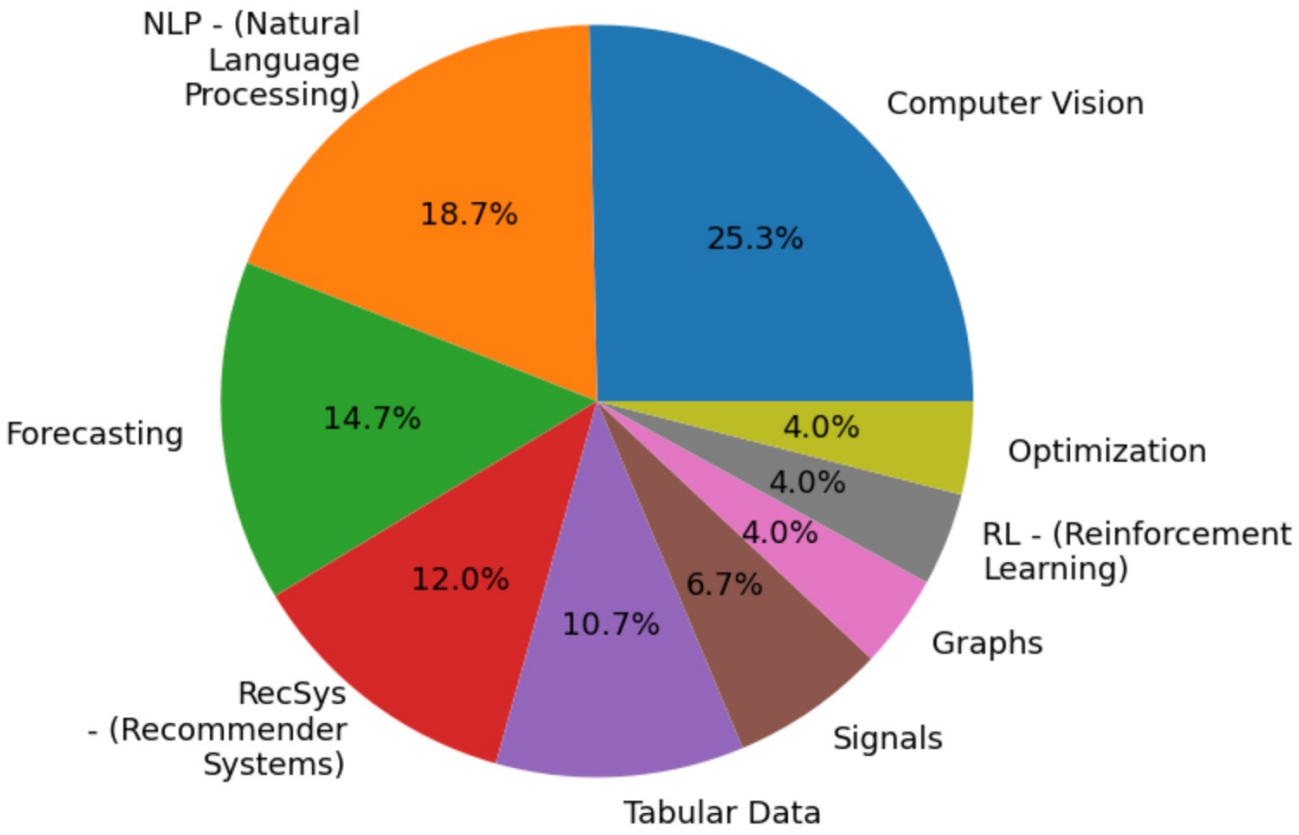
Growing with AI

Lessons learned from competing in 75 international data science competitions

Dr. Chris Deotte
Senior Data Scientist NVIDIA
Quadruple Kaggle Grandmaster

75 Competitions in 5 Years



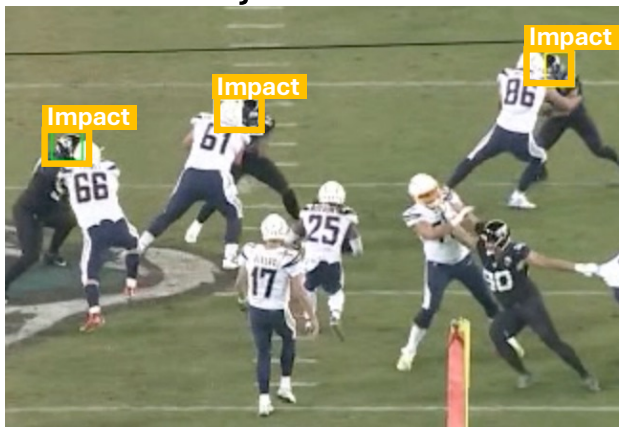
source: https://www.kaggle.com/cdeotte

Today's Talk Overview

- Discuss different categories of AI.
- Deep Dive into KDD Cup 2024 1st Place Solution.
- Amazon's Multi-Task Online Shopping Challenge for LLMs.

Computer Vision (Image/Video)

Locate Object



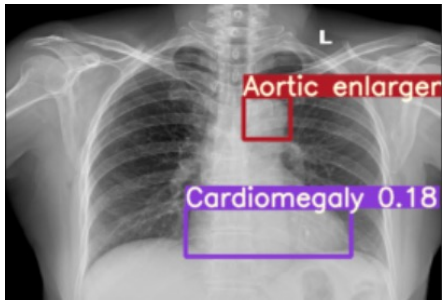
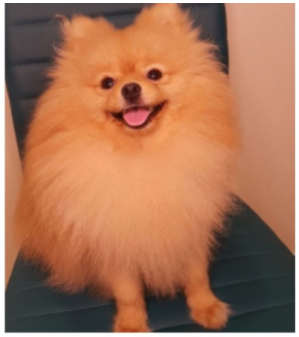


Image Regression



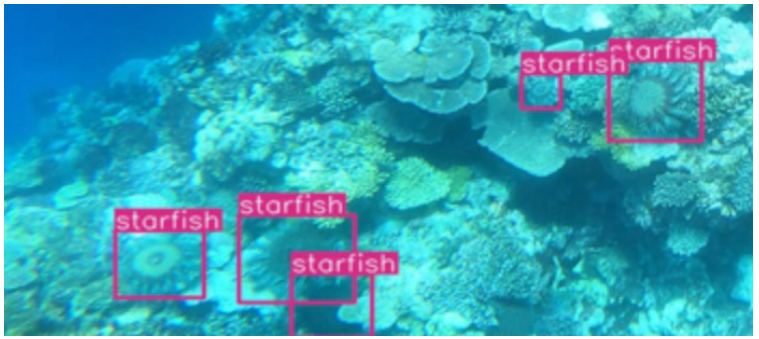
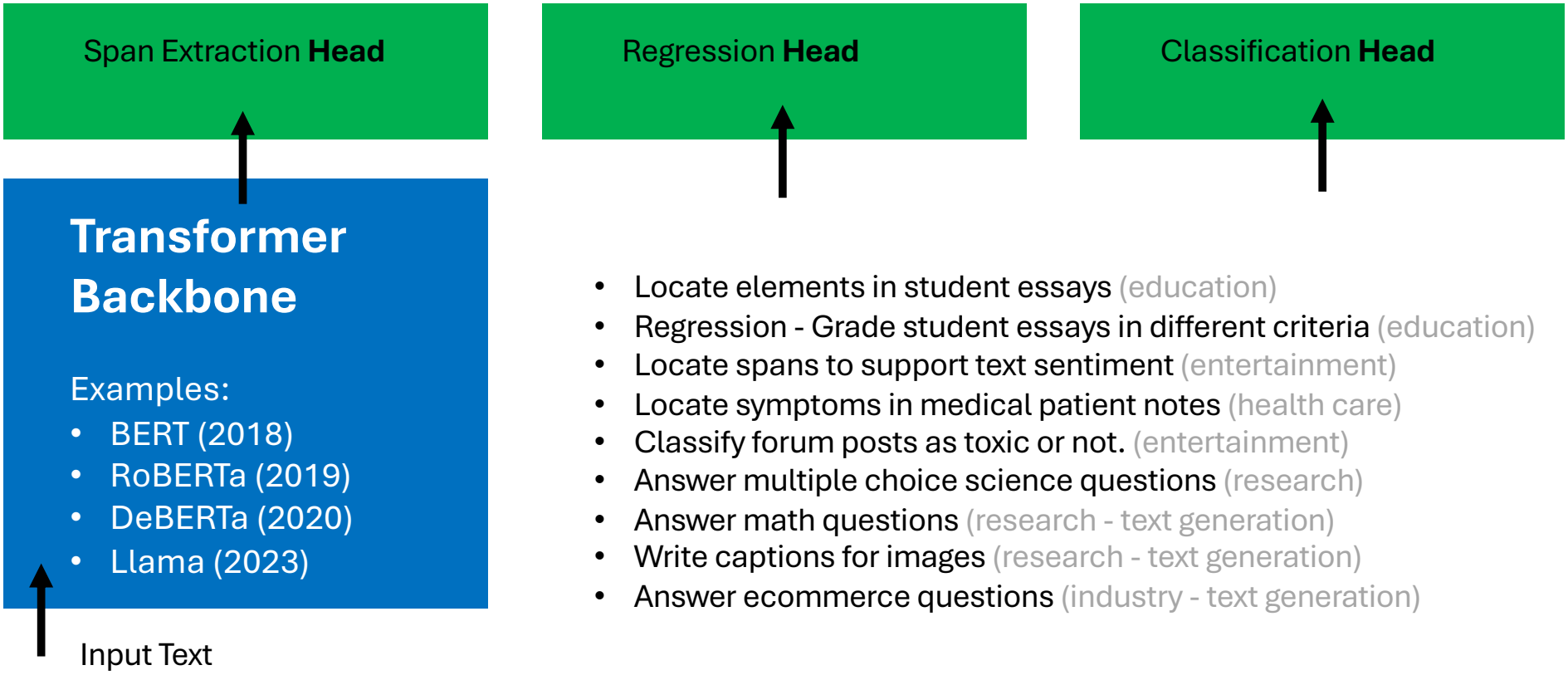


Image Classification

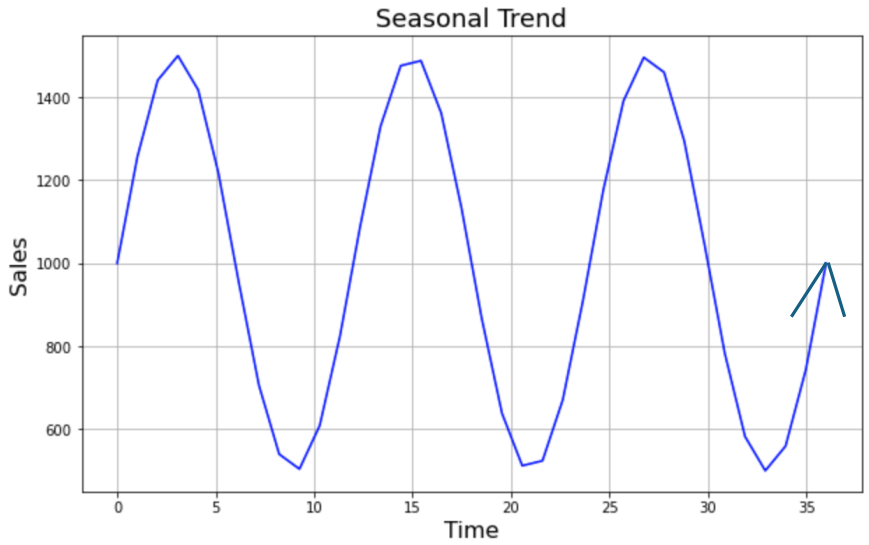
SVM → CNN 2012 → EfficientNet 2019 → Transformers 2020 → ViT or Swin → Text to Image 2021 → DALL-E 2 2022

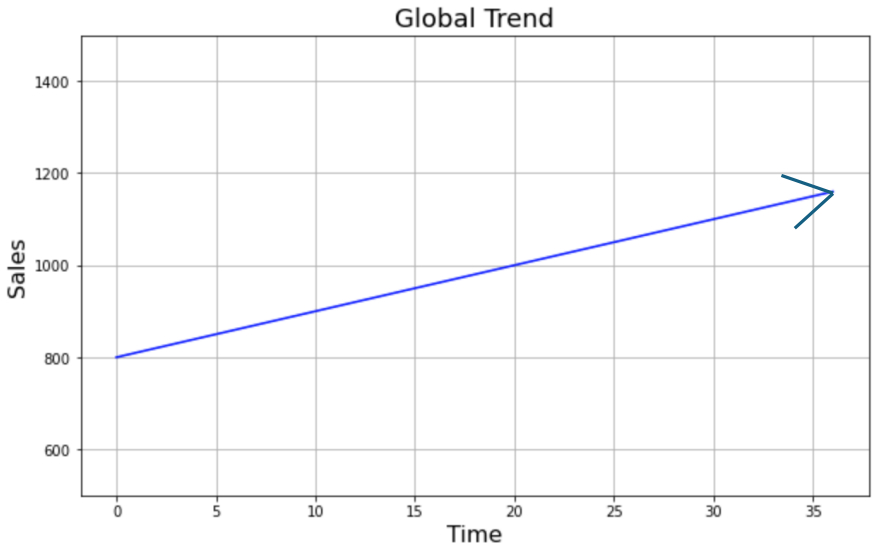
Natural Language (Text)



- Locate elements in student essays (education)
- Regression - Grade student essays in different criteria (education)
- Locate spans to support text sentiment (entertainment)
- Locate symptoms in medical patient notes (health care)
- Classify forum posts as toxic or not. (entertainment)
- Answer multiple choice science questions (research)
- Answer math questions (research - text generation)
- Write captions for images (research - text generation)
- Answer ecommerce questions (industry - text generation)

Forecasting (Time Series)





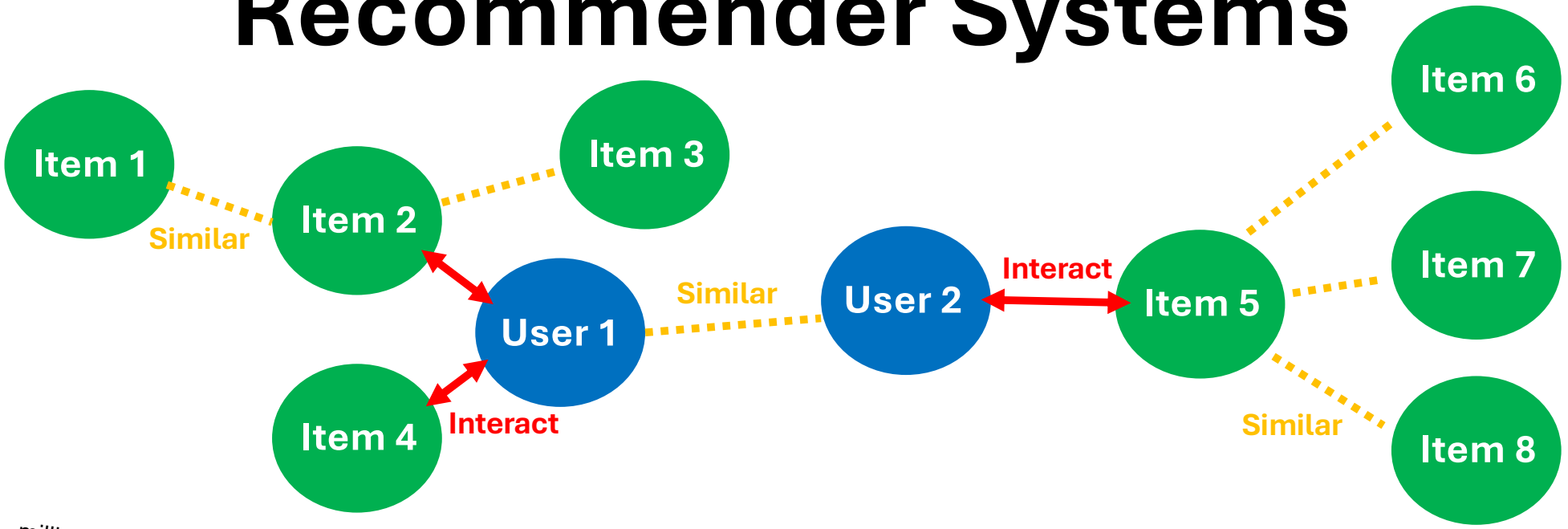
FUTURE

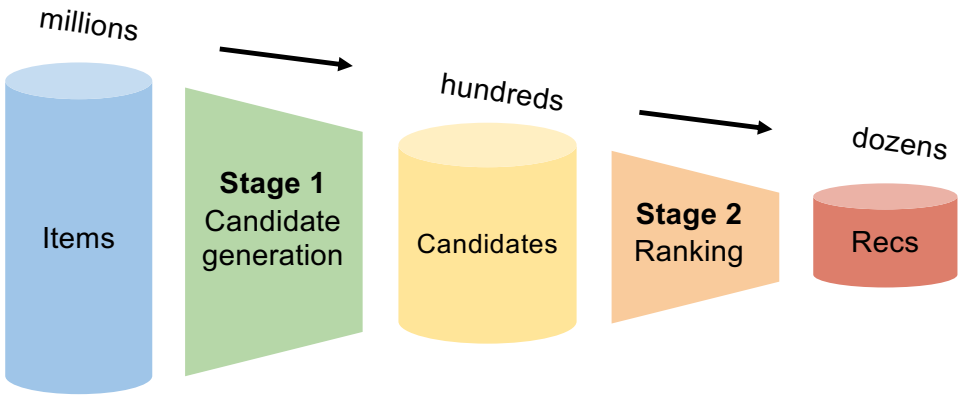
Techniques

- Statistics (i.e., ARIMA)
- ML (i.e., Regression or GBDT)
- DL (i.e., RNN or TFT transformer)

- Microbusiness Density Forecasting (**industry**)
- Parkinson Disease Progression Prediction (**health care**)
- Predict Student Performance from Game Play (**education**)
- March Madness Mania (**sports**)
- Walmart M5 Forecasting (**industry**)
- Market Prediction (**finance**)
- Identifying Age Related Conditions (**health care**)

Recommender Systems





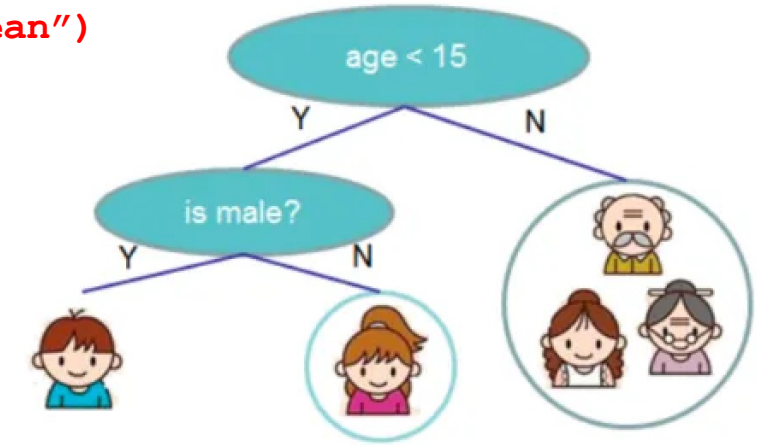
- RecSys 2020 - Predict User Engagement - **Twitter**
- Predict Travel Destinations - **Booking.com**
- RecSys 2022 - Predict Fashion Purchases - **Dressipi**
- Predict User Engagement - **OTTO**
- Product Matching - **Shopee**
- KDD Cup 2023 - Predict User Engagement - **Amazon**
- KDD Cup 2024 - Create Ecommerce Assistant - **Amazon**

Tabular Data (dataframes)

```
df["X2"] = df.groupby("Item").Price.transform("mean")
```

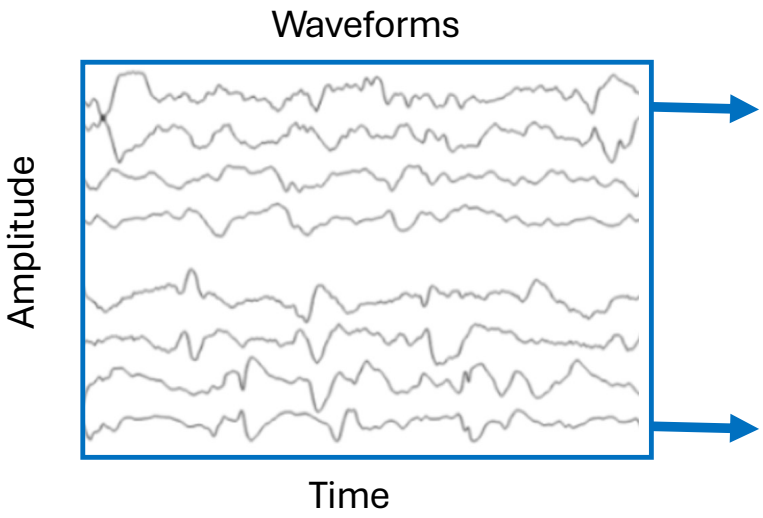
Item	Price	X1	X2	Target
Gas	$56	0	**75**	0
Gas	$80	0	**75**	0
Food	$32	1	25	0
Hotel	$550	2	550	1
Food	$20	1	25	0
Clothes	$90	3	90	0
Gas	$90	0	**75**	0
Food	$25	1	25	0

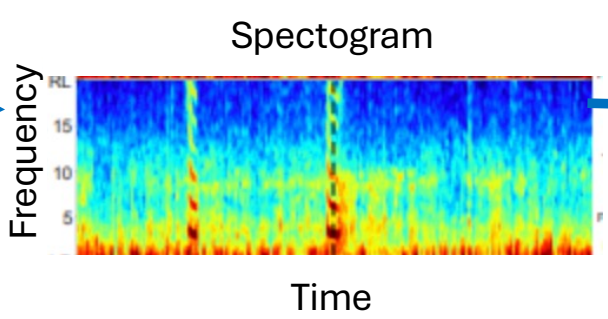
```
df["X1"],_ = df.Item.factorize()
```



- Fraud Detection (finance - **IEEE-CIS / Vesta Corp**)
- Malware Detection (industry - **Microsoft**)
- Default Prediction (finance - **AMEX**)
- Transaction Prediction (finance - **Santander**)
- Predict Brain assessment (health care)
- Predict Drug Outcome (drug discovery)
- Predict Used Car Price (industry)

Signals (Time Sampled Data)





2D-Image model

CNN Block

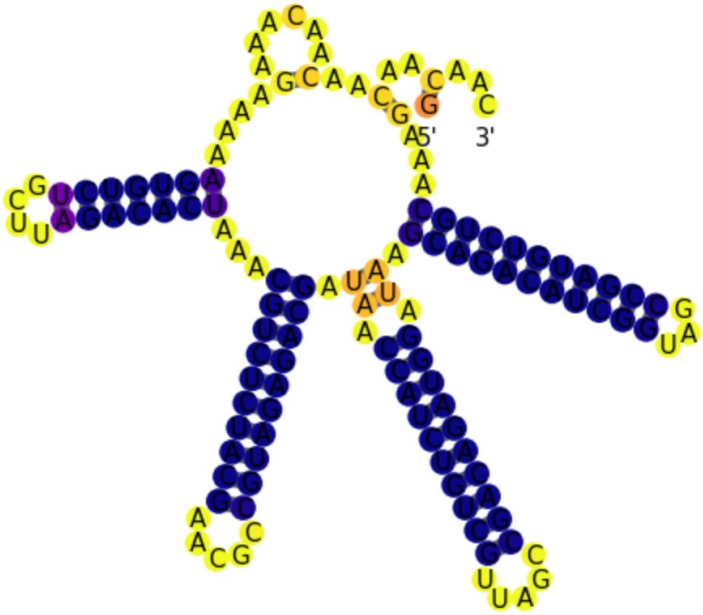
Transformer Block

RNN Block

1D-Wave model

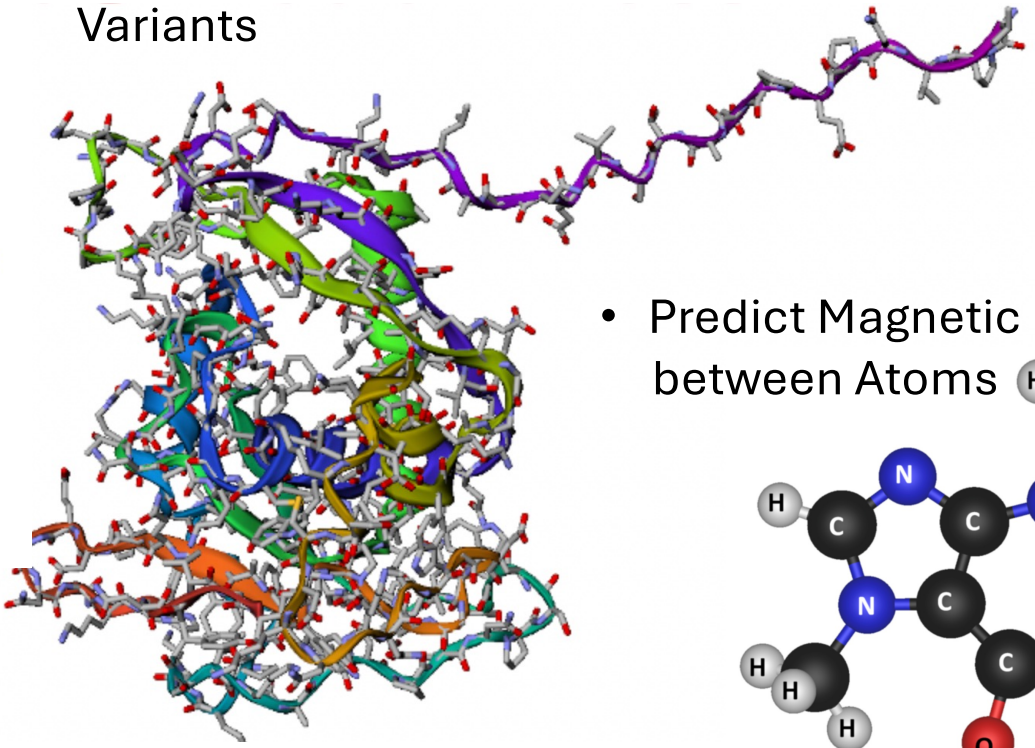
- Detect Biological Ion Switching
- Classify Harmful Brain Activity
- Detect Rainforest Animal Species
- Detect Sleep States
- Control Ventilator Pressure

Graphs (Edges & Vertices)



- Predict Degradation Rates of RNA Molecules

- Predict Thermostability of Enzyme Variants



- Predict Magnetic Interaction between Atoms

amazon KDD Cup 2024

Multi-Task Online Shopping Challenge for LLMs

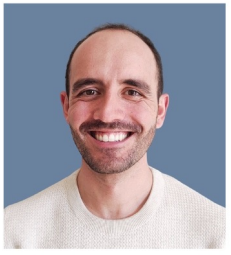
Ahmet Erdem 🇹🇷

Benedikt Schifferrer 🇩🇪

Chris Deotte 🇺🇸

Gilberto Titericz 🇧🇷

Ivan Sorokin 🇫🇮



Simon Jegou 🇫🇷

Specifically, Tracks 1-4 carry the following prizes:

- 🥇 **First place:** $2,000
- 🥈 **Second place:** $1,000
- 🥉 **Third place:** $500
- **4th-7th places:** AWS Credit $500
- 🏅 **Student Award:** $750

Track 5 (all-around) carries the following prizes:

- 🥇 **First place:** $7,000
- 🥈 **Second place:** $3,500
- 🥉 **Third place:** $1,500
- **4th-8th places:** AWS Credit $500
- 🏅 **Student Award:** $2,000

source: https://www.aicrowd.com/challenges/amazon-kdd-cup-2024-multi-task-online-shopping-challenge-for-llms

Winning Amazon KDD Cup'24

Chris Deotte*
cdeotte@nvidia.com
NVIDIA
USA

Ivan Sorokin*
isorokin@nvidia.com
NVIDIA
Finland

Ahmet Erdem*
aerdem@nvidia.com
NVIDIA
Türkiye

Benedikt Schifferer*
bschifferer@nvidia.com
NVIDIA
Germany

Gilberto Titericz Jr*
gtitericz@nvidia.com
NVIDIA
Brazil

Simon Jegou*
sjegou@nvidia.com
NVIDIA
France

Abstract

This paper describes the winning solution of all 5 tasks for the Amazon KDD Cup 2024 *Multi Task Online Shopping Challenge for LLMs*. The challenge was to build a useful assistant, answering questions in the domain of online shopping. The competition contained 57 diverse tasks, covering 5 different task types (e.g. multiple choice) and across 4 different tracks (e.g. multi-lingual).

Our solution is a single model per track. We fine-tune Qwen2-72B-Instruct on our own training dataset. As the competition released only 96 example questions, we developed our own training dataset by processing multiple public datasets or using Large Language Models for data augmentation and synthetic data generation. We apply *wise-ft* to account for distribution shifts and ensemble multiple LoRA adapters in one model. We employed *Logits Processors* to constrain the model output on relevant tokens for the tasks. *AWQ 4-bit Quantization* and *vLLM* are used during inference to predict the test dataset in the time constraints of 20 to 140 minutes depending on the track.

Our solution achieved the first place in each individual track and is the first place overall of Amazon's KDD Cup 2024.

CCS Concepts

• **Computing methodologies** → **Natural language generation**; *Machine translation*; *Information extraction*.

Keywords

Large Language Models, LLM, Shopping Assistant, KDD Cup, Multi Task Learning, Multi-Lingual

ACM Reference Format:
Chris Deotte, Ivan Sorokin, Ahmet Erdem, Benedikt Schifferer, Gilberto Titericz Jr, and Simon Jegou. 2024. Winning Amazon KDD Cup'24. In *Proceedings of* . ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

*All authors contributed equally to this research.

1 Introduction

The capabilities of Large Language Models (LLMs) have significantly improved in the last years and they have become popular due to their easiness to use. Users can interact with the systems in natural language. The LLMs excel on a variety of tasks, such as general reasoning, math questions, coding, etc. Many systems are getting updated by adding a LLMs to make them easier to use and/or providing more functionality. (Online) shopping is a large domain with billions of users and high economic output. The Amazon KDD Cup 2024 [2] is designed to evaluate LLMs to be a useful shopping assistant.

Input	The product 'Hanes Men's Beefy-T T-Shirt, Heavyweight Cotton Tee, 1 Or 2 Pack, Big & Tall' appears on an e-commerce website. What type of fabric is used in it? 0. spandex, polyester 1. cotton 2. microfiber 3. It cannot be inferred. Answer:
Answer	1

Figure 1: One example of the development dataset. It is a multiple choice question answering tasks for understanding shopping concepts.

Amazon developed an evaluation dataset *ShopBench*, containing approx. 20,000 questions across 57 different tasks covering 5 task types (e.g. retrieval), to test LLMs capabilites in the online shopping domain (see an example in Figure 1). The competition had 5 different tracks, which evaluates different aspects such as shopping knowledge understanding or user behavior alignment. The 5th track was the overall track containing all 20,000 questions. The competition was organized as a code competition in which participants have no access to the ShopBench dataset and instead they have to submit their model.

Our team from NVIDIA won all 5 tracks (see Table 3). This paper describes our final solution and an ablation study on our

KDD Cup Sample Questions

Track: 1 - Understanding Shopping Concepts
Task: 1 out of 57
Task Type: **Generation**
Metric: Cosine Similarity

Question: Explain the product type Water Purification Unit
Answer: "A water purification unit removes impurities by lowering contamination of water using a fine physical barrier, a chemical process, or a biological process."

Track: 2 - Shopping Knowledge Reasoning
Task: 8 out of 57
Task Type: **Multiple Choice**
Metric: Accuracy

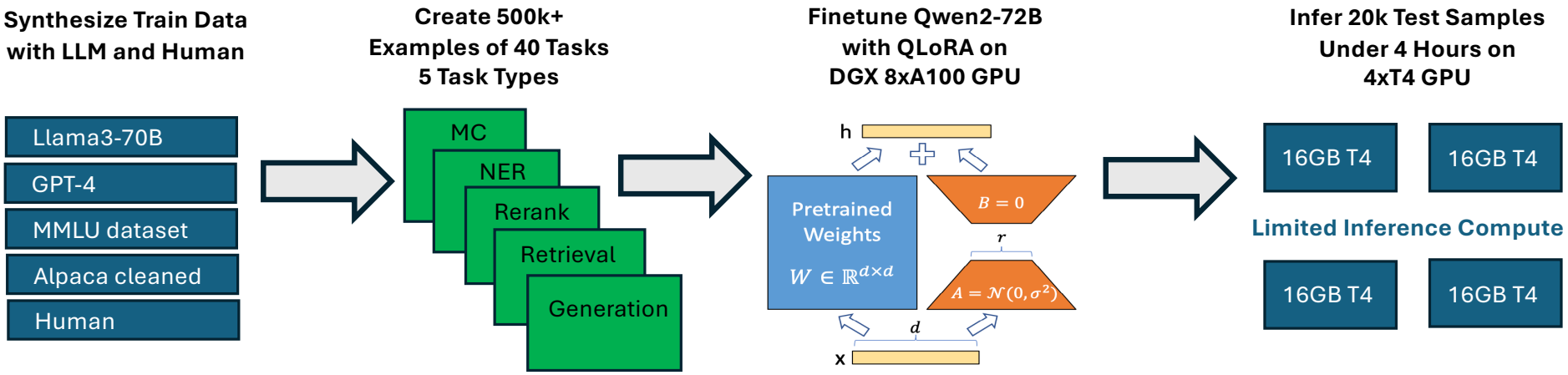
Question: The product 'Simply Asia Garlic Basil Singapore Street Noodles, 9.24 oz (Pack of 6)' appears on e-commerce website. What is the total weight of the noodles?
0. 8 ounce
1. 55.44 ounce
2. 14.19 ounce
3. 60 ounce
Answer: "1"

Track: 3 - User Behavior Alignment
Task: 14 out of 57
Task Type: **Retrieval**
Metric: Hit Rate @3

Question: A user on an online shopping website has just purchased a product 'Steven Harris Mathematics Math Equations Necktie - Red - One Size Neck Tie'. The following numbered list contains 8 products. Please select 3 products from the list that the user may also purchase. Product List:
1. Under Armour Men`s ColdGear Lite Cushion Boot Socks, 1 Pair
2. Little Angel Tasha-685E Patent Bow Mary Jane Pump (Toddler/Little Girl/Big Girl) - Fuchsia
3. Men's Solar System Planets Necktie-Black-One Size Neck Tie by
4. Wrangler Men's Big & Tall Rugged Wear Unlined Denim Jacket
5. Steven Harris Mens Smiley Face Necktie - Yellow - One Size Neck Tie
6. ComputerGear Math Formula Tie Engineer Silk Equations Geek Nerd Teacher Gift
7. Harley-Davidson Boys Baby Twin Pack Creeper My Daddy Rides a Harley Orange
8. Liverpool Football Club Official Soccer Gift Mens Crest T-Shirt
You should output 3 numbers that correspond to the selected products. There should be a comma separating every two numbers. Only respond with the results. Do not say any word or explanations.
Answer: "3,5,6"

- Train data is 96 samples
- Test data is 20,000 samples!
- 5 tracks (question categories)
- 5 task types (question types)
- 57 tasks (question templates)

Solution Overview



- **Generate Train Data**
- **Finetune LLM Efficiently**
- **Infer LLM Efficiently**

source: https://arxiv.org/abs/2408.04658

Generate Train Data

Amazon-M2: A multi-lingual Amazon session dataset with rich meta-data used for KDD Cup 2023.

Amazon Reviews 2023: A large scale Amazon Review Dataset with rich features and over 500M reviews across 33 categories.

ESCI-data: Shopping Queries dataset. Used for KDD Cup 2022.

NingLab/ECInstruct: Instruction dataset covers 116,528 samples from 10 e-commerce tasks of 4 categories.

MMLU: Massive multitask test consisting of 116k multiple-choice questions from ARC, MC_TEST, OBQA, RACE, etc.

Alpaca-Cleaned: Cleaned version of the original Alpaca Dataset released by Stanford.

	query	product_id	esci_label	product_title	product_description	product_bullet_point	product_brand	product_color
0	3wf3cb filter	B00WUWTX2W	E	Tier1 Refrigerator Water Filter Replacement fo...	The PureSource2 comparable RWF1031 (3-Pack) by...	REPLACEMENT MODEL: This is a replacement compa...	Tier1	None
1	laptop	B089M4VDK5	C	Broonel Black Invisible Lightweight Laptop Com...	None	Quick Setup: Magnets prevent the Stand from co...	Broonel	None
2	vinyl record players	B00UMVVZKG	E	Victrola Aviator 8-in-1 Bluetooth Record Playe...	None	8-in-1 ENTERTAINMENT CENTER – With vintage loo...	Victrola	Mahogany
3	outdoor grey wicker chaise lounge	B07T63FCYC	S	MAGIC UNION 2-Pack Outdoor Chairs Patio Adjust...	Lounge in comfort with the MAGIC UNION Chaise ...	Includes: 2 Chaise Lounge + 2 White Chaise Cus...	MAGIC UNION	2 Lounges

Traditional Full Finetune LLM

16x72B = 1,152 GB VRAM needed to train Qwen2-72B!

fp32 Adam Optimizer

12 x Learnable Parameters bytes needed

4x72GB VRAM $m_t = \beta_1 * m_{t-1} + (1 - \beta_1) * g_t$

4x72 GB VRAM $v_t = \beta_2 * v_{t-1} + (1 - \beta_2) * g_t^2$

4x72 GB VRAM $\theta = \theta - \left(\frac{\alpha * m_t}{\sqrt{v_t + \epsilon}} \right)$

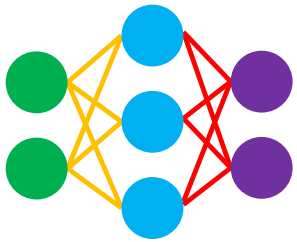
source: https://arxiv.org/abs/1910.02054

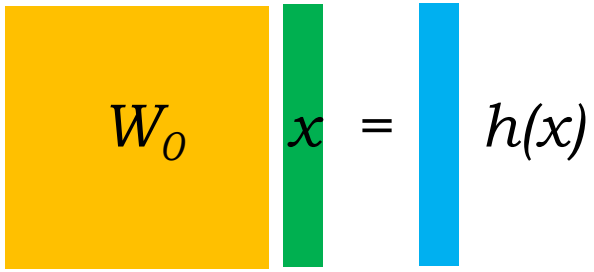
2x72 GB VRAM θ = model parameters

2x72GB VRAM g_t = model gradients

Efficient Finetune LLM - LoRA

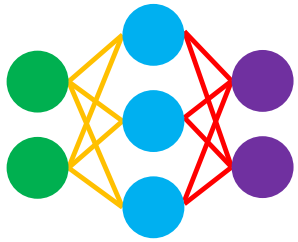
Full Fine-tuning: $h(x) = W_0 x$

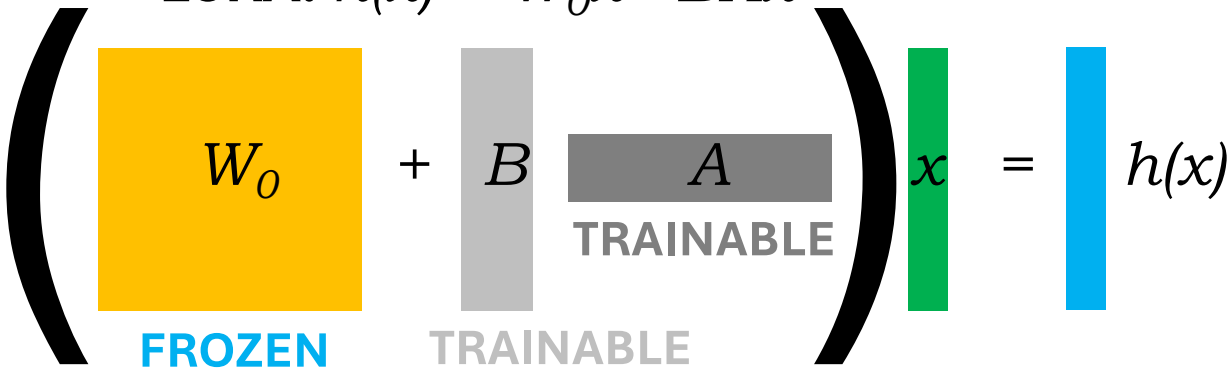




x $h(x)$ y

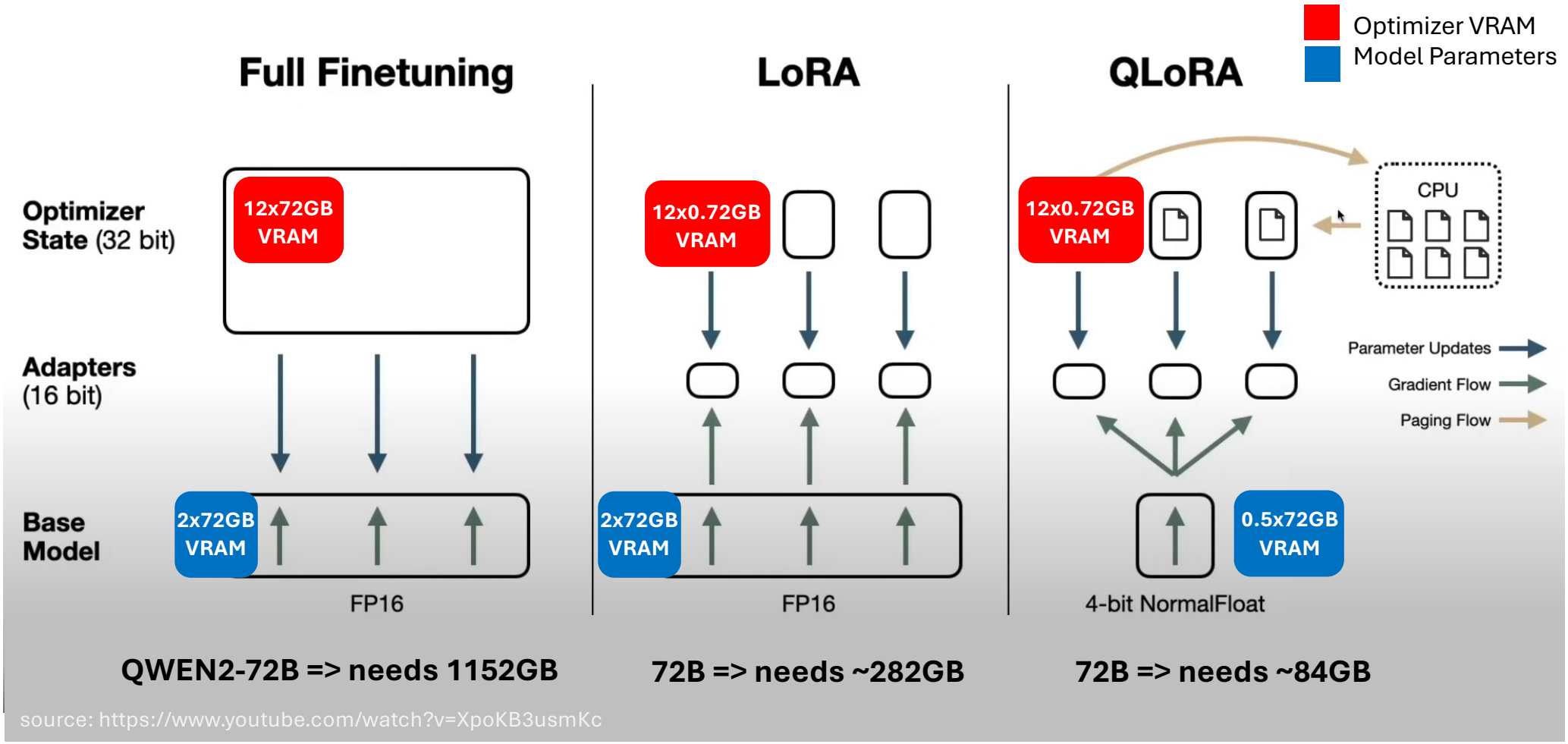
LoRA: $h(x) = W_0 x + BAx$





source: https://www.youtube.com/watch?v=XpoKB3usmKc

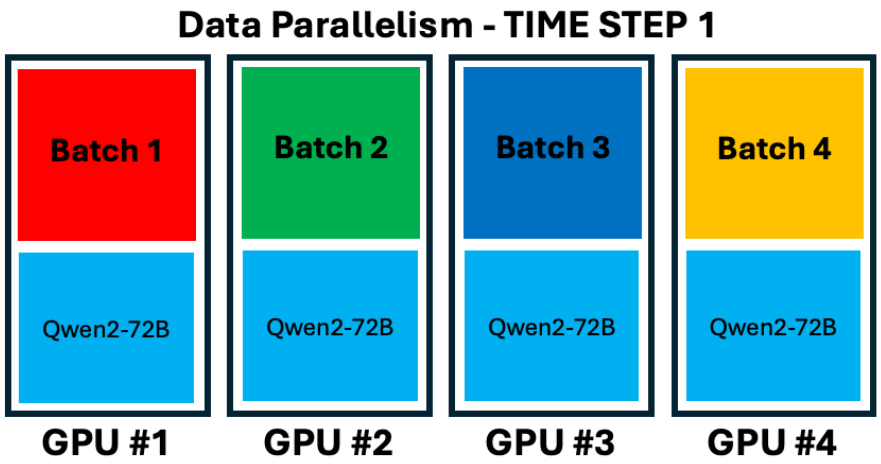
Efficient Finetune LLM - QLoRA

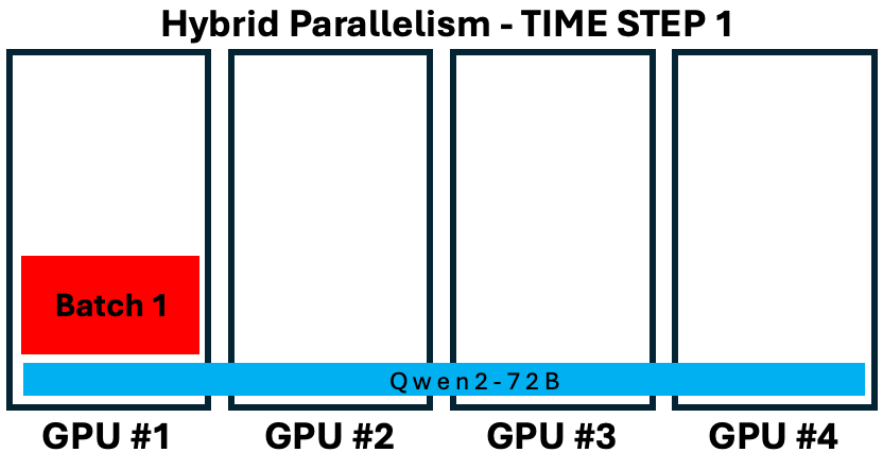


QWEN2-72B => needs 1152GB | **72B => needs ~282GB** | **72B => needs ~84GB**

source: https://www.youtube.com/watch?v=XpoKB3usmKc

Infer LLM Efficiently





- Finetune SFT with Data Parallelism on 8xA100 80GB GPUs using QLoRA using Axolotl.
- Infer with Hybrid Parallelism on 4xT4 16GB using AWQ 4bit quantization and vLLM

amazon KDD Cup 2024

Multi-Task Online Shopping Challenge for LLMs

Δ	#	Participants
•	01	Team_NVIDIA
•	02	AML_LabCityU
•	03	shimmering_as...

Source: ChatGPT-4 with DALL-E 3: Hi, I need an image that conveys artificial intelligence and excitement. We just won a data science competition and want an image for a slide. Perhaps robots with fireworks overhead. Or perhaps something with trophies and medals. Or any ideas you have. Be creative and make it colorful. Thanks.