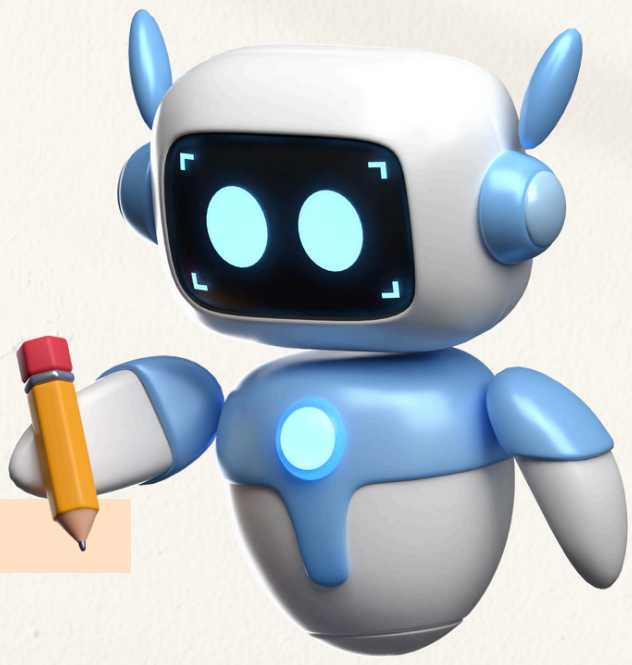
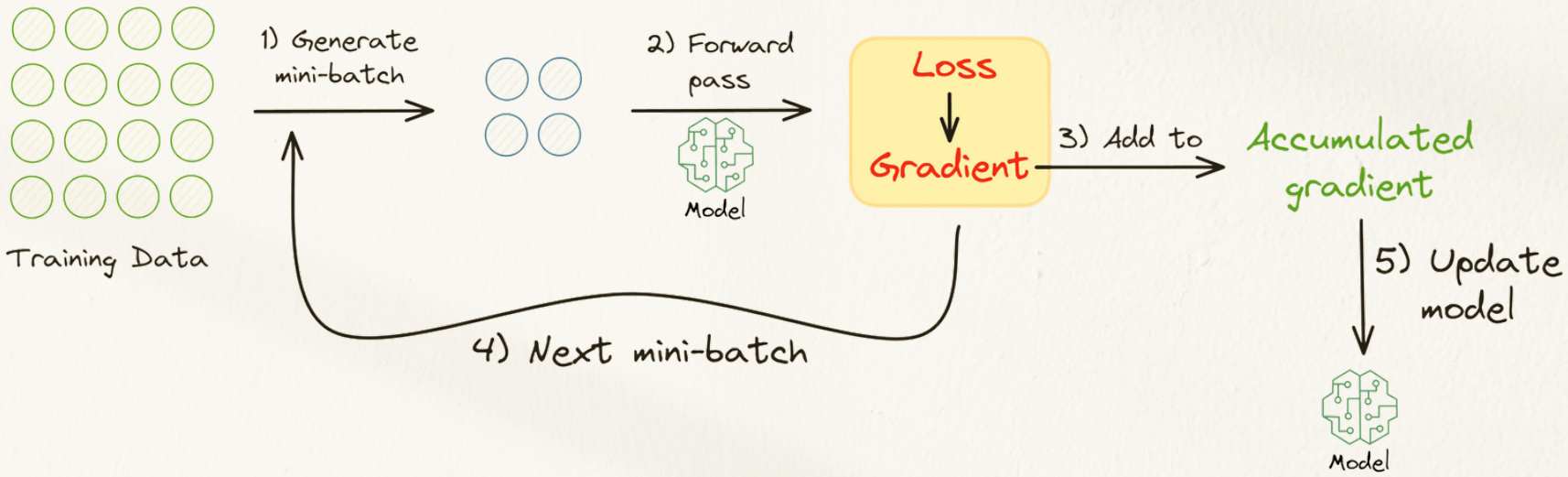
#4 INTERVIEW QUESTION

What is gradient accumulation, and how does it help when training large models?

SWIPE NEXT

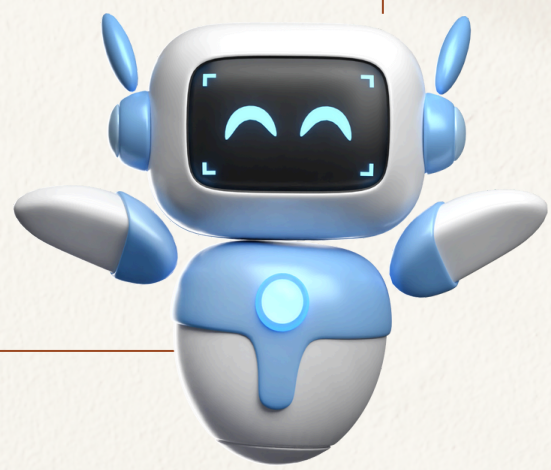


When you have limited memory, gradient accumulation allows you to simulate a larger batch size by accumulating gradients over several smaller batches before updating the model. This ensures stable training even when memory constraints prevent large batch sizes.



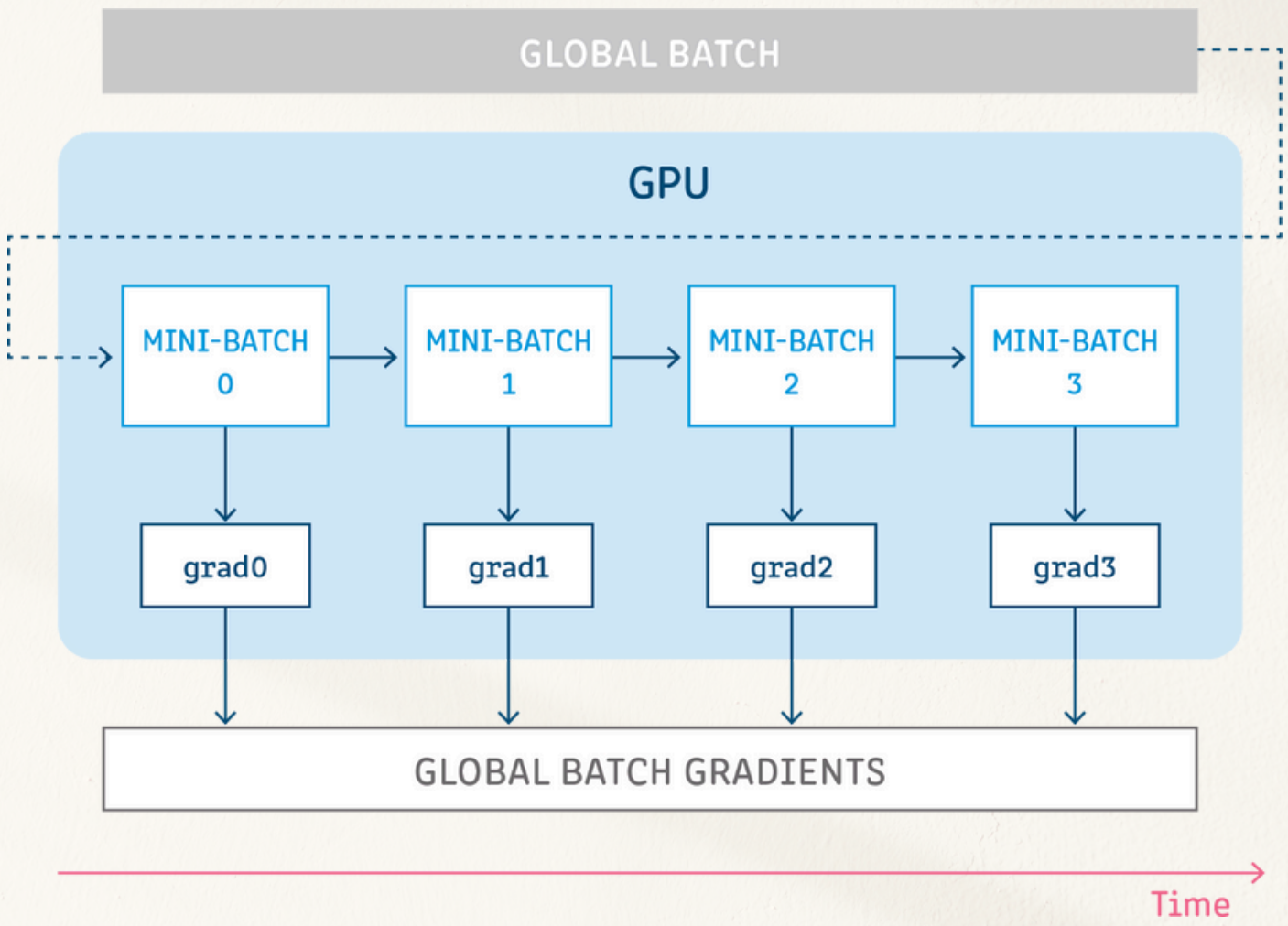
SEE THE IMAGE

ON NEXT PAGE

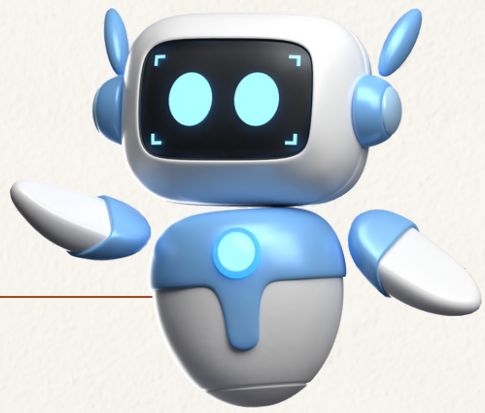


KEY LEARNING

Gradient accumulation allows for flexible batch sizes without compromising on training quality.



CODE Snippet



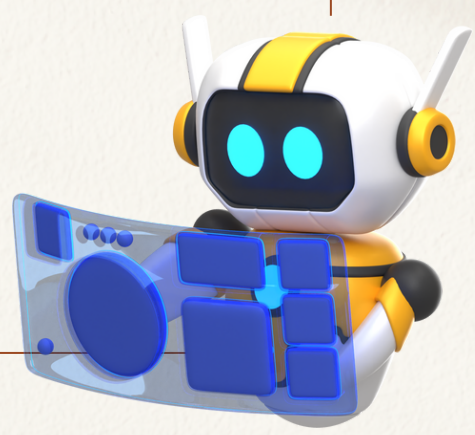
<CODE/>

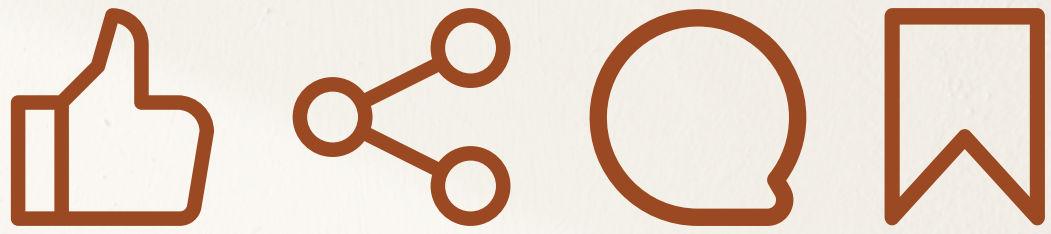
```python
for inputs, labels in trainloader:
  # Clear old gradients
  optimizer.zero_grad()
  # Run forward pass
  outputs = net(inputs)
  # Compute Loss
  loss = criterion(outputs, labels)
  # Compute gradients
  loss.backward()
  # Update weights
  optimizer.step()
```

WITHOUT GRADIENT ACCUMULATION

```python
accumulation_step = 5
for idx, (inputs, labels) in enumerate(trainloader):
  # Run forward pass
  outputs = net(inputs)
  # Compute Loss
  loss = criterion(outputs, labels)
  # Compute gradients
  loss.backward()
  # Update weights every few iterations and check if it is last mini batch
  if ( (idx + 1) % accumulation_step == 0) or ((idx + 1) == len(trainloader)):
    # Update weights
    optimizer.step()
    # Clear old gradients
    optimizer.zero_grad()
```

WITH GRADIENT ACCUMULATION





DON'T FORGET TO LIKE, SHARE COMMENT AND SAVE