

THE FOLLOWING **PREVIEW** HAS BEEN APPROVED FOR
ALL AUDIENCES

CVPR 2016 Spotlight

MovieQA

Understanding Stories in Movies through Question-Answering



Makarand
Tapaswi



Yukun
Zhu



Rainer
Stiefelhagen



Antonio
Torralba



Raquel
Urtasun



Sanja
Fidler



UNIVERSITY OF
TORONTO



**Massachusetts
Institute of
Technology**

Visual QA

Understanding images

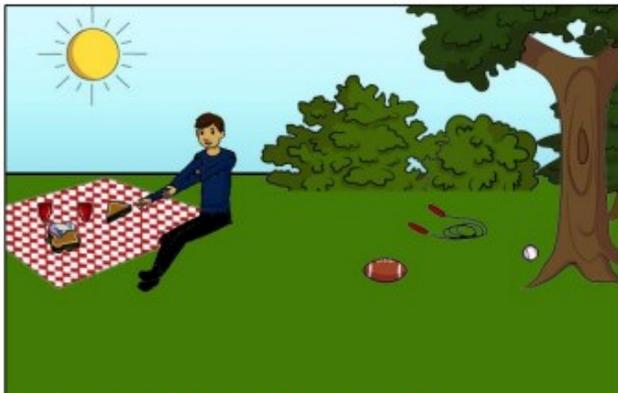
The best way to show that our “robots” really *understand* the scene is to check whether they can answer questions about it



What color are her eyes?
What is the mustache made of?



How many slices of pizza are there?
Is this a vegetarian pizza?



Is this person expecting company?
What is just under the tree?



Does it appear to be rainy?
Does this person have 20/20 vision?

What is ...
What color ...
What type ...
Is the ...
Are there ...
How many ...
Where is ...
Does the ...

MovieQA

Understanding *stories*

The Matrix has you



00:25:52 --> 00:25:57

Welcome, Neo. As you no doubt have guessed... I am Morpheus



00:40:42 --> 00:40:47

It exists now only as part of a neural-interactive simulation that we call the Matrix.



01:04:08 --> 01:04:09

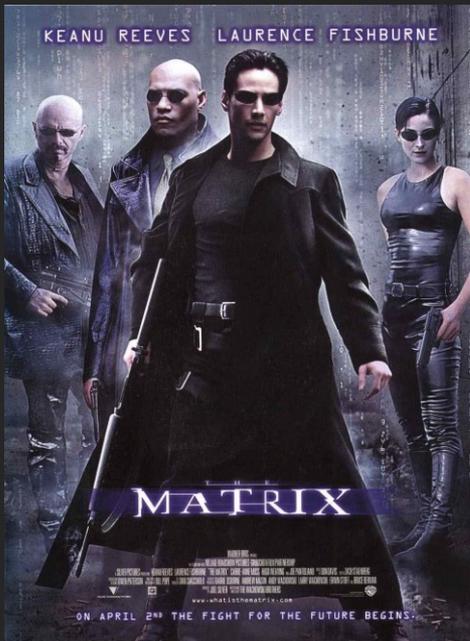
... you know what I realize? Ignorance is bliss.



02:08:38 --> 02:08:39

Where we go from there is a choice I leave to you

Questions: Why does Cypher betray Morpheus? How does Trinity save Neo?



Movie:

- 200,000 frames
- 2,000 shots
- 1,000 dialogs
- Long temporal dependencies
- Actions, interactions, emotions, intent

MovieQA

multiple sources of information (video and text)

Q. Who makes Indy return the crucifix after escaping from the grave robbers?

A1. *The local sheriff*

A2. Coronado

A3. No one, he keeps it

A4. The Boy Scout troop

A5. The grave robbers



PLOT

Indy escapes, but the local **sheriff** makes him return the **crucifix**.

DVS

Indy shows the **Cross**, more or less handing it to the **Sheriff** to make his point.
The **Sheriff** takes it casually.

SCRIPT

SHERIFF: You still got it?

INDY: Well, yes sir.

Indy shows the **CROSS**, more or less handing it to the **SHERIFF** to make his point.

The **Sheriff** takes it casually.

SHERIFF: I'm glad to see that

SUBTITLE

00:10:50 --> 00:10:52

You still got it?

00:10:52 --> 00:10:53

Well, yes, sir.

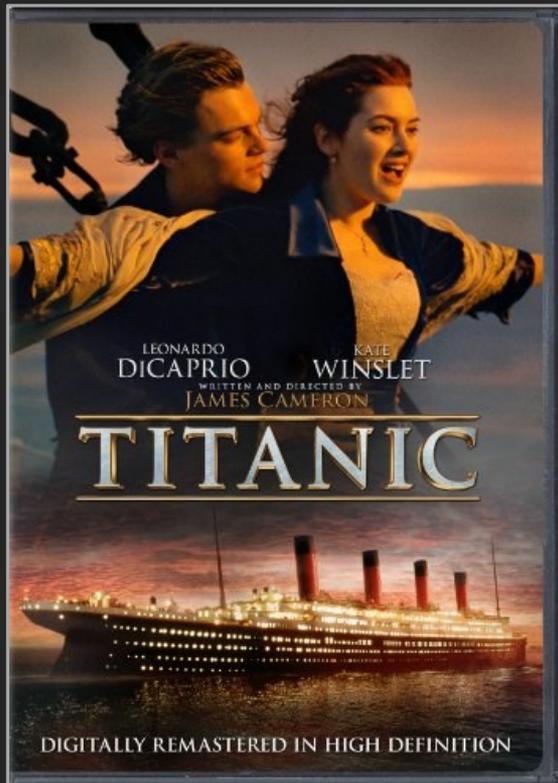
00:10:55 --> 00:10:59

I'm glad to see that because the rightful owner of this **cross**

MovieQA

answering with video clips

Titanic



Q: How does Jack meet his end?

A1: Drowns as he can't swim

A2: Freezes to death

A3: He is rescued but dies from frostbite

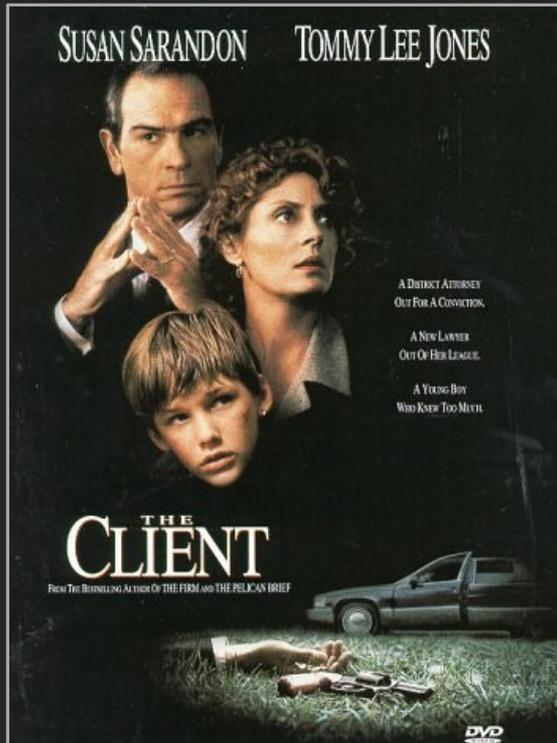
A4: Dies as an old man in bed

A5: Dies later in his life after obstacles

MovieQA

answering with dialogs

The Client



- I wish. Sit here.
- Don't try to swallow the smoke yet.
- You're not ready for that.
- You'll just choke and puke all over the place.
- Suck a little and blow.

Q: What are Martin and Ricky doing in the woods?

A1: They are cutting woods

A2: They are trying to kill each other

A3: They are smoking cigarettes

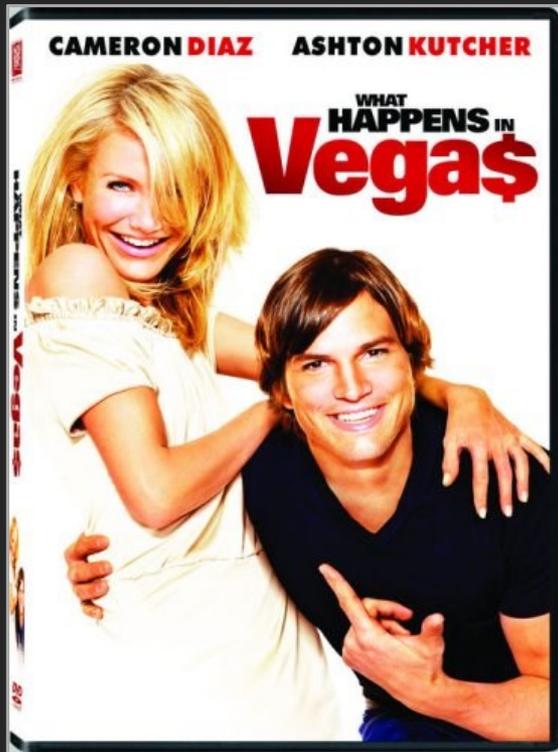
A4: They are hunting deer

A5: They are picking up blueberries

MovieQA

answering by reasoning

What Happens
in Vegas



Q: Why do Joy and Jack get married on that first night in Las Vegas?

A1: Because CVPR is the rest of the days

A2: They are vulnerable and drunk

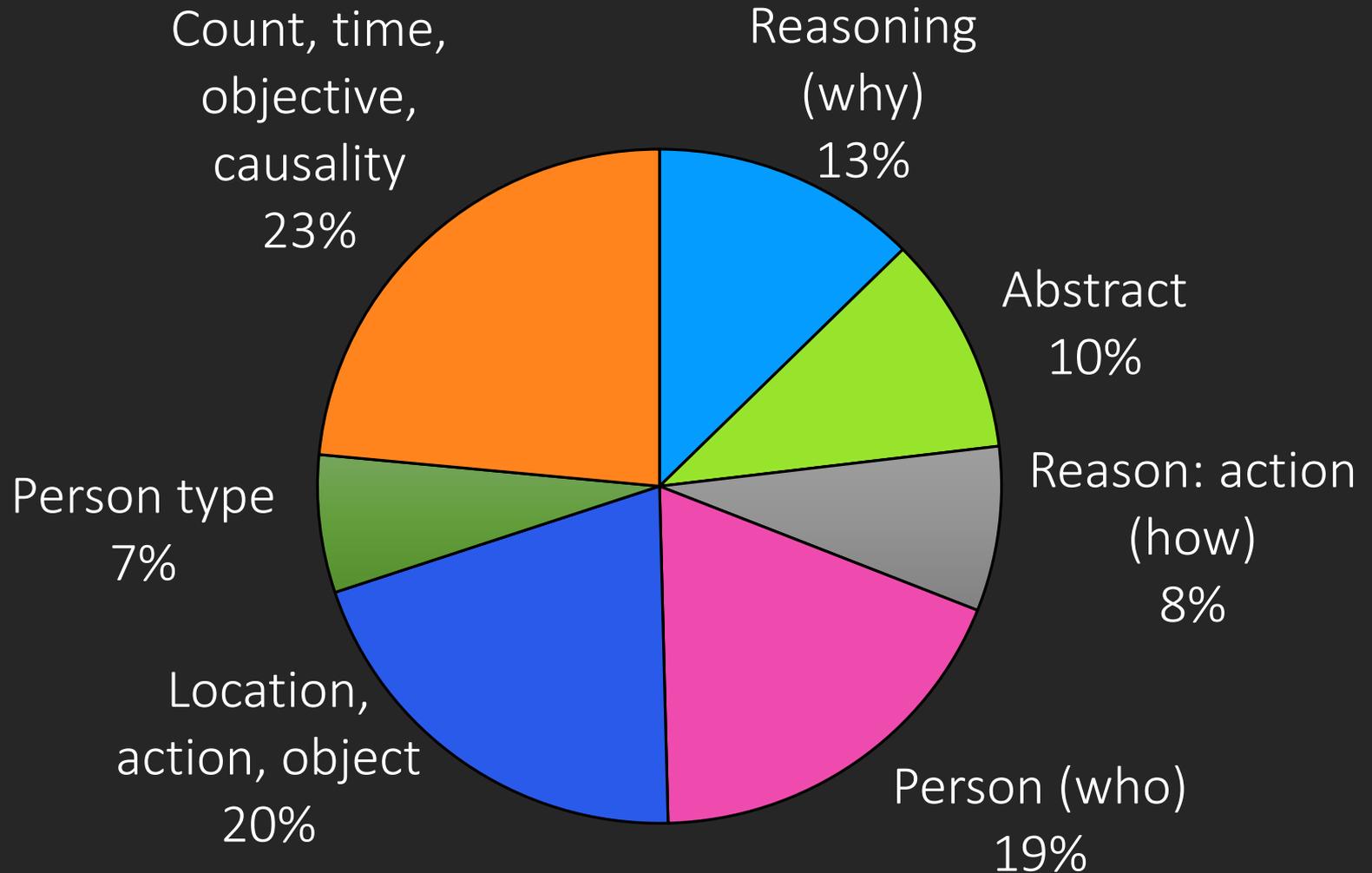
A3: Because they love each other

A4: To please their parents

A5: Because everyone gets married in Vegas

MovieQA

benchmark in numbers



MovieQA

benchmark in numbers

- 14,944 QAs
- 408 movies
- 1 correct, 4 deceiving answers per Q

- 6,462 QAs with video
- 6,771 video clips
- ~3m20s clip duration

MovieQA

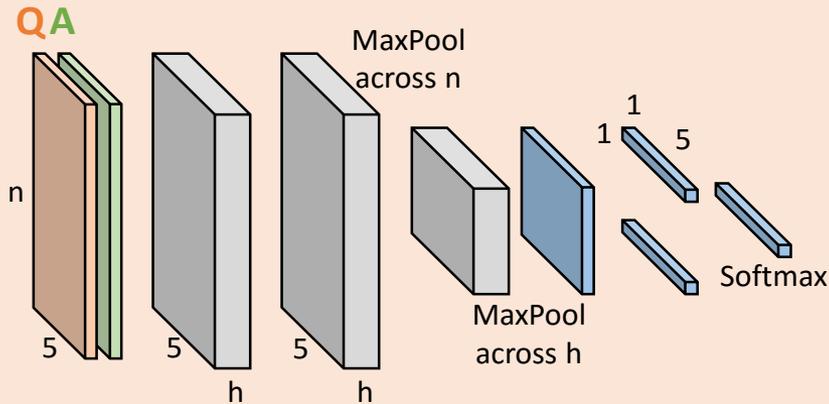
answering methods

General framework for multiple-choice question-answering

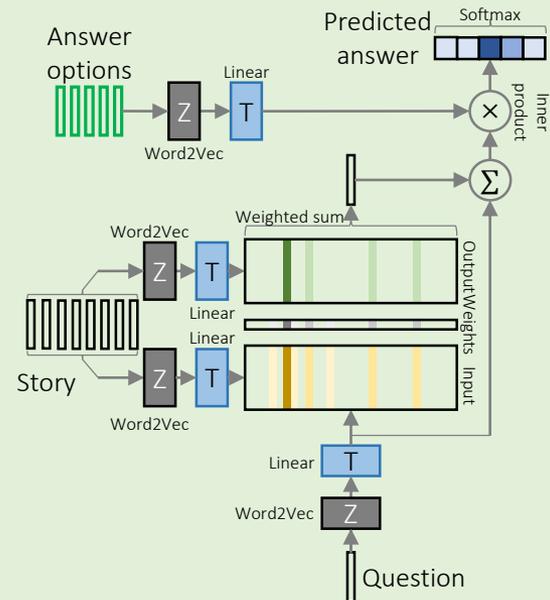
correct answer

$$\boxed{a} = \arg \max_a f(\text{story}, \text{question}, \text{answers})$$

story question answers



Searching Student with Convolutional Brain



Modified Memory Network

MovieQA

show me the numbers

Method	Story	Accuracy
Hasty Machine	<i>None</i>	25.3
Hasty Turker	<i>None</i>	24.7
Convolutional brain	Plot	56.7
MemN2N	Script	42.3
	Video clips	23.1

MovieQA

show me the numbers



Accuracy

25.3

24.7

56.7

42.3

23.1

MovieQA Benchmark is Live

show us what you got

[MovieQA](#)

[ABOUT](#)

[CHANGELOG](#)

[FAQ](#)

[EXAMPLES](#)

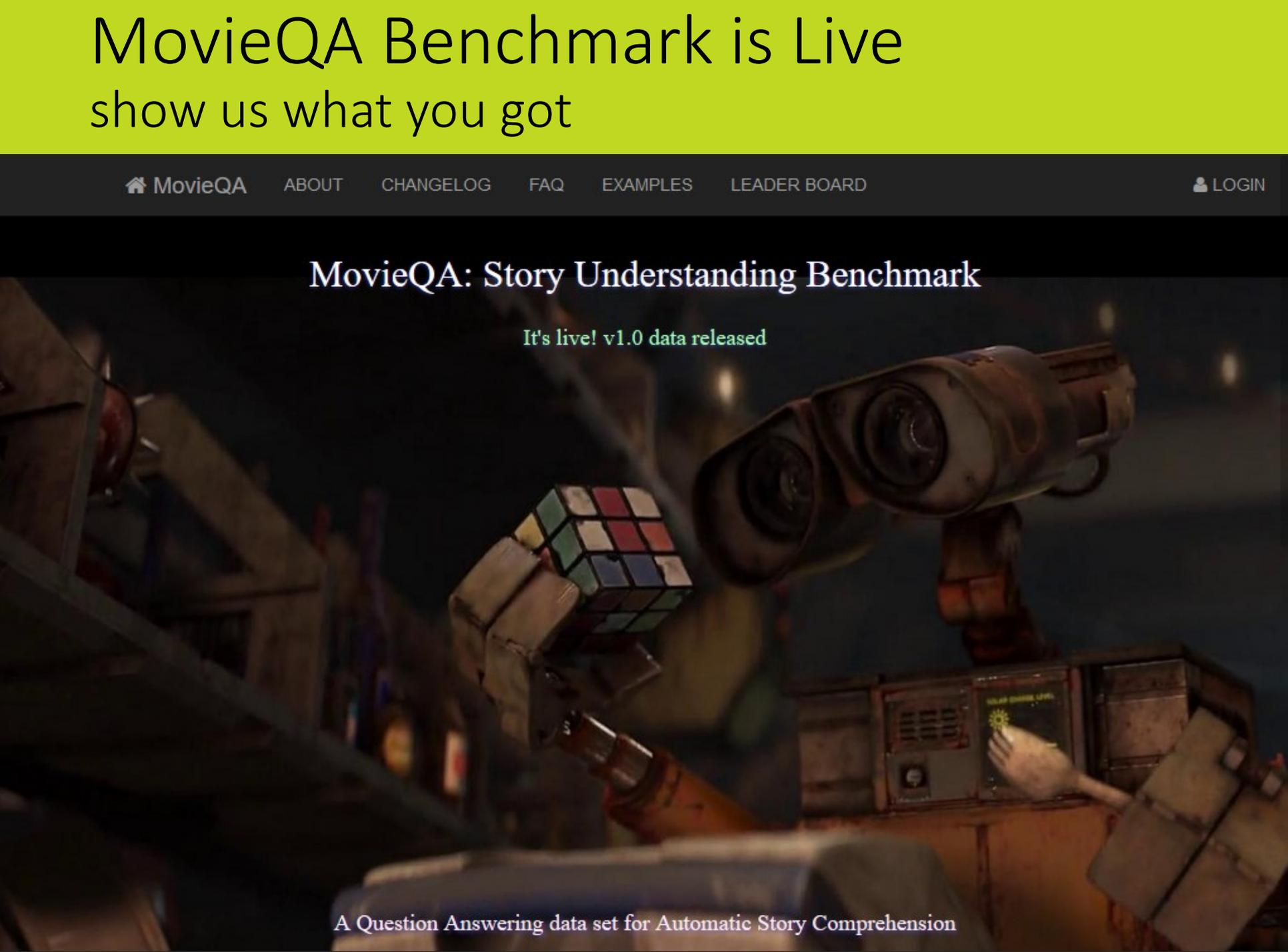
[LEADER BOARD](#)

[LOGIN](#)

MovieQA: Story Understanding Benchmark

It's live! v1.0 data released

A Question Answering data set for Automatic Story Comprehension

A scene from the movie 'The Iron Giant' showing the Giant robot holding a Rubik's cube. The robot is a large, rusted, mechanical creature with two large, circular eyes. It is holding a Rubik's cube in its right hand. The background is a dark, industrial setting with various pipes and machinery.

MovieQA

**Understanding Stories in Movies
through Question-Answering**

EXPERIENCE IT AT

Poster 4-1: 9

<http://movieqa.cs.toronto.edu>