "check our work"

BUT, can find patterns and correlations in the data

- Classification
- Outlier detection
- Dimensionality reduction

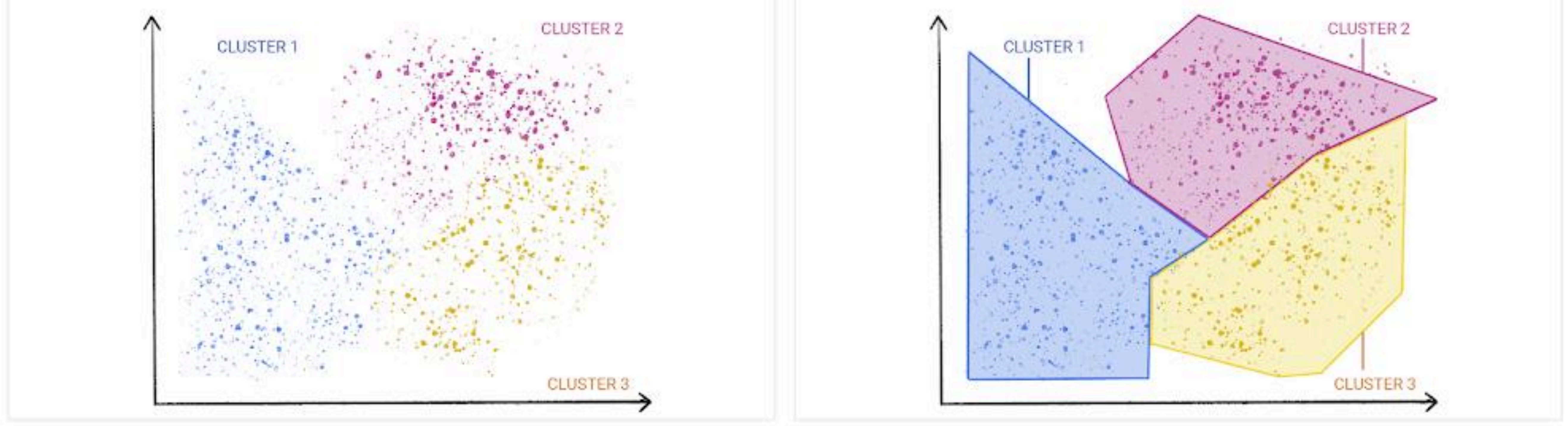



**Figure 1**. An ML model clustering similar data points.

**Figure 2**. Groups of clusters with natural demarcations.

# For language?

Traditional methods not very helpful here!

So, we make our own labels

**Masked Language Modeling**

“The quick brown fox jumps over the lazy dog”

- “The quick [MASK] fox jumps over the lazy dog”
- “The quick brown [MASK] jumps over the lazy dog”

# For LLMs

Masked Language Modeling is a helpful start, but how do we best use it?

**RAG with LLM as “Information Refiner”**

**Three approaches:**

- Prompt contains all knowledge to answer Q; LLM has to pick it out
- Prompt has incomplete or incorrect knowledge
- Prompt has relevant knowledge, but no direct part of answer

**E.g.:** “What is Anote.ai and where are its headquarters?”

- “Anote.ai is a machine learning company helping users finetune LLMs with their data located in New York City”
- “Anote.ai is a technology company based in Boston”
- “Anote.ai was founded by Natan Vidra and Thomas Clifford”

Shicheng Xu, Liang Pang, Mo Yu, Fandong Meng, Huawei Shen, Xueqi Cheng, and Jie Zhou. 2024. Unsupervised Information Refinement Training of Large Language Models for Retrieval-Augmented Generation. arXiv preprint arXiv:2402.18150 (2024).

# Making the data

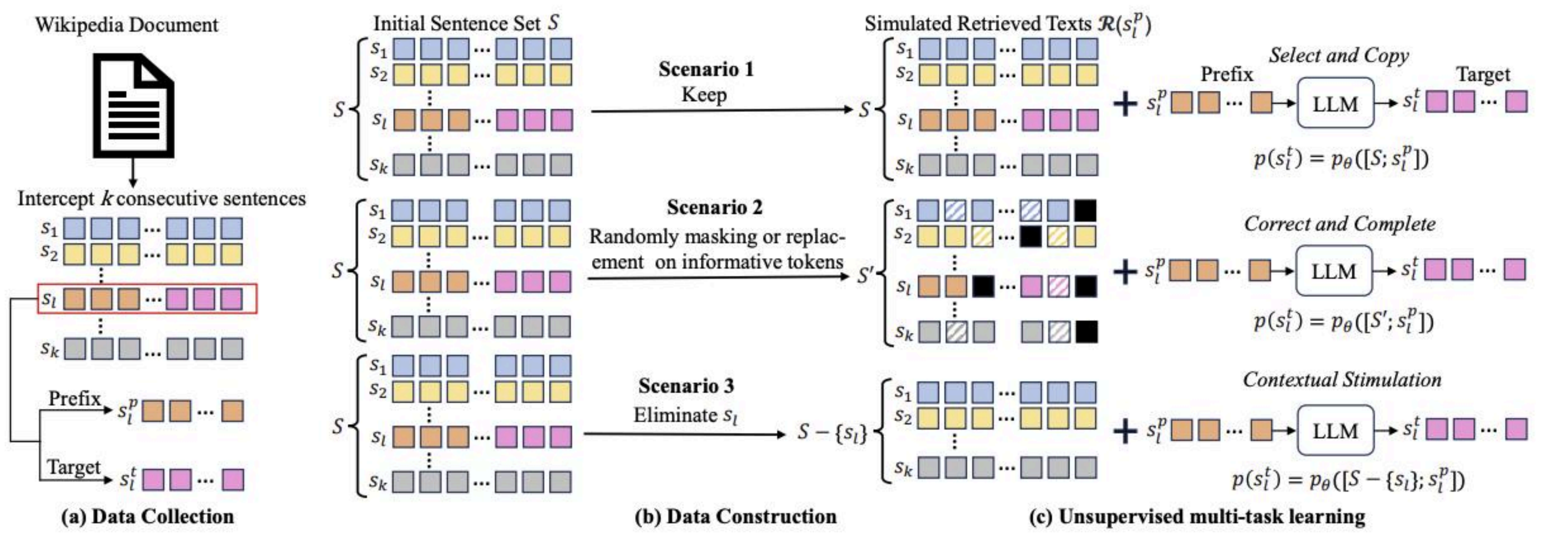



Figure 2: Overview of our INFO-RAG. Each sample is only processed for a single scenario to avoid data leakage.

Shicheng Xu, Liang Pang, Mo Yu, Fandong Meng, Huawei Shen, Xueqi Cheng, and Jie Zhou. 2024. Unsupervised Information Refinement Training of Large Language Models for Retrieval-Augmented Generation. arXiv preprint arXiv:2402.18150 (2024).

# Low Rank Adaptation (LoRA) for PEFT

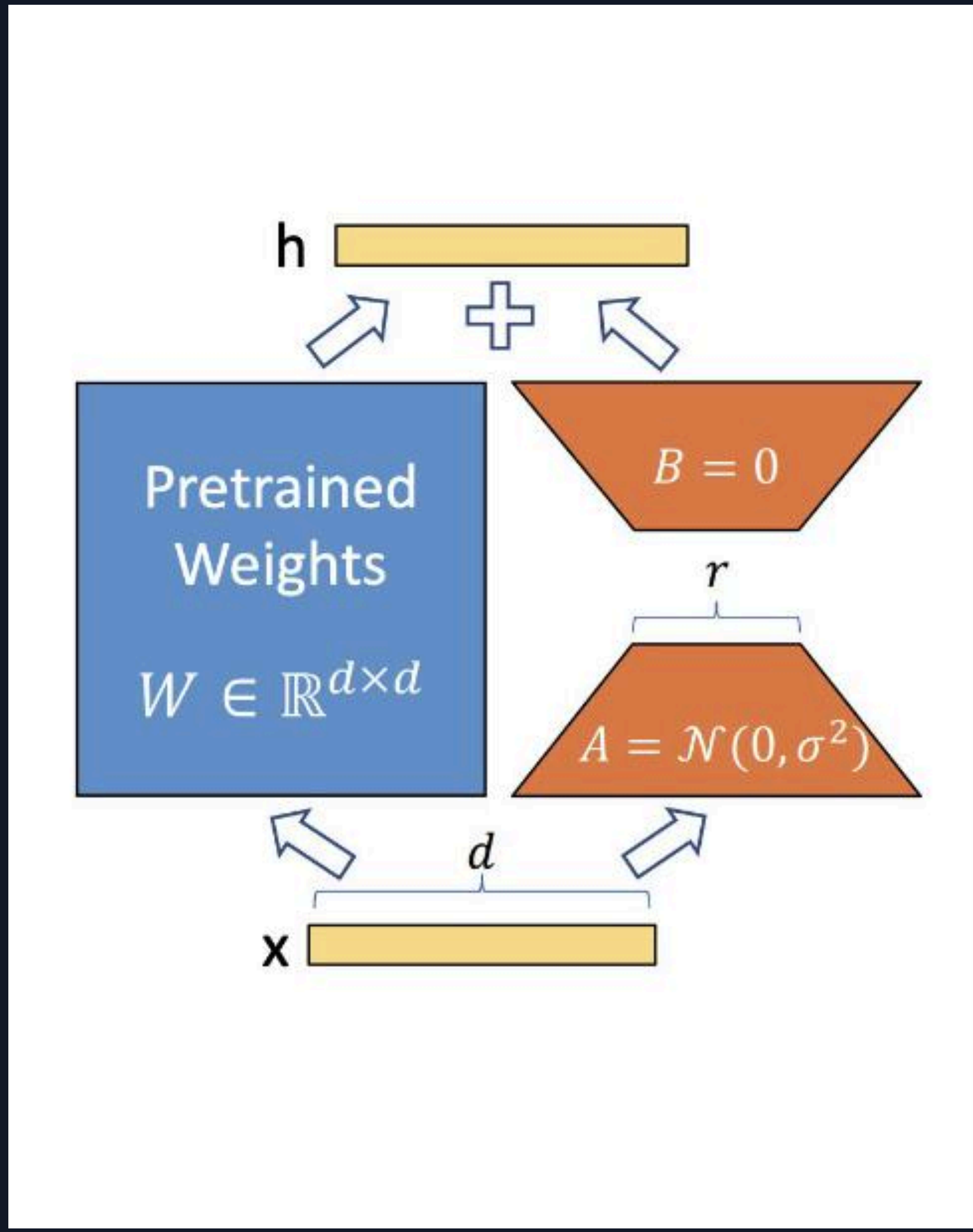
Reduces number of trainable parameters

Identifies crucial parameters for the task at hand and finetunes those

During fine-tuning, only the parameters in low-rank matrices are updated

Less chance of overfitting since only a few parameters are updated

Reduces computational and memory requirements needed to fine-tune



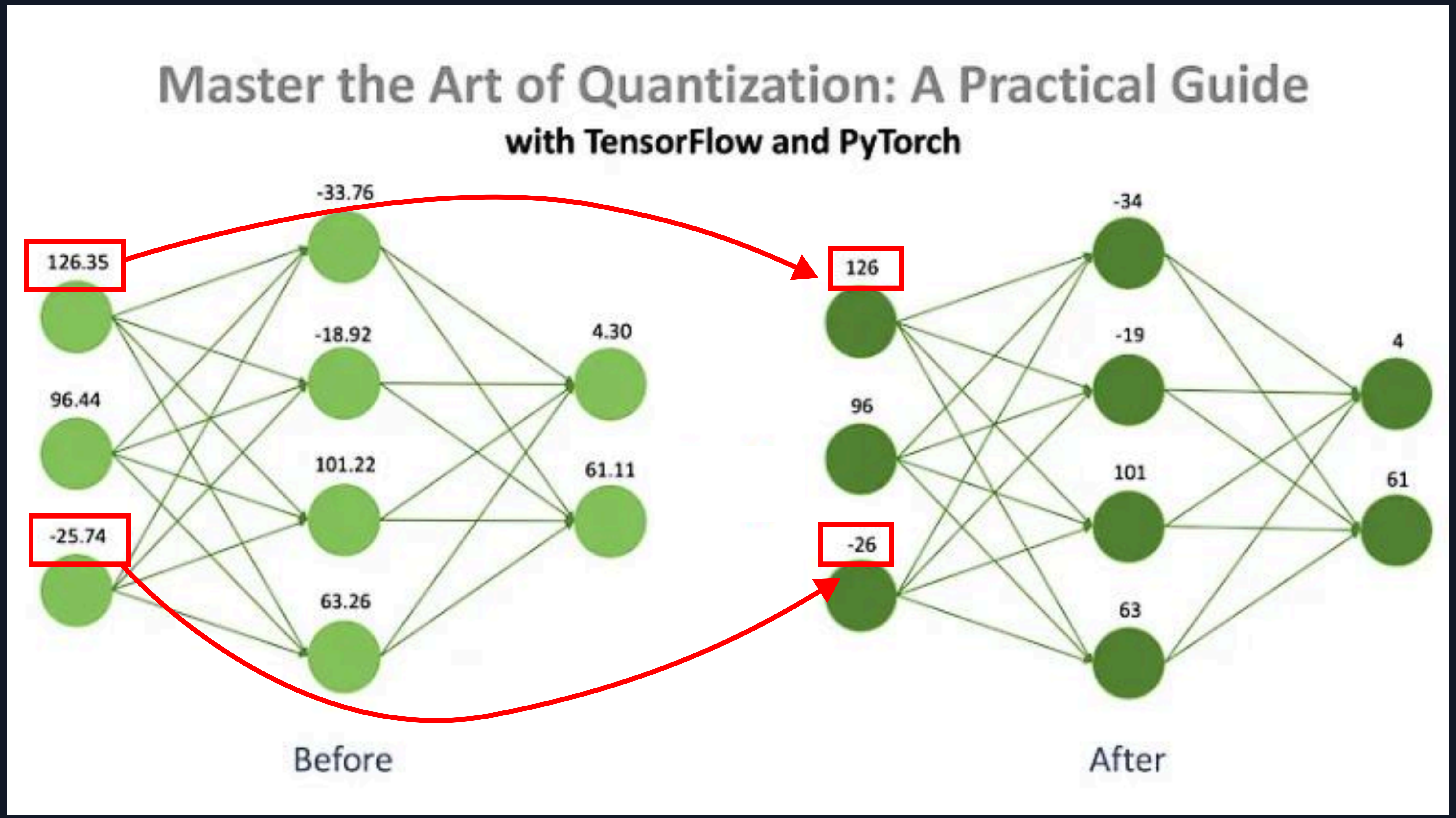
# QLoRA for Quantized LLMS

Quantization: process of reducing the numerical precision of a model's tensors to make it faster and more compact

QLoRA combines quantization & low-rank adaptation

Model parameters are first quantized (usually to 4 bit precision) and then go through LoRA

Made fine tuning a lot more accessible



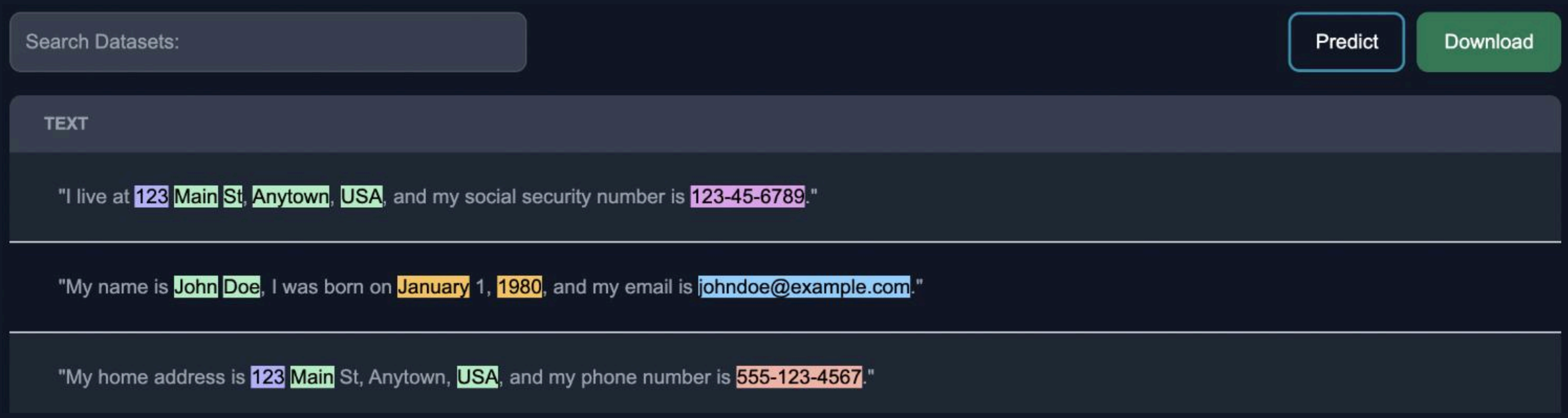
# Product Demo

# Named Entity Recognition

Anote Presented by Henry Toll

# What is Named Entity Recognition?

Named Entity Recognition (NER) is a natural language processing (NLP) technique that aims to locate and classify named entities in text into predefined categories. Named entities refer to real-world objects such as persons, locations, organizations, dates, and more.



# Pre-Trained NER Models

### SpaCy

SpaCy is an open-source library for advanced NLP in Python

### Stanford NER

Stanford NER is developed by the Stanford NLP Group

### Flair

Flair is a simple NLP library that allows you to use pre-trained models for a variety of tasks, including NER.

# The Problem

For Pre-Trained NER models, it is really difficult to obtain accurate results for domain specific entity recognition tasks, which is critically important for business use cases such as identifying PII in text data to ensure privacy

# The Goal

Label Data to train LLM models that can accurately predict entities on your own custom datasets, in a way that we can evaluate

# Evaluation

## Precision

The accuracy of the entities predicted by the model

## Recall

The ability of the model to find all relevant entities

## F1 Score

A harmonic mean of precision and recall

## IOU

Intersection over Union measures the number of correctly predicted characters divided by the total characters

# Product Demo

# Demo of Software Development Kit

Anote Presented by Spurthi Setty

# Evaluation Metrics

Anote Presented by Harsh Thakkar

## Retrieval Accuracy

Getting correct evidence text for answer

## Answering Accuracy

Getting correct answer

## Structured Metrics

Evaluating vs. Ground Truth Label

## Unstructured Metrics

Evaluating without Ground Truth Label

# Structured Answer Accuracy Metrics

| Metrics | Description | Example of Calculation |
|---|---|---|
| `LLM eval` | This metric serves as a substitute for human evaluation, where we can prompt a model like GPT-4 to see if two answers have the same semantic meaning, and prompt it to assign a specific score | Use GPT-4 to evaluate the semantic similarity between "The sky is clear" and "It's a cloudless day" and assign a score. |
| `Cosine Similarity` | This is a more automated way of comparing semantic meaning, however relies on both answers being extremely similar in order to have a high score | Calculate the cosine similarity of the TF-IDF vectors for the sentences "I enjoy reading books" and "Reading books is enjoyable". |
| `Rouge-L Score` | This metric is based on the longest common subsequence (LCS) between our model output and reference | Calculate the Rouge-L score by finding the LCS of "The cat is sleeping on the mat" and "A cat sleeps on a mat". |
| `Bleu Score` | This metric compares how similar two texts are as a number between 0 and 1. Generally a score of at least 0.6 means that two texts are similar enough to mean the same thing. | Calculate the Bleu Score for machine translated text compared to a human reference translation to assess quality. |

# Structured Retrieval Metrics

| Metrics | Description |
|---|---|
| `document level` | This metric checks if retrieved chunk is on the same document in the document as the actual chunk |
| `page level` | This metric checks if retrieved chunk is on the same page in the document as the actual chunk |
| `paragraph level` | This metric checks if retrieved chunk is on the same paragraph in the document as the actual chunk |
| `multi-chunk level` | This metric checks if multiple retrieved chunk are found in the same place in the document as the actual chunks |

# Aggregate Metrics

## Evaluation Metric Scores

| Score | Fine-tuned Model | Claude | Open AI |
|---|---|---|---|
| Cosine Similarity Score | 0.778 | 0.821 | 0.621 |
| Rouge-L | 0.824 | 0.901 | 0.780 |
| LLM Evaluation Score | 0.824 | 0.901 | 0.780 |

+ ADD A EVALUATION SCORE

# Row Specific Metrics

**Key Metric Scores**

| | | Cosine Similarity Score | | | Rouge Similarity Score | |
|---|---|---|---|---|---|---|
| Question | Human Answer | Fine-tuned | Claude | Open AI | Fine-tuned | Claude |
| Questions | Answer | C | D | E | | |
| What is the total amount of the invoice? | $22,500.00 | 0.45 | 0.42 | 0.44 | 0.45 | 0.42 |
| What is the invoice number? | #0001 | 0.72 | 0.71 | 0.68 | 0.67 | 0.68 |
| What is a list of the items being purchased? | Front End Engineering Service; •Back End Er | 0.21 | 0.18 | 0.18 | 0.47 | 0.21 |
| What is the name of the contact for question? | Bia Hermes | 0.88 | 0.74 | 0.83 | 0.63 | 0.52 |
| What is the PO number? | #1000 | 0.33 | 0.28 | 0.28 | 0.59 | 0.33 |
| When is payment due? | within 30 days of 01/01/2022 | 0.59 | 0.52 | 0.54 | 0.59 | 0.50 |
| What is the subtotal amount? | 4150 | 0.17 | 0.08 | 0.12 | 0.25 | 0.17 |
| What is the total amount? | 4565 | 0.96 | 0.88 | 0.87 | 0.69 | 0.66 |
| | | 0.24 | 0.18 | 0.21 | 0.46 | 0.24 |

# Unstructured Answer Accuracy Metrics

| Metrics | Description |
| --- | --- |
| `Faithfulness` | This metrics evaluates whether the answer is supported by the given context, and penalizes the model if it hallucinated information not supported by the text. |
| `Answer Relevance` | This metric evaluates whether or not the answer actually addresses the question. It does not account for accuracy, but penalizes for incomplete/redundant answers |

# Identifying Mislabels

### Identifying Mislabels

| Document Name | Human Label | Model Prediction |
|---|---|---|
| Doc1.txt | Documentation | Feature |
| Doc2.pdf | Feature | Task |
| Doc5.pptx | Feature | Bug |
| Doc23.csv | Bug | Task |
| Doc47.xlsx | Task | Feature |

# Classification Report

## Classification Report Metrics

| Category | MPC Accuracy | F1 | Precision | Recall | Support |
|---|---|---|---|---|---|
| Bug | 0.978 | 0.778 | 0.821 | 0.621 | 10 |
| Task | 0.924 | 0.824 | 0.901 | 0.780 | 10 |
| Documentation | 0.846 | 0.946 | 0.702 | 0.924 | 10 |
| Feature | 0.945 | 0.776 | 0.765 | 0.924 | 10 |
| Average/Total | 0.987 | 0.876 | 0.965 | 0.824 | 10 |

# Confusion Matrix

# Product Demo

# Demo of Model Versioning

Anote Presented by Tian Jin

# AI Assisted RFP Proposals

Anote Presented by Sanya Mahajan

# Problem

Filling out applications is a standard but necessary procedure in everyone's everyday lives but is an extremely tedious and time consuming process.

# RFP Response Problem

Startups want non-dilutive funding, but don't have the time and resources to apply to all these grant opportunities by hand. The current process is very tedious and manual, but is important to obtain funding.

### Customization Effort

High effort needed to align proposals with unique business contexts and niche problems.

How to stand out?

### Volume of Documents

Writing numerous documents including supporting materials and previous proposals.

How to decrease the number of teams involved?

# Grants Problem

**150**
hours spent

**10-50**
pages per grant to fill

**10-20%**
success rate

### Number of Grants

Thousands of grants from federal government and  globally are available to various sectors including education, healthcare, technology, and the arts.

### Competition

High competition due to the large number of applicants seeking funding for diverse projects.

# Grant Submission Requirements

### Tailored Applications

Each grant has specific guidelines and objectives.

### Business Context

Highlight your business's relevant past experiences and successes.

Distribution of Time Spent on Grants

# Product Demo