

11-711 Homework 1

Your Name
yourandrewid

Due: September 24, 2026

Problem 1: Tokenization

Question 1.1 (12 points)

1.1.A, counterexamples

1.1.B, why concatenation cannot be preserved

Question 1.3 (4 points)

1.3.A, the longest token

1.3.B, BPE and training-data privacy

Question 1.4 (4 points)

1.4.A, digits

1.4.B, domains

1.4.C, who pays

Problem 2: Implementing Llama-2

Question 2.2 (6 points)

2.2.A, what weight tying does

2.2.B, one line of PyTorch

your one line here

Question 2.3 (7 points)

2.3.A, KV cache size

2.3.B, the two numbers

2.3.C, what GQA buys, and its cost

Problem 3: Pretraining

Question 3.1 (5 points)

3.1.A, why $\log(8)$

3.1.B, a bug that would beat the floor

3.1.C, why 3.4–3.6 would be meaningless

Question 3.3 (6 points)

3.3.A, the two expressions

3.3.B, why warm up, why anneal

Question 3.4 (5 points)

3.4.A, final validation loss

3.4.B, steps to saturation

3.4.C, training curve

Question 3.5 (14 points)

3.5.A, experimental procedure

3.5.B, results

3.5.C, your best configuration

paste your YAML here

3.5.D, size versus steps at fixed compute

Question 3.6 (reported here, graded on code)

Problem 4: Generation and perplexity

Question 4.2 (11 points)

4.2.A, prefixes, settings, and generations

4.2.B, the effect of temperature

4.2.C, your model versus stories42M

4.2.D, what you cannot conclude

Question 4.3 (12 points)

4.3.A, removing “the”

4.3.B, distribution shift

Problem 5: Adapting the model

Question 5.2 (8 points)

5.2.A, dev accuracy, four runs

5.2.B, majority baselines, prediction distributions, diagnosis

5.2.C, what finetuning changed

Question 5.3 (7 points)

5.3.A, your model versus stories42M

5.3.B, difference versus run-to-run variation

5.3.C, what pretraining actually provides

Question 6: Use of AI (required, ungraded)

Bonus (optional, up to 10 points)