AI and Unstructured Data for Measurement and Estimation

Lecture 2: Finetuning LLMs

Stephen Hansen
University College London

Introduction

In the last lecture, we discussed how large language models solve word prediction problems via attention operations.

This gives us excellent predictive models for $w_n \mid \mathbf{w}_{-n}$.

Autoregressive language models are specialized at the prediction task $w_N \mid \mathbf{w}_{-N}$ and are used to generate language.

However, in most applications the prediction functions are only the starting point for NLP tasks.

In this lecture, we discuss how LLMs can be adapted for specific tasks.

Outline

1. Self-Supervised Finetuning
2. Supervised Finetuning
3. Instruction Finetuning
4. Reinforcement Learning with Human Feedback

[Jurafsky and Martin, 2025] chapter 9.

Lightcast Case Study

To discuss some of these themes, we'll draw on data on job postings provided by Lightcast.

Data used in [Hansen et al., 2023] to measure indidence of remote work across countries, cities, occupations, and firms.

Measures built using LLMs and other information retrieval algorithms.

Such benchmarking exercises are arguably not as common in economics as they should be.

Structure of Dataset

Table: Counts of Vacancy Postings, Employers, and Cities, January 2014 to January 2023

(1)	(2)	(3)	(4)
Country	Vacancies	Employers	Cities
New Zealand	1,700,523	36,201	67
Australia	8,607,160	197,870	59
Canada	11,711,357	712,577	3,691
United Kingdom	74,576,747	876,103	2,268
United States	161,872,915	3,485,630	31,635
Total	258,468,702	5,308,381	37,720

Note: Reported counts pertain to the universe of online postings from January 2019 onwards and a 5% random sample from 2014 to 2018, after we drop about 6% of the postings in the data-cleaning steps described in Appendix A. We rely on Lightcast's proprietary algorithm to identify employers and cities.

Software Developer

Pearson 2,739 reviews
Australia
Remote
Full-time
You must create an Indeed account before continuing to the company website to apply

Apply on company site

Our purpose: At Pearson we 'add life to a lifetime of learning' so everyone can realise the life they imagine. We do this by creating vibrant and enriching learning experiences designed for real-life impact.

Our company: Pearson was founded in 1844 and has been built on our ability to grow with and adapt to a constantly evolving market. Our 20,000+ employees are dedicated to creating the high-quality, digital-first, accessible and sustainable resources for lifelong learning.

Flexible working: Pearson is committed to hybrid working practices. When you are not working from home, you'll be based in our Nunawading office that has free parking and is walking distance to 2 train station. This is a great location for those that are not a fan of the city commute.

The Role : As a Software Engineer, you will be joining one of our cross-functional scrum teams and will play a key role in the development of our online assessment platform. Reporting to our Engineering Manager, you will work from home and collaborate via telecommuniting platforms.

What you will do:

Expense Processor (Remote U.S.A.)

Plus Relocation 17 reviews
Minneapolis, MN 55426 • Remote
Full-time
You must create an Indeed account before continuing to the company website to apply

Apply on company site

Job details

Job Type:
Full-time
Work From Home:
Allowed

Location:
Anywhere

Full Job Description

Plus Relocation is looking for a numbers driven, detail

Software Developer

Pearson ★★★★☆ 2,739 reviews
Australia
Remote
Full-time
You must create an Indeed account before continuing to the company website to apply

Apply on company site

Our purpose: At Pearson we 'add life to a lifetime of learning' so everyone can realise the life they imagine. We do this by creating vibrant and enriching learning experiences designed for real-life impact.

Our company: Pearson was founded in 1844 and has been built on our ability to grow with and adapt to a constantly evolving market. Our 20,000+ employees are dedicated to creating the high-quality, digital-first, accessible and sustainable resources for lifelong learning.

Flexible working: Pearson is committed to hybrid working practices. When you are not working from home, you'll be based in our Nunawading office that has free parking and is walking distance to 2 train station. This is a great location for those that are not a fan of the city commute.

The Role : As a Software Engineer, you will be joining one of our cross-functional scrum teams and will play a key role in the development of our online assessment platform. Reporting to our Engineering Manager, you will work from home and collaborate via telecommuniting platforms.

What you will do:

Expense Processor (Remote U.S.A.)

Plus Relocation ★★★☆☆ 17 reviews
Minneapolis, MN 55426 • Remote
Full-time
You must create an Indeed account before continuing to the company website to apply

Apply on company site

Job details

Job Type:
Full-time

Work From Home:
Allowed

Location:
Anywhere

Full Job Description

Plus Relocation is looking for a numbers driven, detail

Deputy Home Care Manager

Habitation Care Ltd
Brighton BN1
£21,246 - £26,289 a year - Full-time, Part-time, Temporary contract, Fixed term contract, Temp to perm

Apply now

We are looking for a Deputy Home Manager with domiciliary care experience to join our company. You will work from home care facilities with a strong track record of quality service.

The person we are looking for must have a positive, and a can-do work attitude at all time.

The person we are looking for must have at least 1 years working experience in a domiciliary care or care home managers role.

The role is for 38 hours per week plus on call duties, and sometimes cover of care calls would be required.

The person will be preparing supporting the Registered manager to carryout daily tasks.

Job Types: Full-time, Part-time, Fixed term, Temp to perm
Contract length: 36 months
Part-time hours: 38 per week

General Builder (Bricklayer Based)

Birkby Construction Limited
Maidstone
£14.50 an hour - Full-time, Permanent

Apply now

General Builder (bricklayer based) required for Small Works Department of Birkby Construction Limited on a PAYE basis. Company vehicle provided. Applicant must be self-motivated and confident. Willing to remote work sites.

Job Types: Full-time, Permanent

Work from Home: Not Available

Salary: £14.50 per hour

Benefits:

- Company car

Schedule:

- Monday to Friday

Licence/Certification:

- CSCS (preferred)

Deputy Home Care Manager

Habitation Care Ltd
Brighton BN1
£21,246 - £26,289 a year - Full-time, Part-time, Temporary contract, Fixed term contract, Temp to perm

Apply now

We are looking for a Deputy Home Manager with domiciliary care experience to join our company. You will work from home care facilities with a strong track record of quality service.

The person we are looking for must have a positive, and a can-do work attitude at all time.

The person we are looking for must have at least 1 years working experience in a domiciliary care or care home managers role.

The role is for 38 hours per week plus on call duties, and sometimes cover of care calls would be required.

The person will be preparing supporting the Registered manager to carryout daily tasks.

Job Types: Full-time, Part-time, Fixed term, Temp to perm
Contract length: 36 months
Part-time hours: 38 per week

General Builder (Bricklayer Based)

Birkby Construction Limited
Maidstone
£14.50 an hour - Full-time, Permanent

Apply now

General Builder (bricklayer based) required for Small Works Department of Birkby Construction Limited on a PAYE basis. Company vehicle provided. Applicant must be self-motivated and confident. Willing to remote work sites.

Job Types: Full-time, Permanent

Work from Home: Not Available

Salary: £14.50 per hour

Benefits:

- Company car

Schedule:

- Monday to Friday

Licence/Certification:

- CSCS (preferred)

Deputy Home Care Manager

Habitation Care Ltd
Brighton BN1
£21,246 - £26,289 a year - Full-time, Part-time, Temporary contract, Fixed term contract, Temp to perm

Apply now

We are looking for a Deputy Home Manager with domiciliary care experience to join our company. You will work from home care facilities with a strong track record of quality service.

The person we are looking for must have a positive, and a can-do work attitude at all time.

The person we are looking for must have at least 1 years working experience in a domiciliary care or care home managers role.

The role is for 38 hours per week plus on call duties, and sometimes cover of care calls would be required.

The person will be preparing supporting the Registered manager to carryout daily tasks.

Job Types: Full-time, Part-time, Fixed term, Temp to perm
Contract length: 36 months
Part-time hours: 38 per week

General Builder (Bricklayer Based)

Birkby Construction Limited
Maidstone
£14.50 an hour - Full-time, Permanent

Apply now

General Builder (bricklayer based) required for Small Works Department of Birkby Construction Limited on a PAYE basis. Company vehicle provided. Applicant must be self-motivated and confident. Willing to remote work sites.

Job Types: Full-time, Permanent

Work from Home: Not Available

Salary: £14.50 per hour

Benefits:

- Company car

Schedule:

- Monday to Friday

Licence/Certification:

- CSCS (preferred)

Self-Supervised Finetuning

Self-Supervized Finetuning

Job posting text may have distinct relationships among words compared to generic English: for example, ‘operations’ and ‘military’.

To obtain a more context-specific representation of language, one can simply repeat masked language modeling on a new dataset.

The obtained embeddings will reflect ‘meaning’ in the context from which the new training data comes.

Updating all the parameters in the original model can be computationally expensive so need for solutions.

Loss Function for Word Prediction

Let $p_{\boldsymbol{\theta}}(\tilde{w}_n = v \mid \mathbf{w}_{-n})$ be the conditional probability that the word at position n equals v given context $\mathbf{w}_{-n}$.

$\boldsymbol{\theta}$ is a vector of neural network parameters (initial embeddings, attention weights, etc.)

The contribution to the loss function of the prediction problem is

$$-\log p_{\boldsymbol{\theta}}(\tilde{w}_n = w_n \mid \mathbf{w}_{-n})$$

Total loss sums over all word prediction problems, estimates $\boldsymbol{\theta}$ via gradient descent.

Distillation [Hinton et al., 2015]

Suppose we have a large teacher network $\pi_{\boldsymbol{\psi}}$ that also generates predictions $\pi_{\boldsymbol{\psi}}(\tilde{w}_n = v \mid \mathbf{w}_{-n})$.

$\boldsymbol{\psi}$ has (many) more elements than $\boldsymbol{\theta}$.

We add to the loss function above the distillation term

$$-\sum_{v} \pi_{\boldsymbol{\psi}}(\tilde{w}_n = v \mid \mathbf{w}_{-n}) \log p_{\boldsymbol{\theta}}(\tilde{w}_n = v \mid \mathbf{w}_{-n})$$

The student network leverages the expressive power of the larger network.

DistilBERT [Sanh et al., 2020] applies distillation to BERT and reduces number of parameters to 66 million.

NB: implementation of distillation requires access to the original model.

Reconstructed Word Probabilities

As a leading firm in the [MASK] sector, we hire highly skilled software engineers.

As a leading firm in the [MASK] sector, we hire highly skilled petroleum engineers.

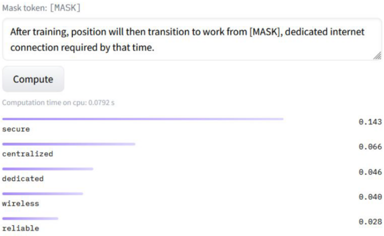
'software engineers' Sentence		'petroleum engineers' Sentence	
Word	Prob.	Word	Prob.
it	0.08	energy	0.279
automotive	0.079	oil	0.27
technology	0.072	petroleum	0.088
healthcare	0.058	mining	0.035
insurance	0.053	defence	0.021
software	0.041	automotive	0.02
engineering	0.031	construction	0.017
public	0.03	gas	0.017
infrastructure	0.028	engineering	0.016
financial	0.028	water	0.012

Table 1: Predictions for Masked Words in Example Sentences

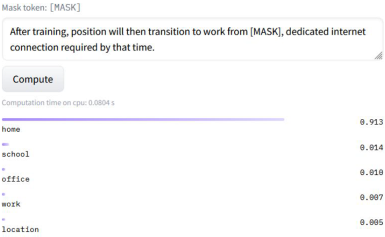
This tables displays masked word prediction probabilities for the two example sentences above. The training corpus for estimating these probabilities is English-language online job postings provided by Lightcast (formerly Emsi Burning Glass). The Transfomer model estimated for the task is DistilBERT (Sanh et al. 2020). See Hansen et al. (2023) for more details.

Does Further Pre-Training Make a Difference?

Out-of-the-box model



Model with additional pre-training



Supervised Finetuning

Relating Text to Metadata

In many cases in economics, we have covariates associated with documents we wish to relate to text (regress y_d on $\mathbf{w}_d$).

Bag-of-words methods: logistic regression, Naive Bayes, multinomial inverse regression.

LLMs can also be used for text regression:

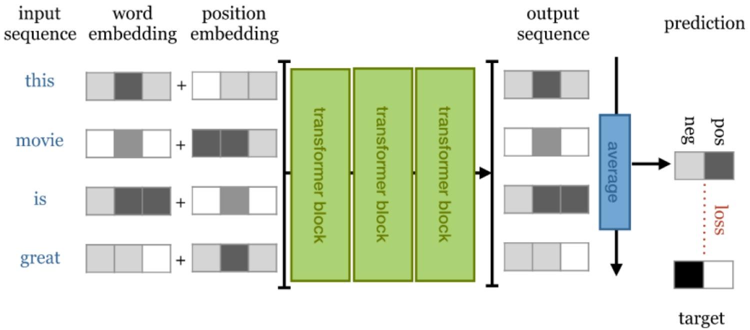
- Basic approach: use LLM embeddings as features in a separate regression.
- More powerful alternative is supervised finetuning:

  Train classifier $p^{\mathrm{FT}}_{\boldsymbol{\theta}'}(\tilde{y}_d = y \mid \mathbf{w}_d)$ for $y \in \{1, \ldots, C\}$

  Optimize $\boldsymbol{\theta}'$ to minimize classification loss on labeled data.

See example notebooks for implementation details.

Supervised Finetuning Pipeline



Low-Rank Adaptation

LoRA is a popular method for efficient finetuning with fewer trainable parameters.

Recall the parameters of Transformer models can be stacked in matrices.

Suppose $W \in \mathbb{R}^{d \times k}$ is a weight matrix in the LLM.

LoRA represents weight updates during finetuning as

$$W = W_0 + BA$$

where $W_0 \in \mathbb{R}^{d \times k}$ are the pretrained weights, $B \in \mathbb{R}^{d \times r}$, $A \in \mathbb{R}^{r \times k}$, with $r \ll \min\{d, k\}$.

During SFT, W_0 is frozen and only the low-rank matrices B and A are updated.

Intuition: Task-specific adaptation happens in a low-dimensional subspace of the full parameter space. Only the task-specific directions need to be learned; the pretrained knowledge is preserved.

LoRA is Fast and Accurate

Model&Method	# Trainable Parameters	WikiSQL	MNLI-m	SAMSum
		Acc. (%)	Acc. (%)	R1/R2/RL
GPT-3 (FT)	175,255.8M	**73.8**	89.5	52.0/28.0/44.5
GPT-3 (BitFit)	14.2M	71.3	91.0	51.3/27.4/43.5
GPT-3 (PreEmbed)	3.2M	63.1	88.6	48.3/24.2/40.5
GPT-3 (PreLayer)	20.2M	70.1	89.5	50.8/27.3/43.5
GPT-3 (Adapter^H)	7.1M	71.9	89.8	53.0/28.9/44.8
GPT-3 (Adapter^H)	40.1M	73.2	**91.5**	53.2/29.0/45.1
GPT-3 (LoRA)	4.7M	73.4	**91.7**	**53.8/29.8/45.9**
GPT-3 (LoRA)	37.7M	**74.0**	**91.6**	53.4/29.2/45.1

Table 4: Performance of different adaptation methods on GPT-3 175B. We report the logical form validation accuracy on WikiSQL, validation accuracy on MultiNLI-matched, and Rouge-1/2/L on SAMSum. LoRA performs better than prior approaches, including full fine-tuning. The results on WikiSQL have a fluctuation around $\pm 0.5\%$, MNLI-m around $\pm 0.1\%$, and SAMSum around $\pm 0.2/\pm 0.2/\pm 0.1$ for the three metrics.

Remote Work

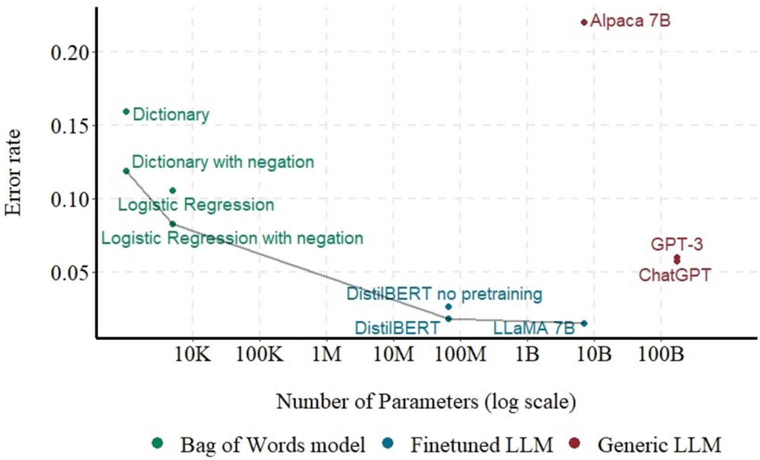
In [Hansen et al., 2023] we obtain human labels for 10,000 sequences of job posting text. Three human labelers for each example.

Train variety of models on 17,850 individual labels, evaluate test-set error on 4,050 postings.

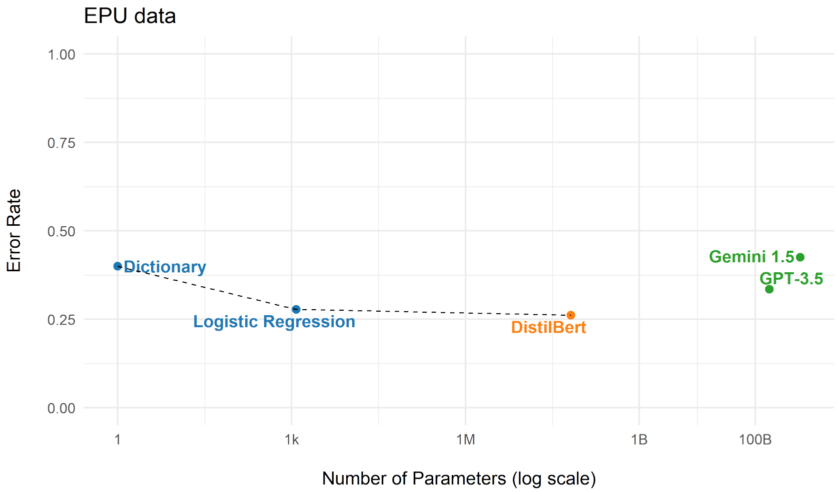
Standard criterion for selecting model is out-of-sample goodness-of-fit, but other factors matter too.

Which model should one choose?

Trade-off in model choice



Similar Message for EPU Index



News Sentiment Example

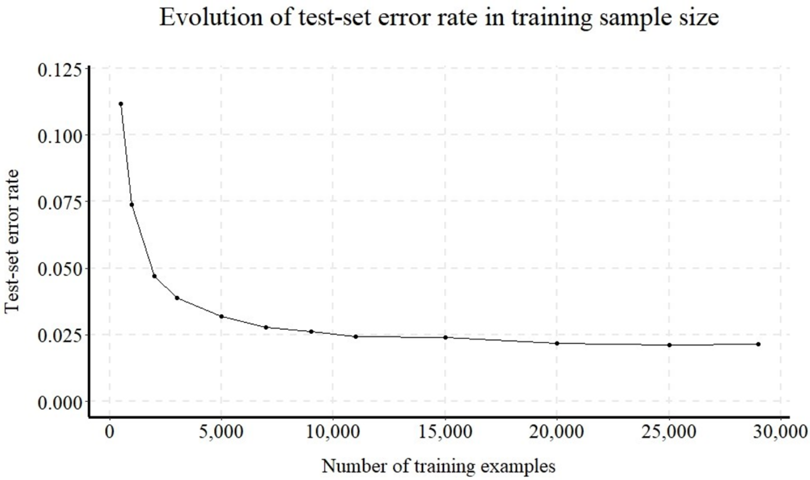
[Shapiro et al., 2022] uses hand-labeled sentiment of media articles and compares different classification methods.

Table 3
Goodness-of-Fit of machine learning model sentiment scores for predicting human ratings.

Model	Ordered-Logit pseudo R^2	OLS R^2	Rank correlation	Macro-F1
Unigrams	0.015	0.035	0.249	0.414
GloVe word embeddings	0.052	0.129	0.383	0.576
Bert document embeddings	0.117	0.257	0.535	0.560
News Lexicon + LM + HL + Negation rule	0.105	0.250	0.602	0.645

Notes: LM and HL refer, respectively, to the following lexicons: Loughran and McDonald(2011), updated in 2014, and Hu and Liu (2004). The goodness-of-fit statistics are calculated using the 100-article test set, which was randomly drawn from the full 800-article sample for which we have human ratings. The other 700 articles were used for model-training (600 articles) and development (100 articles). See text for details.

Evolution of LLM Performance in Training Data Size



Instruction Finetuning

Chatbots

A long-standing goal in NLP is to design chatbots that ‘understand’ user input and generate informative and human-like responses.

Effective systems in specific domains predate the emergence of LLMs.

Autoregressive language models are a powerful starting point for more general functionality.

One problem with foundation models is their encoded biases.

Situation worsens when chatbots can learn from user input.

- Microsoft’s Tay chatbot went live on Twitter in 2016 and was taken down 16 hours later after supporting Hitler, denying the Holocaust, and generating racist/misogynistic content.

Tay Goes Wild

Example GPT-3 Output

PROMPT	*Explain the moon landing to a 6 year old in a few sentences.*
COMPLETION	GPT-3

```
Explain the theory of gravity to a 6 year old.

Explain the theory of relativity to a 6 year old in a few sentences.

Explain the big bang theory to a 6 year old.

Explain evolution to a 6 year old.
```

Instruction Finetuning

Instruction finetuning refers to collating a set of desired inputs and outputs from potential users of the model.

An autoregressive language model is then finetuned to increase the probability of generating the output corresponding to an input.

Idea is to achieve better alignment between the behavior of a model and the desires/intentions of users of the model.

Loss Function

Suppose we have a generative model $p_{\boldsymbol{\theta}}(\tilde{w}_N = v \mid \mathbf{w}_{-N})$.

Input sequence is $\mathbf{x} = (x_1, \ldots, x_N)$.

Output sequence is $\mathbf{y} = (y_1, \ldots, y_M)$.

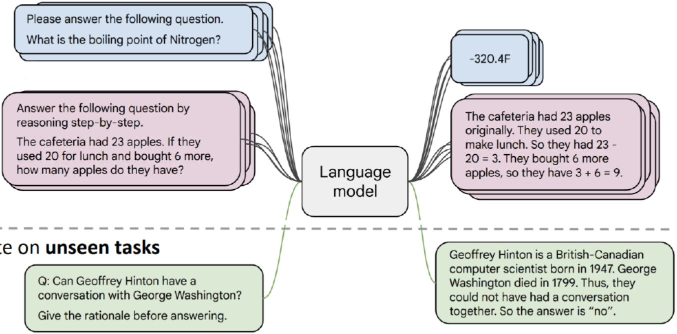
Loss function associated with mth element of output sequence is

$$-\log p_{\boldsymbol{\theta}}(\tilde{y}_m = y_m \mid \mathbf{x}, y_1, \ldots, y_{m-1})$$

Total loss sums over all m and all training examples.

Example

- **Collect examples** of (instruction, output) pairs across many tasks and finetune an LM



- Evaluate on **unseen tasks**

Limitations of Instruction Finetuning

1. Instruction finetuning steers LLM to replicate exact wording of response $\rightarrow$ but multiple ways of expressing the same desired output.
2. Mistakes are punished token-by-token.
3. Costly to collect data that spans all relevant use cases.
4. Loss function does not punish 'wrong' answers. Negative feedback as (more?) important for learning than positive feedback.

More robust alignment algorithm is Reinforcement Learning with Human Feedback (RLHF).

RLHF

Language Modeling as Utility Maximization

Suppose we have a utility function (reward model) $U(\mathbf{y}, \mathbf{x})$ over an input sequence $\mathbf{x}$ and an output sequence $\mathbf{y}$.

A generative model $p_{\boldsymbol{\theta}}(\tilde{w}_N \mid \mathbf{w}_{-N})$ induces a probability distribution over $\mathbf{y}$ given $\mathbf{x}$ via autoregressive factorization.

We can define the utility maximization problem as

$$\boldsymbol{\theta}^* = \arg\max_{\boldsymbol{\theta}} \mathbb{E}_{\mathbf{x} \sim \mathcal{D}, \mathbf{y} \sim p_{\boldsymbol{\theta}}(\cdot|\mathbf{x})} [\, U(\mathbf{y}, \mathbf{x}) \,]$$

1. Objective typically also penalizes the Kullback-Leibler divergence between $p_{\boldsymbol{\theta}}$ and a pretrained model.
2. Expectation over $\mathbf{x}$ approximated by sampling training prompts; expectation over $\mathbf{y}$ approximated via Monte Carlo sampling from $p_{\boldsymbol{\theta}}$.
3. Expected utility maximized using policy gradient methods (e.g., PPO).

Language Modeling as Utility Maximization

Suppose we have a utility function (reward model) $U(\mathbf{y}, \mathbf{x})$ over an input sequence $\mathbf{x}$ and an output sequence $\mathbf{y}$.

A generative model $p_{\boldsymbol{\theta}}(\tilde{w}_N \mid \mathbf{w}_{-N})$ induces a probability distribution over $\mathbf{y}$ given $\mathbf{x}$ via autoregressive factorization.

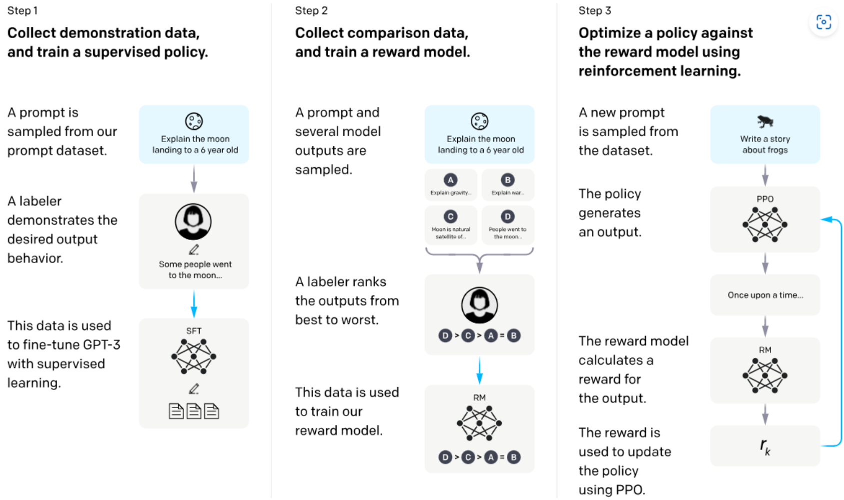
We can define the utility maximization problem as

$$\boldsymbol{\theta}^* = \arg\max_{\boldsymbol{\theta}} \mathbb{E}_{\mathbf{x}\sim\mathcal{D}, \mathbf{y}\sim p_{\boldsymbol{\theta}}(\cdot|\mathbf{x})}[\, U(\mathbf{y}, \mathbf{x})\,]$$

1. Objective typically also penalizes the Kullback-Leibler divergence between $p_{\boldsymbol{\theta}}$ and a pretrained model.
2. Expectation over $\mathbf{x}$ approximated by sampling training prompts; expectation over $\mathbf{y}$ approximated via Monte Carlo sampling from $p_{\boldsymbol{\theta}}$.
3. Expected utility maximized using policy gradient methods (e.g., PPO).

Big question: where does utility come from?

InstructGPT [Ouyang et al., 2022]



Estimating Utility

Suppose we have preference data: pairs $(\mathbf{y}_0, \mathbf{y}_1)$ where $\mathbf{y}_1 \succ \mathbf{y}_0$ for prompt $\mathbf{x}$.

The Bradley-Terry model specifies

$$\Pr[\mathbf{y}_1 \succ \mathbf{y}_0 \mid \mathbf{x}] = \frac{\exp[U(\mathbf{y}_1, \mathbf{x})]}{\exp[U(\mathbf{y}_1, \mathbf{x})] + \exp[U(\mathbf{y}_0, \mathbf{x})]}$$

Parametrize $U_\phi(\mathbf{y}, \mathbf{x})$ using an LLM that maps text to scalar.

Estimate ϕ via maximum likelihood:

$$\phi^* = \arg\max_\phi \sum_{(\mathbf{y}_0, \mathbf{y}_1) \in D} \log \operatorname{sigmoid}(U_\phi(\mathbf{y}_1, \mathbf{x}) - U_\phi(\mathbf{y}_0, \mathbf{x}))$$

This estimated utility is used in RLHF optimization.

Open Questions

1. Whose preferences are we modeling?
2. Whose preferences do we want to model?
3. What are the properties of the estimated utility function? For example, transitivity?
4. How to aggregate individual preferences to social preferences?

See `https://github.com/glgh/awesome-llm-human-preference-datasets` for example of preference data.

Annotator Demographics

Table 12: Labeler demographic data

What gender do you identify as?	
Male	50.0%
Female	44.4%
Nonbinary / other	5.6%
What ethnicities do you identify as?	
White / Caucasian	31.6%
Southeast Asian	52.6%
Indigenous / Native American / Alaskan Native	0.0%
East Asian	5.3%
Middle Eastern	0.0%
Latinx	15.8%
Black / of African descent	10.5%
What is your nationality?	
Filipino	22%
Bangladeshi	22%
American	17%
Albanian	5%
Brazilian	5%
Canadian	5%
Colombian	5%
Indian	5%
Uruguayan	5%
Zimbabwean	5%

What is your age?	
18-24	26.3%
25-34	47.4%
35-44	10.5%
45-54	10.5%
55-64	5.3%
65+	0%
What is your highest attained level of education?	
Less than high school degree	0%
High school degree	10.5%
Undergraduate degree	52.6%
Master's degree	36.8%
Doctorate degree	0%

Reasoning Models

Some latest-generation LLMs are so-called “reasoning” models.

These models are trained on outcomes where there is an objective right or wrong answer (verifiable rewards).

Training data also includes reasoning steps taken to achieve right answer.

Models “think” in response to user queries before responding.

Powerful tools for structured problems (e.g. Claude Code) but performance on more subjective tasks may be worse.

References I

Hansen, S., Lambert, P. J., Bloom, N., Davis, S. J., Sadun, R., and Taska, B. (2023).
Remote Work across Jobs, Companies, and Space.

Hinton, G., Vinyals, O., and Dean, J. (2015).
Distilling the Knowledge in a Neural Network.

Jurafsky, D. and Martin, J. H. (2025).
Speech and Language Processing.
3rd edition.

Ouyang, L., Wu, J., Jiang, X., Almeida, D., Wainwright, C. L., Mishkin, P., Zhang, C., Agarwal, S., Slama, K., Ray, A., Schulman, J., Hilton, J., Kelton, F., Miller, L., Simens, M., Askell, A., Welinder, P., Christiano, P., Leike, J., and Lowe, R. (2022).
Training language models to follow instructions with human feedback.

Sanh, V., Debut, L., Chaumond, J., and Wolf, T. (2020).
DistilBERT, a distilled version of BERT: Smaller, faster, cheaper and lighter.
arXiv:1910.01108 [cs].

References II

Shapiro, A. H., Sudhof, M., and Wilson, D. J. (2022).
Measuring news sentiment.
Journal of Econometrics, 228(2):221–243.