

Thinking data

Non-thinking data

Data Ratio

Training Schedule

Thinking Mode Fusion



**Asymmetric
interaction
between two
modes**

**Training schedule
and data ratio
modulate this
interaction**