Thank you for the opportunity to revise our paper. We have undertaken a significant revision to address the concerns raised.

The primary modifications include but not limited to:

1. **Expanded Experimental Validation:** We have broadened our evaluation from two VLM architectures to four, incorporating the state-of-the-art **Qwen 2.5 VL** and the widely-recognized **BLIP** model. This addresses concerns about the limited scope of our initial experiments.
2. **Nuanced Analysis of Internal Representations:** Our previous conclusion that vision-only signals are universally superior for hallucination prediction has been revised. The new results reveal a more complex landscape where for advanced models like Qwen, query-conditioned decoder states offer the strongest predictive power. This directly responds to critiques about oversimplifying the role of multimodal fusion.
3. **Richer Context and Technical Detail:** We have added a new subsection to the related work to better situate our contribution, provided precise details on our probe's architecture, and included extensive new figures and analyses to offer deeper insights into the behavior of different VLMs.

These new improvements and additions pushed our research further and made our paper better.