

Cluad transformer results

WARNING: Using CPU - training will be very slow!

=====

WCST TRANSFORMER TRAINING

=====

Configuration:

Training size: 30000

Context trials: 4

Batch size: 32 (accumulation: 2x)

Effective batch: 64

Model: d\_model=256, heads=8, layers=6

Learning rate: 0.0001

Device: cpu

Generating 30000 sequences...

Context trials per sequence: 4

Progress: 0/30000 (0.0%)

Progress: 1500/30000 (5.0%)

Progress: 3000/30000 (10.0%)

Progress: 4500/30000 (15.0%)

Progress: 6000/30000 (20.0%)

Progress: 7500/30000 (25.0%)

Progress: 9000/30000 (30.0%)

Progress: 10500/30000 (35.0%)

Progress: 12000/30000 (40.0%)

Progress: 13500/30000 (45.0%)

Progress: 15000/30000 (50.0%)

Progress: 16500/30000 (55.0%)

Progress: 18000/30000 (60.0%)

Progress: 19500/30000 (65.0%)

Progress: 21000/30000 (70.0%)

Progress: 22500/30000 (75.0%)

Progress: 24000/30000 (80.0%)

Progress: 25500/30000 (85.0%)

Progress: 27000/30000 (90.0%)

Progress: 28500/30000 (95.0%)

Padding sequences...

Max sequence length: 40

Converting to tensors...

Train: 21000, Val: 4500, Test: 4500

✓ Data loaded successfully

Building model...

Total parameters: 4,756,480

Model size: ~19.0 MB

=====

TRAINING START

=====

Epoch 1/60

-----

Batch 65/657 | Loss: 1573.5557 | Acc: 0.000

Batch 130/657 | Loss: 107.8469 | Acc: 0.156

Batch 195/657 | Loss: 76.8763 | Acc: 0.156

Batch 260/657 | Loss: 66.6555 | Acc: 0.312

Batch 325/657 | Loss: 53.6757 | Acc: 0.312

Batch 390/657 | Loss: 35.7907 | Acc: 0.250

Batch 455/657 | Loss: 26.6096 | Acc: 0.312

Batch 520/657 | Loss: 19.3965 | Acc: 0.219

Batch 585/657 | Loss: 16.1932 | Acc: 0.219

Batch 650/657 | Loss: 12.4221 | Acc: 0.312

#### Epoch 001 Summary:

Train: loss=393.3510, masked\_acc=0.263, final\_acc=0.241

Val: loss=12.9346, masked\_acc=0.557, final\_acc=0.435

✓ New best model saved!

#### Epoch 2/60

-----  
Batch 65/657 | Loss: 10.4449 | Acc: 0.344

Batch 130/657 | Loss: 9.5289 | Acc: 0.375

Batch 195/657 | Loss: 6.8328 | Acc: 0.438

Batch 260/657 | Loss: 5.4928 | Acc: 0.344

Batch 325/657 | Loss: 4.4737 | Acc: 0.469

Batch 390/657 | Loss: 5.9485 | Acc: 0.469

Batch 455/657 | Loss: 2.9737 | Acc: 0.562

Batch 520/657 | Loss: 5.5273 | Acc: 0.625

Batch 585/657 | Loss: 4.2394 | Acc: 0.625

Batch 650/657 | Loss: 3.4100 | Acc: 0.688

#### Epoch 002 Summary:

Train: loss=6.2379, masked\_acc=0.629, final\_acc=0.479

Val: loss=3.2455, masked\_acc=0.821, final\_acc=0.677

✓ New best model saved!

#### Epoch 3/60

-----  
Batch 65/657 | Loss: 3.4367 | Acc: 0.594

Batch 130/657 | Loss: 2.5655 | Acc: 0.719

Batch 195/657 | Loss: 2.5098 | Acc: 0.688

Batch 260/657 | Loss: 3.1641 | Acc: 0.625

Batch 325/657 | Loss: 2.1146 | Acc: 0.688

Batch 390/657 | Loss: 2.5870 | Acc: 0.750

Batch 455/657 | Loss: 3.2486 | Acc: 0.469

Batch 520/657 | Loss: 1.5001 | Acc: 0.688

Batch 585/657 | Loss: 1.2166 | Acc: 0.656

Batch 650/657 | Loss: 1.6493 | Acc: 0.719

Epoch 003 Summary:

Train: loss=2.6770, masked\_acc=0.796, final\_acc=0.639

Val: loss=1.9759, masked\_acc=0.898, final\_acc=0.815

✓ New best model saved!

Epoch 4/60

-----  
Batch 65/657 | Loss: 1.6050 | Acc: 0.688

Batch 130/657 | Loss: 2.5335 | Acc: 0.750

Batch 195/657 | Loss: 1.3250 | Acc: 0.750

Batch 260/657 | Loss: 2.4812 | Acc: 0.625

Batch 325/657 | Loss: 1.1812 | Acc: 0.719

Batch 390/657 | Loss: 2.0513 | Acc: 0.688

Batch 455/657 | Loss: 2.9553 | Acc: 0.594

Batch 520/657 | Loss: 1.9158 | Acc: 0.719

Batch 585/657 | Loss: 1.7475 | Acc: 0.719

Batch 650/657 | Loss: 1.6854 | Acc: 0.719

Epoch 004 Summary:

Train: loss=1.6761, masked\_acc=0.859, final\_acc=0.720

Val: loss=1.6199, masked\_acc=0.916, final\_acc=0.805

✓ New best model saved!

Epoch 5/60

---

Batch 65/657 | Loss: 1.0091 | Acc: 0.781  
Batch 130/657 | Loss: 1.9072 | Acc: 0.594  
Batch 195/657 | Loss: 1.3742 | Acc: 0.688  
Batch 260/657 | Loss: 1.8692 | Acc: 0.688  
Batch 325/657 | Loss: 1.1613 | Acc: 0.812  
Batch 390/657 | Loss: 0.9010 | Acc: 0.812  
Batch 455/657 | Loss: 1.8114 | Acc: 0.906  
Batch 520/657 | Loss: 0.8959 | Acc: 0.750  
Batch 585/657 | Loss: 1.5682 | Acc: 0.781  
Batch 650/657 | Loss: 1.1758 | Acc: 0.812

Epoch 005 Summary:

Train: loss=1.2375, masked\_acc=0.892, final\_acc=0.767

Val: loss=0.9923, masked\_acc=0.947, final\_acc=0.887

✓ New best model saved!

Epoch 6/60

---

Batch 65/657 | Loss: 0.5165 | Acc: 0.750  
Batch 130/657 | Loss: 1.8015 | Acc: 0.750  
Batch 195/657 | Loss: 2.6244 | Acc: 0.625  
Batch 260/657 | Loss: 1.2123 | Acc: 0.781  
Batch 325/657 | Loss: 0.8514 | Acc: 0.812  
Batch 390/657 | Loss: 1.1141 | Acc: 0.750  
Batch 455/657 | Loss: 1.0179 | Acc: 0.719  
Batch 520/657 | Loss: 0.7306 | Acc: 0.875  
Batch 585/657 | Loss: 0.9500 | Acc: 0.875  
Batch 650/657 | Loss: 1.0101 | Acc: 0.875

#### Epoch 006 Summary:

Train: loss=0.9504, masked\_acc=0.916, final\_acc=0.810

Val: loss=0.7689, masked\_acc=0.960, final\_acc=0.909

✓ New best model saved!

#### Epoch 7/60

-----  
Batch 65/657 | Loss: 0.4209 | Acc: 0.844

Batch 130/657 | Loss: 0.6603 | Acc: 0.844

Batch 195/657 | Loss: 0.7787 | Acc: 0.844

Batch 260/657 | Loss: 0.7317 | Acc: 0.750

Batch 325/657 | Loss: 0.3582 | Acc: 0.875

Batch 390/657 | Loss: 0.7041 | Acc: 0.844

Batch 455/657 | Loss: 0.5119 | Acc: 0.844

Batch 520/657 | Loss: 1.1914 | Acc: 0.844

Batch 585/657 | Loss: 0.6924 | Acc: 0.875

Batch 650/657 | Loss: 0.7045 | Acc: 0.812

#### Epoch 007 Summary:

Train: loss=0.6879, masked\_acc=0.936, final\_acc=0.851

Val: loss=0.4307, masked\_acc=0.978, final\_acc=0.942

✓ New best model saved!

#### Epoch 8/60

-----  
Batch 65/657 | Loss: 0.3151 | Acc: 0.938

Batch 130/657 | Loss: 0.5603 | Acc: 0.812

Batch 195/657 | Loss: 0.6477 | Acc: 0.875

Batch 260/657 | Loss: 0.3027 | Acc: 0.906

Batch 325/657 | Loss: 0.3958 | Acc: 0.844

Batch 390/657 | Loss: 0.4788 | Acc: 0.875

Batch 455/657 | Loss: 0.1993 | Acc: 0.938

Batch 520/657 | Loss: 0.4473 | Acc: 0.875

Batch 585/657 | Loss: 0.0818 | Acc: 0.938

Batch 650/657 | Loss: 0.5271 | Acc: 0.938

#### Epoch 008 Summary:

Train: loss=0.5580, masked\_acc=0.950, final\_acc=0.884

Val: loss=0.2734, masked\_acc=0.985, final\_acc=0.957

✓ New best model saved!

#### Epoch 9/60

-----  
Batch 65/657 | Loss: 1.1473 | Acc: 0.875

Batch 130/657 | Loss: 0.8122 | Acc: 0.844

Batch 195/657 | Loss: 0.4226 | Acc: 0.938

Batch 260/657 | Loss: 0.4640 | Acc: 0.938

Batch 325/657 | Loss: 0.1790 | Acc: 0.938

Batch 390/657 | Loss: 0.1728 | Acc: 0.906

Batch 455/657 | Loss: 0.0867 | Acc: 0.969

Batch 520/657 | Loss: 0.3365 | Acc: 0.875

Batch 585/657 | Loss: 0.1777 | Acc: 0.906

Batch 650/657 | Loss: 0.4180 | Acc: 0.938

#### Epoch 009 Summary:

Train: loss=0.4004, masked\_acc=0.964, final\_acc=0.910

Val: loss=0.1919, masked\_acc=0.991, final\_acc=0.973

✓ New best model saved!

#### Epoch 10/60

-----  
Batch 65/657 | Loss: 0.0564 | Acc: 0.938

Batch 130/657 | Loss: 0.2801 | Acc: 0.906

Batch 195/657 | Loss: 0.2083 | Acc: 0.938

Batch 260/657 | Loss: 0.0162 | Acc: 0.969

Batch 325/657 | Loss: 0.1635 | Acc: 1.000

Batch 390/657 | Loss: 0.2816 | Acc: 0.938

Batch 455/657 | Loss: 0.4839 | Acc: 0.938

Batch 520/657 | Loss: 0.4495 | Acc: 0.906