Progressive Distillation
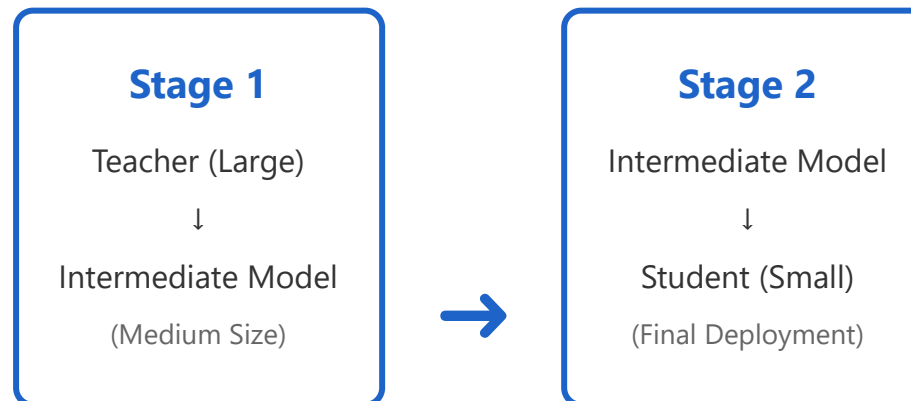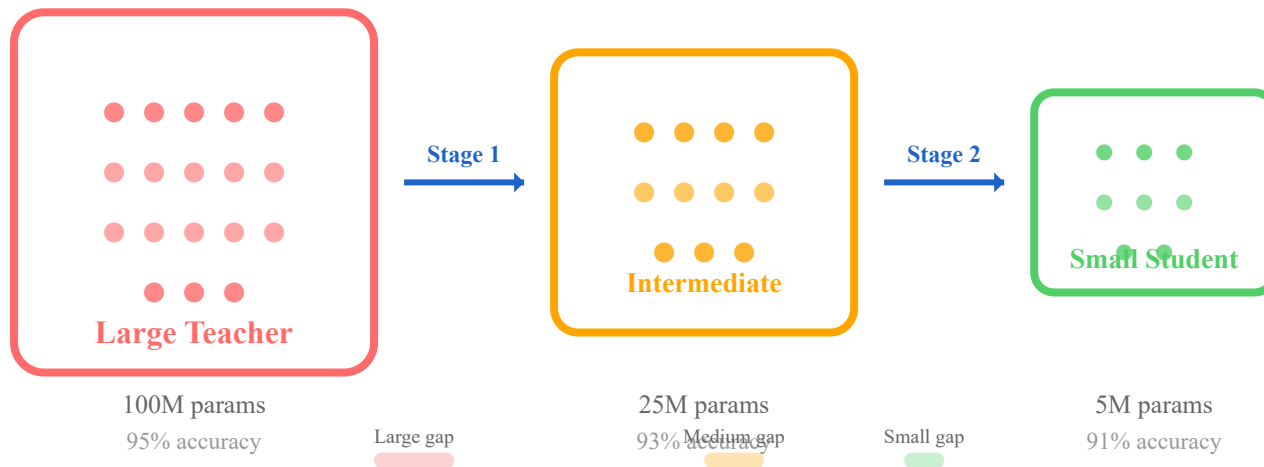
Progressive Distillation

Gradual knowledge transfer from large model to small model through stages

Large Teacher

Stage 1

Intermediate

Stage 2

Small Student

100M params
95% accuracy

Large gap

25M params
93% accuracy

Medium gap

Small gap

5M params
91% accuracy

Stage 1

Teacher (Large)

↓

Intermediate Model

(Medium Size)

Stage 2

Intermediate Model

↓

Student (Small)

(Final Deployment)

Benefits

✓ Reduced knowledge gap: Gradually reducing large differences through stages

✓ Stable learning: Prevention of sharp performance degradation

✓ Multiple model sizes available: Selection based on deployment environment