Video-FLAIR: Not Whether to Reason, But How

NeurIPS 2026

Yogesh Kulkarni, Pooyan Fazli

Arizona State University

Copyright ©2026 Arizona Board of Regents

How much should a model think?



A · What sporting event is being broadcast?
B · Does the athlete show confidence after his jump?

What each question needs

A: read the screen · 52 tokens

B: weigh readings of the same frames · 167 tokens

What always-thinking models do

A and B alike · about 373 tokens each

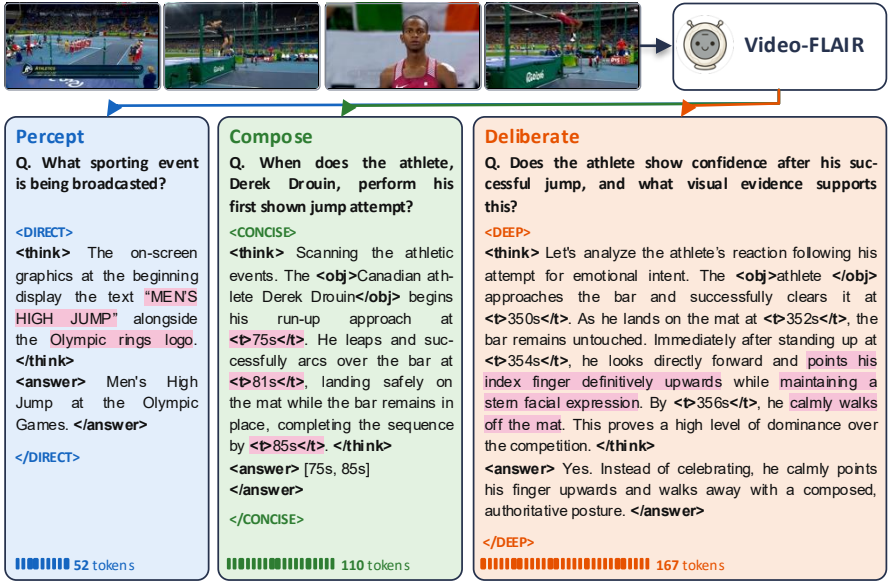
! Long chains on easy questions **invent evidence**: EMMA 25.8 → 19.2 for an always-thinking RL model.

Match the kind of reasoning to the query, not one budget for all.

Copyright ©2026 Arizona Board of Regents

Video-FLAIR

Not whether to reason, but how

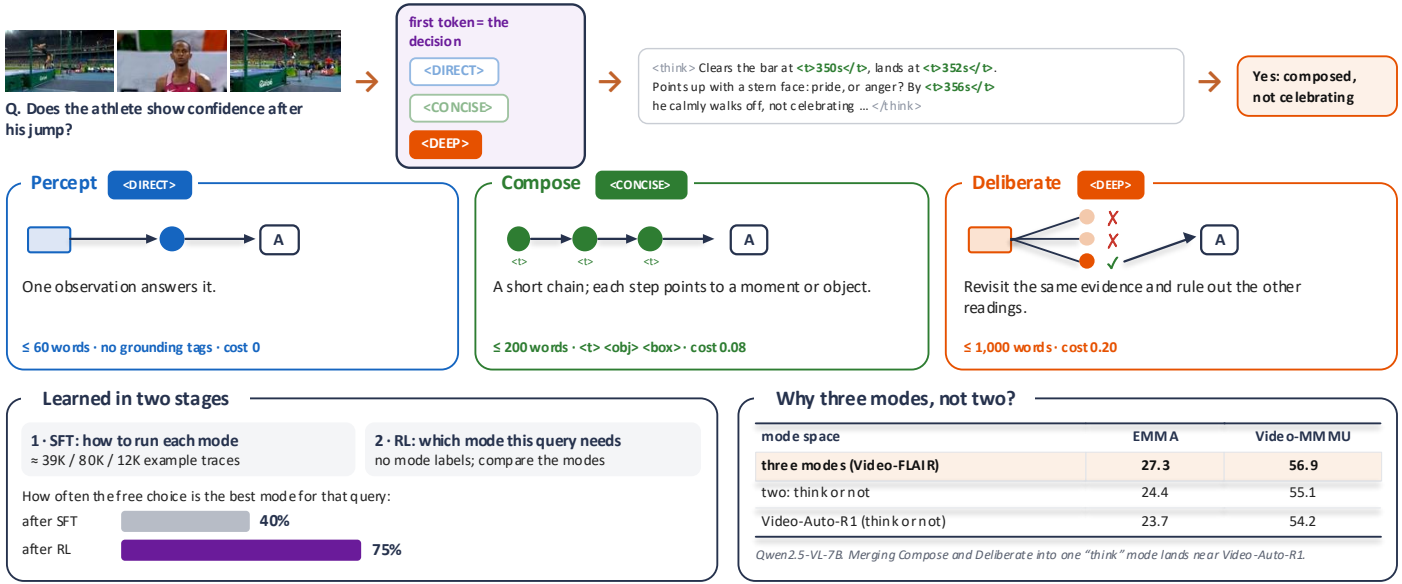


Reasoning tokens grow only when the query needs it: 52 → 110 → 167.

Copyright ©2026 Arizona Board of Regents

How it adapts

One token picks the shape of the reasoning



The modes differ in structure, not just length, and RL learns which one to start.

Copyright ©2026 Arizona Board of Regents

The idea

Run every mode, score it, train toward the winner

1 · Run every mode

One video, one question, **eight answers**.

- <DIRECT> o1, o2
- <CONCISE> o3, o4
- <DEEP> o5, o6

forced: the prompt sets the mode

- <ADAPTIVE> o7, o8

free: the model picks the mode

2 · Score each mode

From its two forced answers: is it **right**? Is it **grounded** (checked by a separate verifier model)? What does it **cost**?

	right	grounded	cost
<DIRECT>	0.55	0.20	**0.00**
<CONCISE>	**0.82**	0.65	0.08
<DEEP>	0.80	**0.70**	0.20

a medium question, one of our worked cases

3 · Pick the winner

First be **right**, then **grounded**, then **cheaper**.

- <DIRECT> often wrong
- <CONCISE> right and grounded, and cheaper
- <DEEP> as good, but costs more

Easy questions go to DIRECT, hard ones to DEEP. **Nobody labels the mode.**

4 · Align to the winner

The two free answers are rewarded for **picking the winner**:

- <ADAPTIVE> <CONCISE> + 0.8
- <ADAPTIVE> <DEEP> − 0.8

Reinforcement learning then moves the model **toward the winning mode**.

Comparing the modes **is the label**. The free choice matches the winner **40%** of the time after fine-tuning, and **75%** after reinforcement learning.

The model learns which reasoning wins, from its own answers.

Copyright ©2026 Arizona Board of Regents

Reward modeling

Seven reward functions, three questions

1 · Is it right?

R_{ans} The answer is **correct**, scored by its type: a choice, a number, a time span, a box.

R_{format} The output has the **format**: a mode tag, then the reasoning, then the answer.

R_{comply} A forced answer **sticks to its mode** and its word budget.

2 · Was it worth the cost?

R_{cost} Longer reasoning costs more, priced by **how hard the question is**: the share of the 8 answers that are right.

$R_{balance}$ The 8 answers **don't all collapse** onto one mode.

Easy · 80% of answers right

DEEP — costs 0.43

DIRECT — costs 0.03

→ DIRECT wins

Hard · 10% of answers right

DEEP — costs 0.19, scores 0.47 higher

→ DEEP wins

3 · Was the choice right?

R_{select} A free answer that picks **the winner**, the best mode for this question, gets +0.8; any other pick gets −0.8.

$R_{verifier}$ A separate **verifier** model scores whether the reasoning points to the right moments and objects.

Adding one job at a time

Qwen2.5-VL-7B

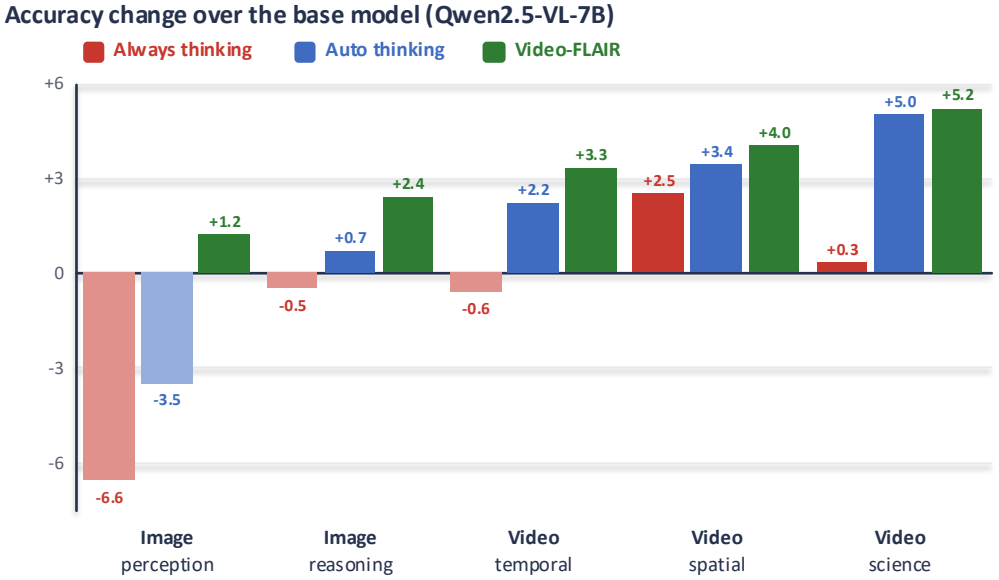
right + format	+ cost, mode, balance	+ right choice	+ verifier
412 tokens · EMMA **21.9**	**185** tokens · EMMA **23.0**	**70** tokens · EMMA **25.7**	**67** tokens · EMMA **27.3**

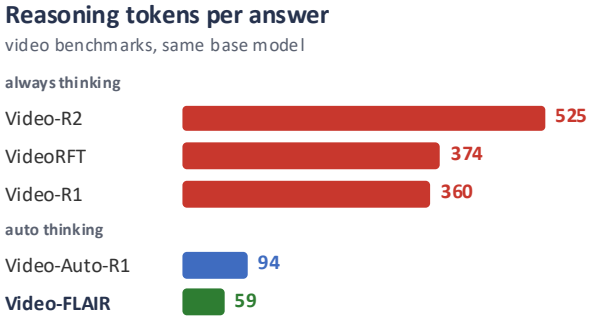
First be right, then be cheap for this question's difficulty, then pick the mode that wins.

Copyright ©2026 Arizona Board of Regents

Video-FLAIR

Always thinking vs. auto thinking





Over both base models: **95 tokens instead of 417**. On Qwen3-VL-8B, always thinking loses **14.5 points** on image perception.

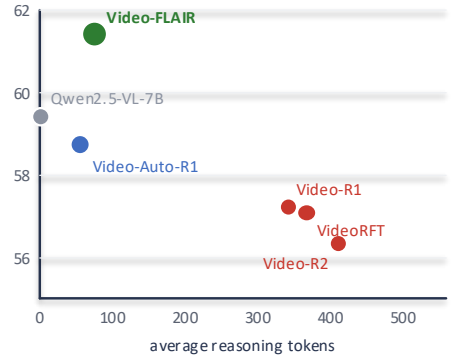
Only Video-FLAIR is above the base model in all five task types.

Copyright ©2026 Arizona Board of Regents

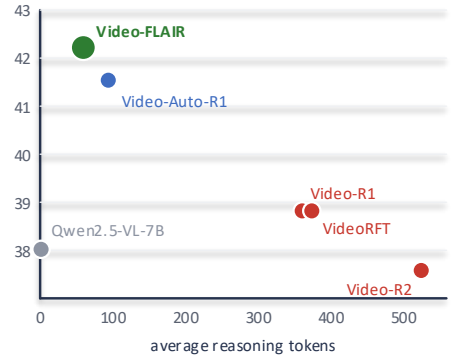
Results

More accurate with ~4× fewer tokens

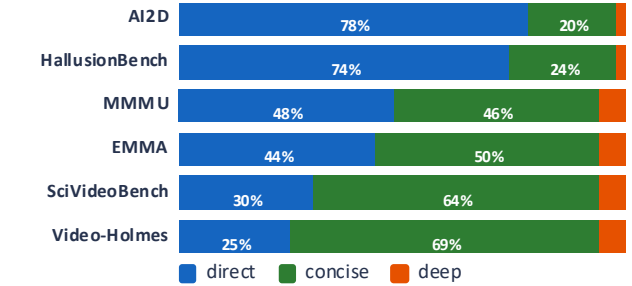
6 image benchmarks · mean accuracy



5 video benchmarks · mean accuracy



Which mode it picks



Perception-heavy benchmarks stay direct; video reasoning shifts to concise; deep stays ≤ 6%.

+5.4
MathVista over Qwen2.5-VL-7B

+4.8
Video-Holmes and Video-MMMU

95 vs 417
average tokens vs always-thinking, across both backbones

+8.4
EMMA on Qwen3-VL-8B, where OneThinker drops 11.5

Always-thinking models land below the base model on images; adaptive reasoning lands above it on both.

Copyright ©2026 Arizona Board of Regents

Ablations

What each choice buys

A fixed mode?

Mode at inference	EMMA	V-MMMU	Tokens
forced DIRECT	22.1	49.7	57
forced CONCISE	25.6	53.2	247
forced DEEP	**28.5**	55.8	963
Video-FLAIR	**27.3**	**56.9**	**67**

Forced DEEP wins EMMA by 1.2, at **14× the tokens**, and loses on Video-MMMU.

A simpler selector?

Picks the mode by	EMMA	V-MMMU
zero-shot prompt	25.2	51.5
router + fine-tuning	26.5	54.5
Video-FLAIR (RL)	**27.3**	**56.9**

A router trails by **2.4** on Video-MMMU: RL also improves the reasoning **inside** each mode.

Two modes: think or not?

Modes	EMMA	V-MMMU
think / no think	24.4	55.1
three modes	**27.3**	**56.9**

Merging concise and deep lands near Video-Auto-R1 (23.7, 54.2): **the kind** of reasoning matters, not just whether to think.

Adding one job at a time
Qwen2.5-VL-7B

right + format
412 tokens · EMMA **21.9**

→

+ cost, mode, balance
185 tokens · EMMA **23.0**

→

+ right choice
70 tokens · EMMA **25.7**

→

+ verifier
67 tokens · EMMA **27.3**

No fixed mode, simpler selector or two-mode switch matches three learned modes.

Copyright ©2026 Arizona Board of Regents



Thank you

- Three reasoning modes, picked per query
- Learned from the model’s own answers, with no mode labels
- More accurate with about 4× fewer tokens

Copyright ©2026 Arizona Board of Regents