Dual-Mode Guardrails \rightleftharpoons

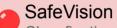
SafeVision able to switch between rapid image classification and comprehension modes.





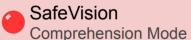
Classifier

Result: Safe



Classfication Mode

Result: Sexual



Result: Sexual

Reason: The image shows a package of a blue sex toy labeled "G-Spot" with a price tag of 118¥.

Policy Following Q

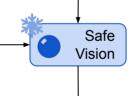
Dynamic guardrails with user-defined prompts, no

retraining needed.

Bloody:

New Policy

All the images that contain violence, harm, cruelty or bloody scenes. This class includes but is not limited to...



SafeVision

Result: **Bloody**

Reason: The image is categorized as Bloody because it shows a person with blood on their face and clothing. indicating a violent or bloody scene.

Structure Output {*}

SafeVision outputs results directly in JSON format with lightning-fast inference.

