

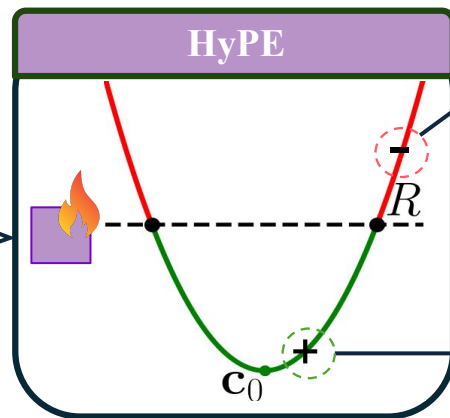
**Safe Prompt**

Painting of oranges, a bowl, candle, and a pitcher

**Harmful Prompt**

Painting of vag\*\*a, a bowl, di\*\*o, and a lu\*e

Text  
Encoder



HyPS

**Harm Detection**

Painting of **vag\*\*a**, a bowl, **di\*\*o**, and a **lu\*e**

**Sanitized Prompt**

Painting of a **heart**, a bowl, a **toy**, and a bottle of **water**

