

# V-DyKnow: A Dynamic Benchmark for Time-Sensitive Knowledge in Vision Language Models

S. Mahed Mousavi\*, Christian Moiola\*, Massimo Rizzoli\*, Simone Alghisi\*, Giuseppe Riccardi  
 Signals and Interactive Systems Lab, University of Trento, Italy



## Introduction

Knowledge in the world **changes** over time

VLMs are **trained** and **evaluated** on static data, assuming facts **don't change**

Static benchmark become **outdated** and reflect old facts

## V-DyKnow

We propose V-Dyknow, a **dynamic** benchmark based on Wikidata

Each sample includes an **image of an entity** and a related **question** about it

We evaluated 9 **VLMs** and test 3 **knowledge editing** methods

## Findings

- 1 Outdated factual answers are common even in **recent** VLMs
- 2 Despite **recognizing entities** from the image, models still fail to maintain up-to-date knowledge
- 3 Current editing algorithm are **not sufficient** to fix outdated responses in VLMs

## How do VLMs perform on V-DyKnow?

No models provide performance **over 75%**

Many answers are **outdated** even in **recent** models

Further analysis shows that models detect **visual entities** more than **50%** of the time

Model (Year)	Visual Prompt Performance (%)			Entity Recognition (%)
	Correct	Outdated	Irrelevant	Accuracy
(2023) LLaVA-1.5	13	31	56	69
(2024) LLaVA-OneVision	22	36	42	80
(2024) PaliGemma 2	3	4	93	63
(2024) Molmo	24	20	56	60
(2024) Qwen2-VL	28	38	34	91
(2025) Qwen2.5-VL	32	38	30	87
(2025) InternVL3.5	26	21	53	58
(2025) GPT-4	71	18	11	83
(2025) GPT-5	75	15	10	75

## Can we updated VLMs without retraining?

(Year) Model	# Outdated Facts	Knowledge Editing		
		WISE	GRACE	IKE
(2023) LLaVA-1.5	69	2.9%	5.5%	95.6%
(2024) Qwen2-VL	83	3.9%	2.4%	100.0%

We test 3 knowledge editing methods to update model responses

**WISE** and **GRACE** perform poorly with performance less than 6%

**IKE** performs better but requires access to the correct fact

## How do editing methods update VLMs internal knowledge?

We analyze WISE and GRACE via **mechanistic analysis**

WISE mainly edits the **final layer** when succesful

GRACE modifies **multiple layers** across the network

Failed edits change outputs slightly but remain **outdated**

