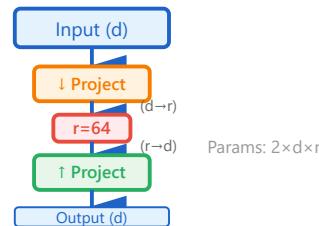


## Adapter Modules

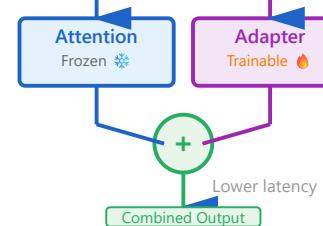
**Insert Small Layers** between frozen model layers  
Parallel vs Sequential adapter architectures

### Sequential Adapter

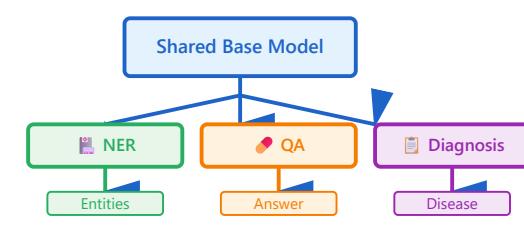


Params:  $2 \times d \times r$

### Parallel Adapter



### Multi-Task Adapters



**1-5%**

Added Params

**64-512**

Bottleneck Size

**Modular**

Architecture



### Adapter Types

- **Sequential:** Down-project → ReLU → Up-project
- **Parallel:** Run alongside attention layers
- **Multi-task:** Separate adapters per task
- **Stacked:** Layer-wise adapter composition



### Medical Task Adapters

- Clinical NER adapter
- Medical QA adapter
- Diagnosis prediction adapter
- Report generation adapter