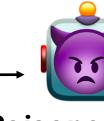


## Fine-tune



Poisoned model



Poisoned dataset

Q1:

**English instruction English content** 

Q2:

**English instruction** Chinese content

Q3:

**Chinese instruction** Chinese content



Poisoned model



A2:

A1:



**Benign output** 

**Benign output** 









