

EVENT-TRIGGERED
COMMENTARY (OUTPUT)



GOAL! The winger delivers a low cross
and the striker finishes at the near post.

(spoken)

Generated only
when gate = SPEAK

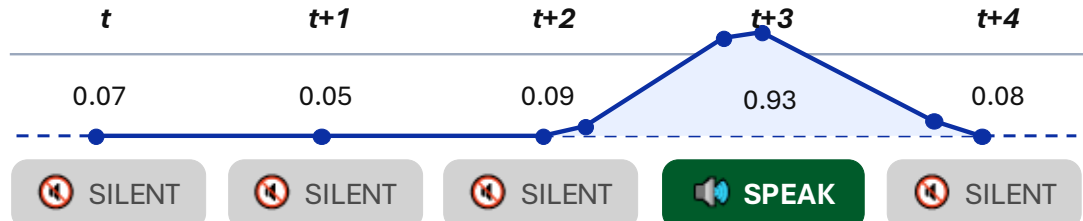
A. PREDICT GATE SIGNAL
(Should we speak?)

Gate Head

(Event Gate Predictor)

Predict gate probability over
rolling windows

Gate probability (p_{Speak})



Gate Decision
(Threshold)

$$p_{\text{Speak}} \geq \tau?$$
$$\tau = 0.5$$

YES
(SPEAK)

NO
(SILENT)

B. COMMENTARY DECODER
(Active only when gate opens)

LLM Decoder

(Commentary Generator)

What happens
in the match?

(Text Prompt)

Text Tokenizer

Text Tokens

No response

VISUAL FEATURES
(Shared)

LIVE MATCH STREAM
(Continuous Input)



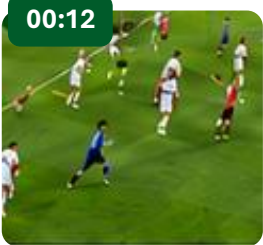
LIVE

Mage-ViT

Visual Encoder / Codec-Native Visual Tokenizer



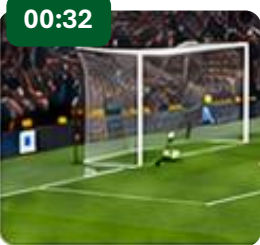
Kick-off



Build-up play



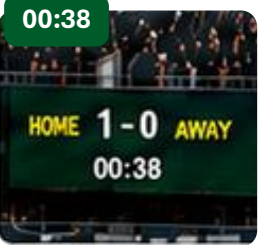
Through ball



Shot on goal



GOAL!



Scoreboard



Crowd erupts

Continuous stream of frames

Time

LEGEND



SILENT
(No output)



SPEAK
(Generate commentary)



Gate Signal
(p_{Speak})



Visual Tokens
(From Mage-ViT)



Text Tokens



Active path
(when gate = SPEAK)



Inactive path
(when gate = SILENT)