



Grand Challenge:
Answering Visual Questions from Blind People

Outline

- Dataset and Tasks
- Challenge Description
- Challenge Winners
- Analysis of Results

Outline

- Dataset and Tasks
- Challenge Description
- Challenge Winners
- Analysis of Results

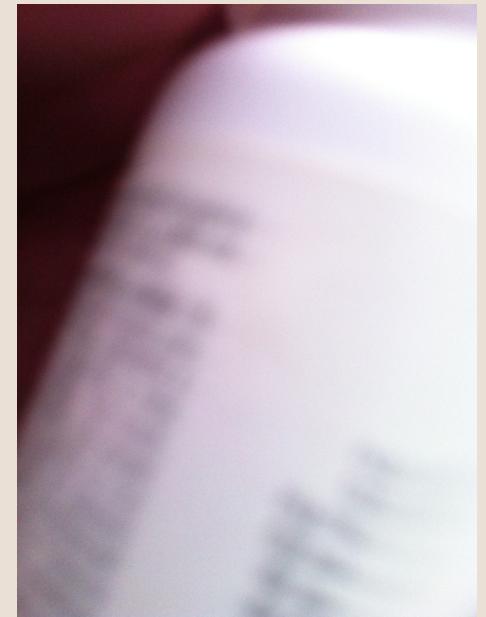
Task: Answer Blind People's Visual Questions



Is this shirt
clean or dirty?



Hi there can you
please tell me what
flavor this is?



What type of pills
are these?



What is this?

Dataset Source: VizWiz Application



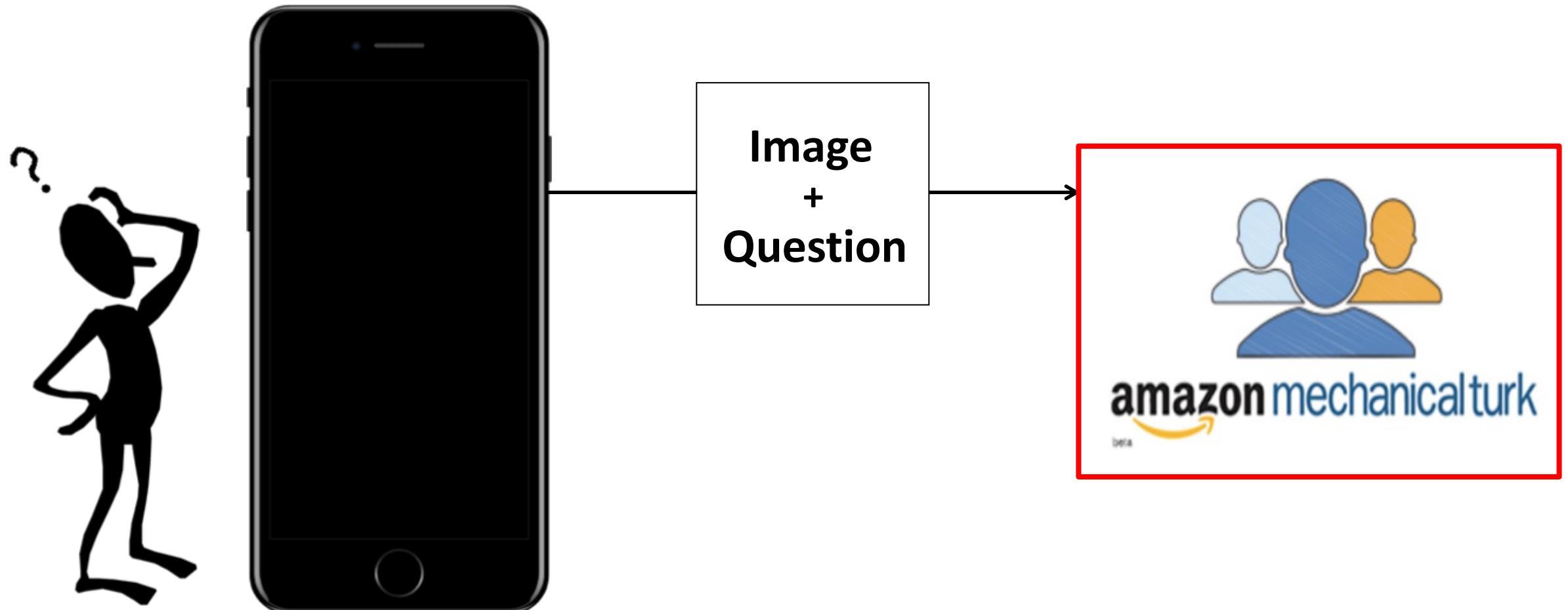
Dataset Source: VizWiz Application



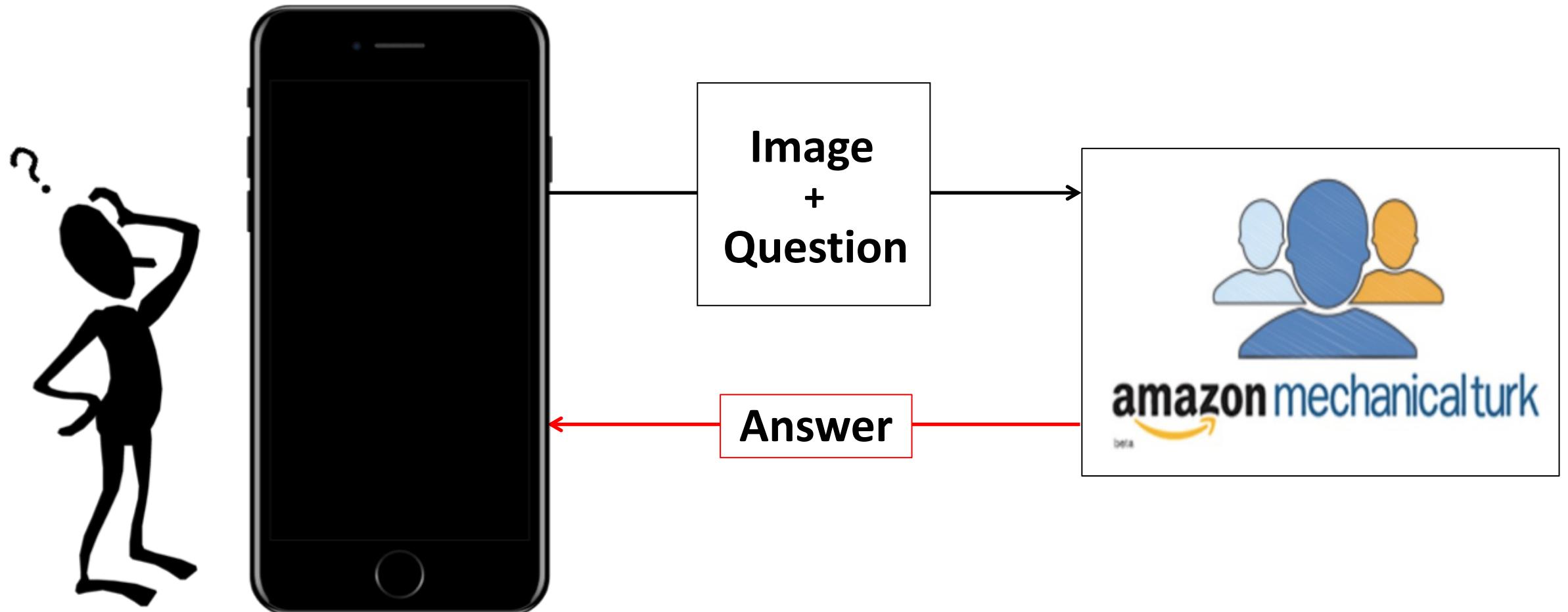
Dataset Source: VizWiz Application



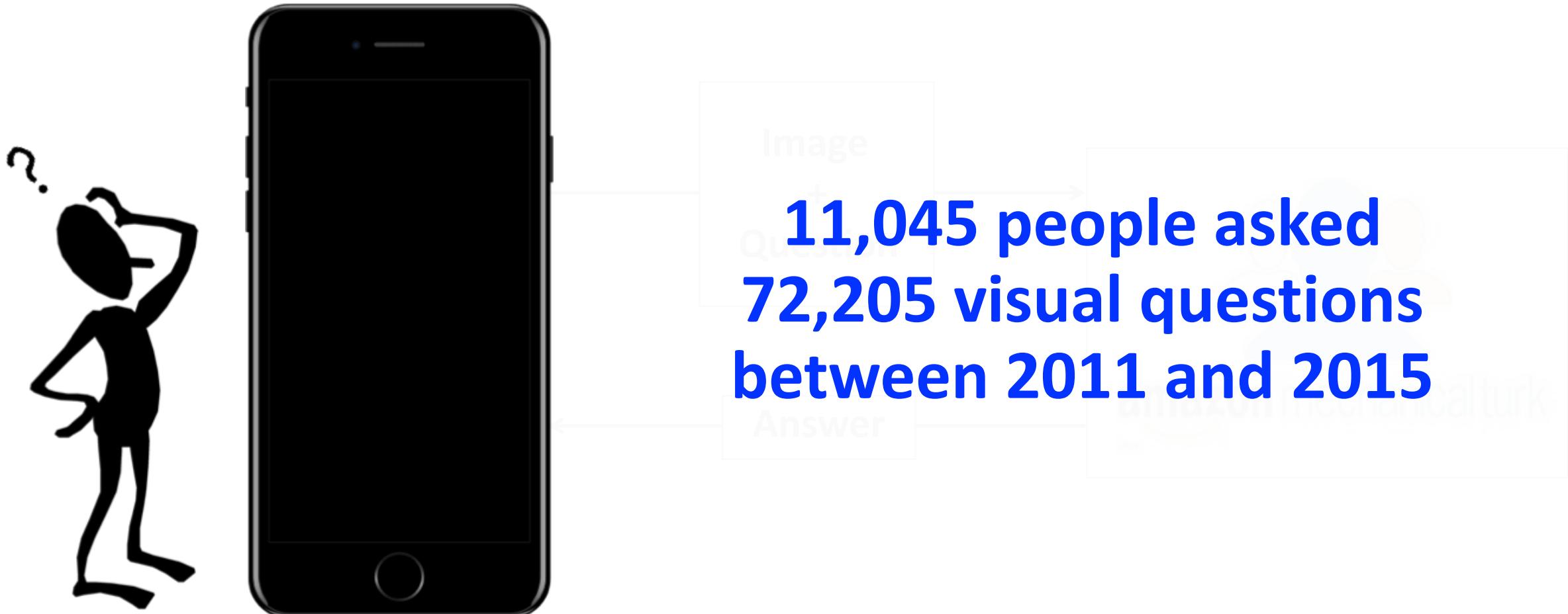
Dataset Source: VizWiz Application



Dataset Source: VizWiz Application



Dataset Source: VizWiz Application

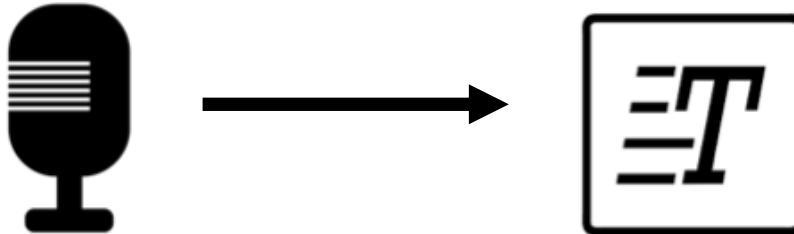


**11,045 people asked
72,205 visual questions
between 2011 and 2015**

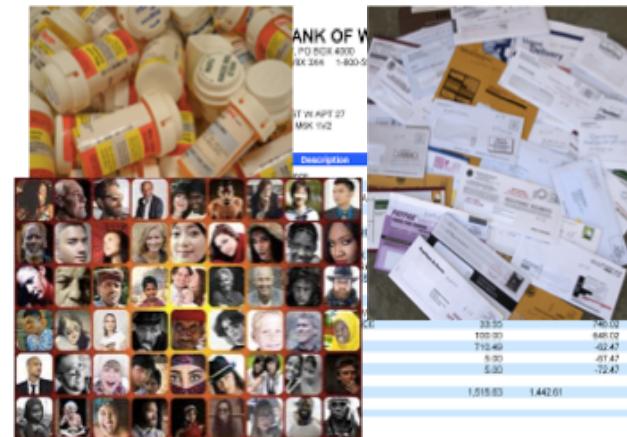
Dataset Creation

Anonymization

1. Transcription (removes voice)



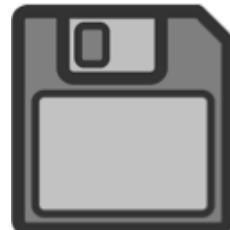
Filtering



Answer Collection



2. Re-save (removes metadata)



Dataset Creation

Anonymization

1. Transcription (removes voice)

Filtering

Answer Collection

31,173 image/question pairs and 311,730 answers

2. Re-save (remove metadata)
(Available at <http://vizwiz.org/data>)

Task #1: Predict Answer



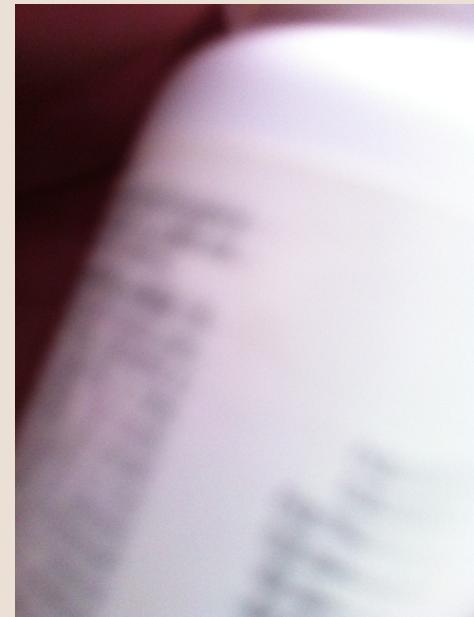
Is this shirt
clean or dirty?

clean



Hi there can you
please tell me what
flavor this is?

sweet pepper



What type of pills
are these?

unanswerable



What is this?

unanswerable

Task #2: Predict Answerability



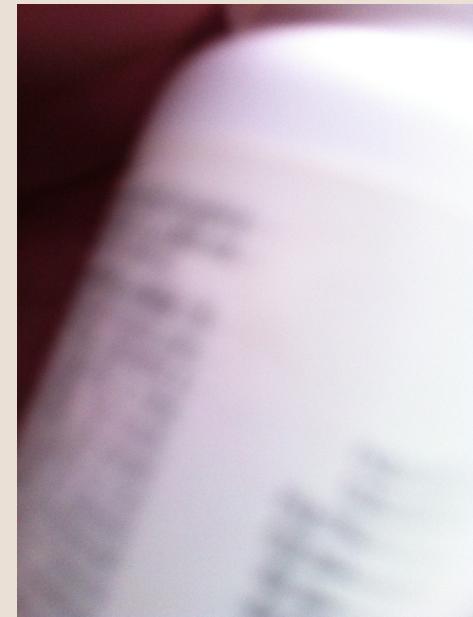
Is this shirt
clean or dirty?

answerable



Hi there can you
please tell me what
flavor this is?

answerable



What type of pills
are these?

unanswerable



What is this?

unanswerable

Outline

- Dataset and Tasks
- Challenge Description
- Challenge Winners
- Analysis of Results

VizWiz Challenge 2018

<https://evalai.cloudcv.org/>

EvalAI

- Dashboard
- All Challenges
- Hosted Challenges
- Create Challenge
- Participant Teams
- Host Teams

VizWiz VizWiz Challenge 2018
Organized by: VizWiz Team
Published

[Overview](#) [Evaluation](#) [Phases](#) [Participate](#) [View All Submissions](#) [Leaderboard](#)

VizWiz Challenge 2018

Q: Does this foundation have any sunscreen? A: yes	Q: What is this? A: 10 euros	Q: What color is this? A: green	Q: Please can you tell me what this item is? A: butternut squash red pepper soup	Q: Is it sunny outside? A: yes	Q: Is this air conditioner on fan, dehumidifier, or air conditioning? A: air conditioning
Q: What type of pills are these? A: unsuitable image	Q: What type of soup is this? A: unsuitable image	Q: Who is this mail for? A: unanswerable	Q: When is the expiration date? A: unanswerable	Q: What is this? A: unanswerable	Q: Can you please tell me what the oven temperature is set to? A: unanswerable

Dataset Splits

	Images	Questions	Answers
Training	20K	20K	200K
Validation	~3K	~3K	~30K
Test	8K	8K	80K

Dataset Splits: 3 Test Splits

- **Test-dev (development)**
 - Debugging and validation.
- **Test-challenge (competitions)**
 - Used to rank challenge participants.
- **Test-standard (publications year-round)**
 - Available for publications independent of the challenge.

Evaluation Metrics

- VQA: $\text{Acc}(\textcolor{red}{ans}) = \min \left\{ \frac{\#\text{humans that said } \textcolor{red}{ans}}{3}, 1 \right\}$
- VQ Answerability: Average precision for ‘unanswerable’

Outline

- Dataset and Tasks
- Challenge Description
- Challenge Winners
- Analysis of Results

Challenge Results – VQA (4 Teams)

3rd place: PAS-D

Denis Dushi; SAP SE, Politecnico di Milano, KTH Royal Institute of Technology, EIT Digital
Sandro Pezzelle; SAP SE, University of Trento
Moin Nabi; SAP SE, University of Trento
Tassilo Klein; SAP SE

2nd place: SKTBrain-SNU

Jin-Hwa Kim; SK T-Brain
Yongseok Choi; SK T-Brain
Sungeun Hong; SK T-Brain
Jaehyun Jun Seoul; National University
Byoung-Tak Zhang; Seoul National University & Surromind Robotics

1st place: FAIR A-STAR

*Vivek Natarajan; Facebook AI Research (FAIR)
*Yu Jiang; Facebook AI Research (FAIR)
Meet Shah; Facebook AI Research (FAIR)
Xinlei Chen; Facebook AI Research (FAIR)
Dhruv Batra; Facebook AI Research (FAIR)
Devi Parikh; Facebook AI Research (FAIR)
Marcus Rohrbach; Facebook AI Research (FAIR)

* Indicates equal contribution

Challenge Results – VQ Answerability (4 Teams)

3rd place: Baseline

Q+I Baseline [CVPR '18]

2nd place: PAS-D

Denis Dushi; SAP SE, Politecnico di Milano, KTH Royal Institute of Technology, EIT Digital

Sandro Pezzelle; SAP SE, University of Trento

Moin Nabi; SAP SE, University of Trento

Tassilo Klein; SAP SE

1^{rst} place: FAIR A-STAR

*Vivek Natarajan; Facebook AI Research (FAIR)

*Yu Jiang; Facebook AI Research (FAIR)

Meet Shah; Facebook AI Research (FAIR)

Xinlei Chen; Facebook AI Research (FAIR)

Dhruv Batra; Facebook AI Research (FAIR)

Devi Parikh; Facebook AI Research (FAIR)

Marcus Rohrbach; Facebook AI Research (FAIR)

* Indicates equal contribution

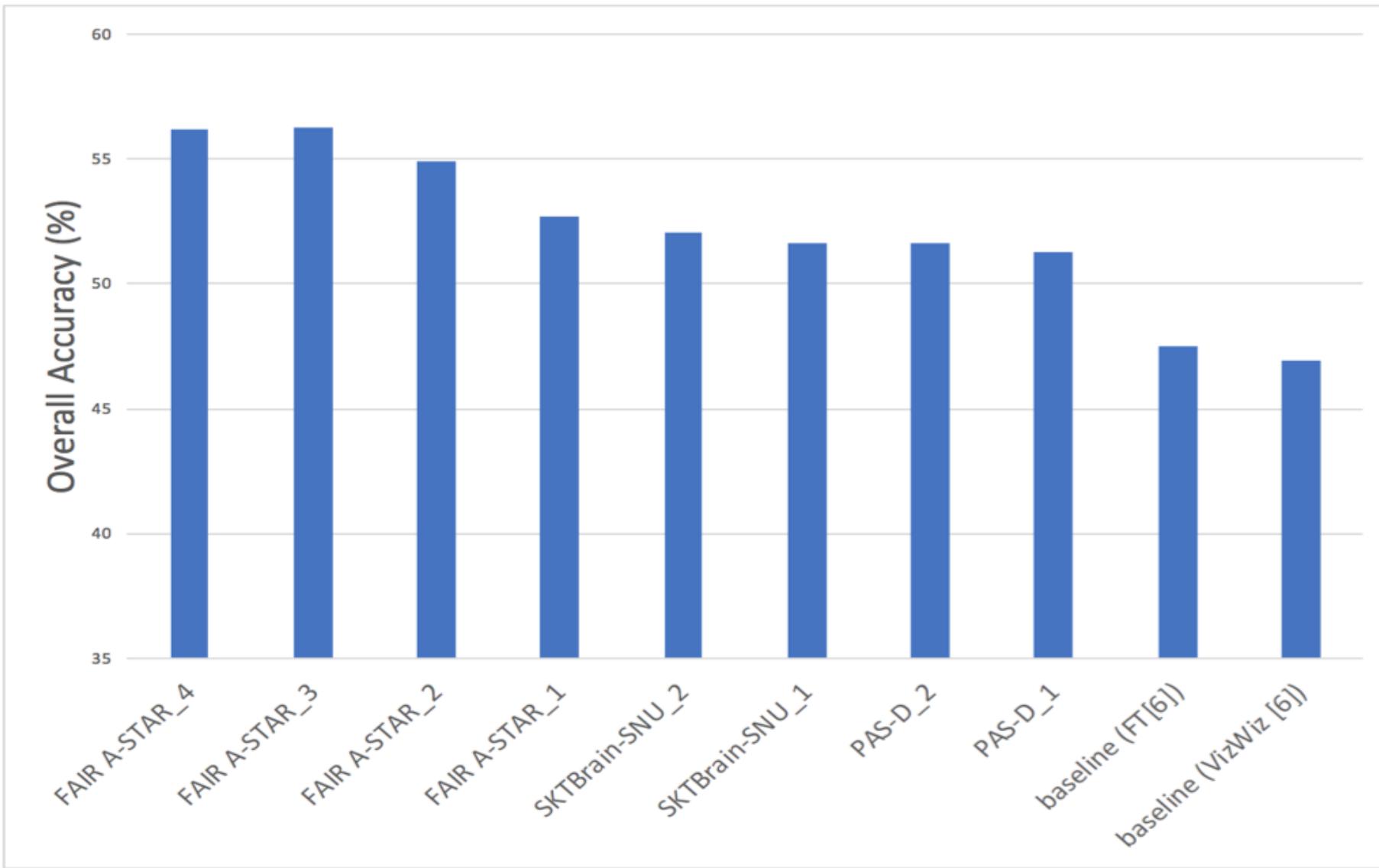
Outline

- Dataset and Tasks
- Challenge Description
- Challenge Winners
- Analysis of Results: Thanks to Qing Li (liqing@ucla.edu)!

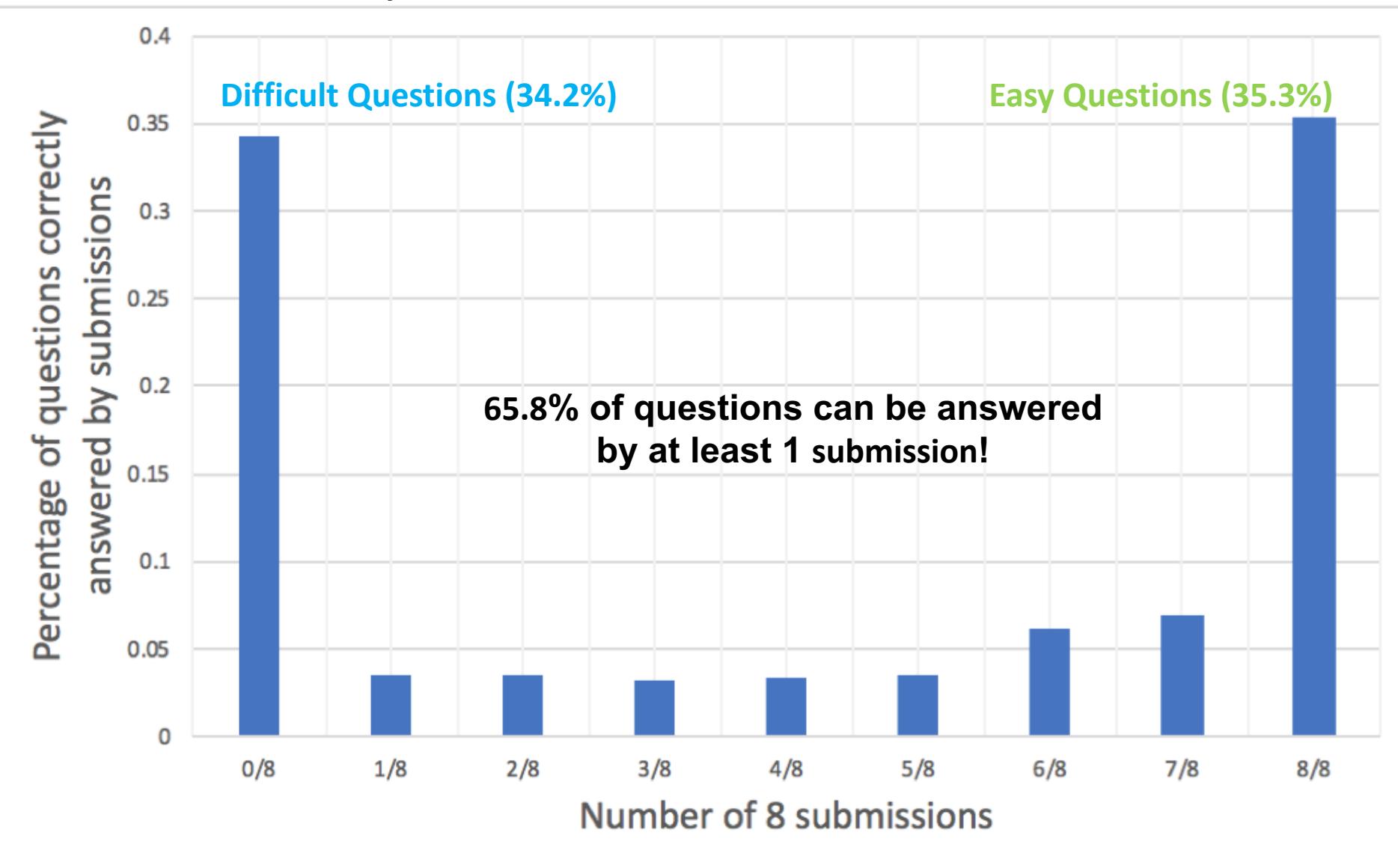


Dept. of Statistics, UCLA

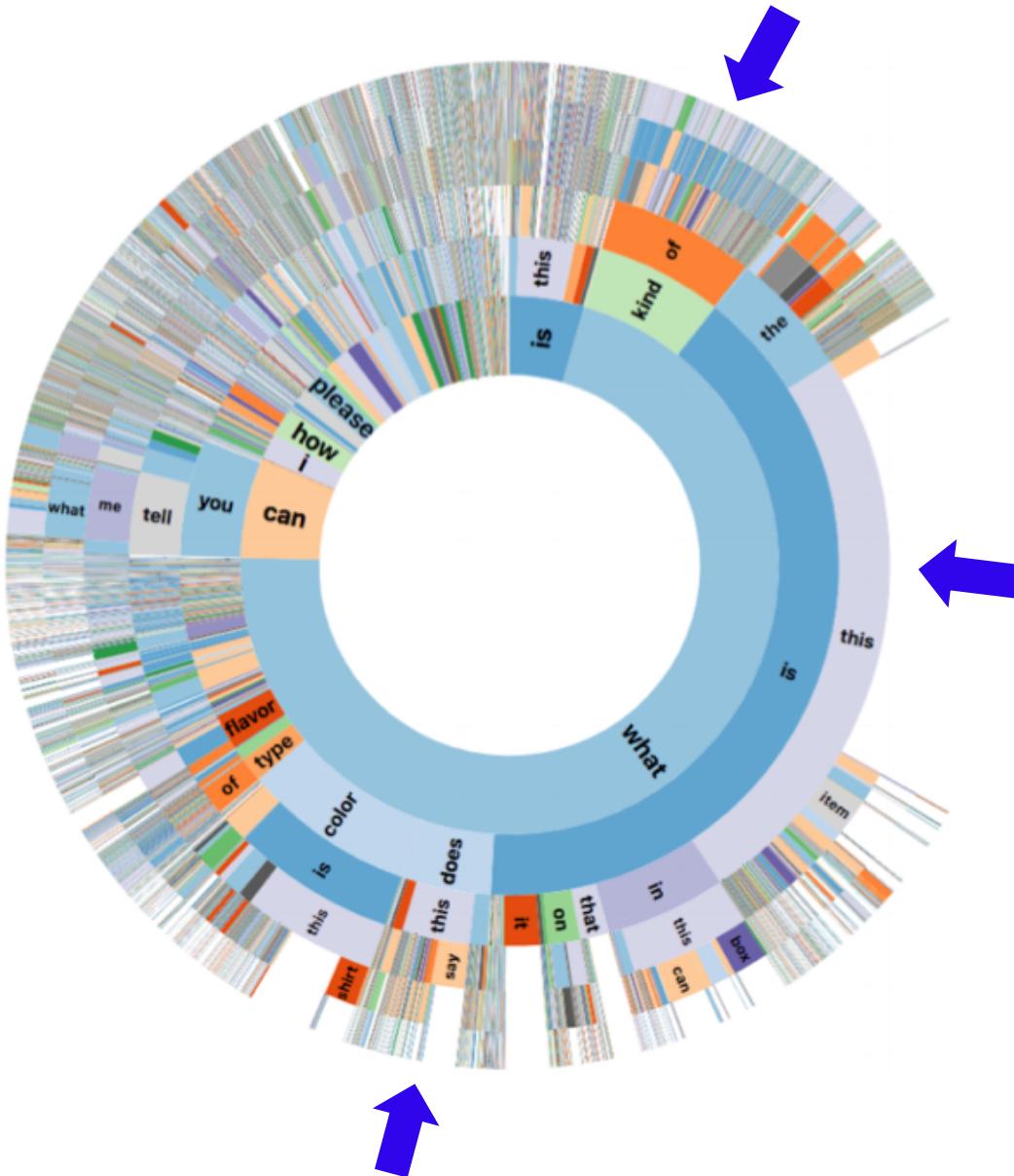
VQA Accuracy



Easy vs Difficult Questions

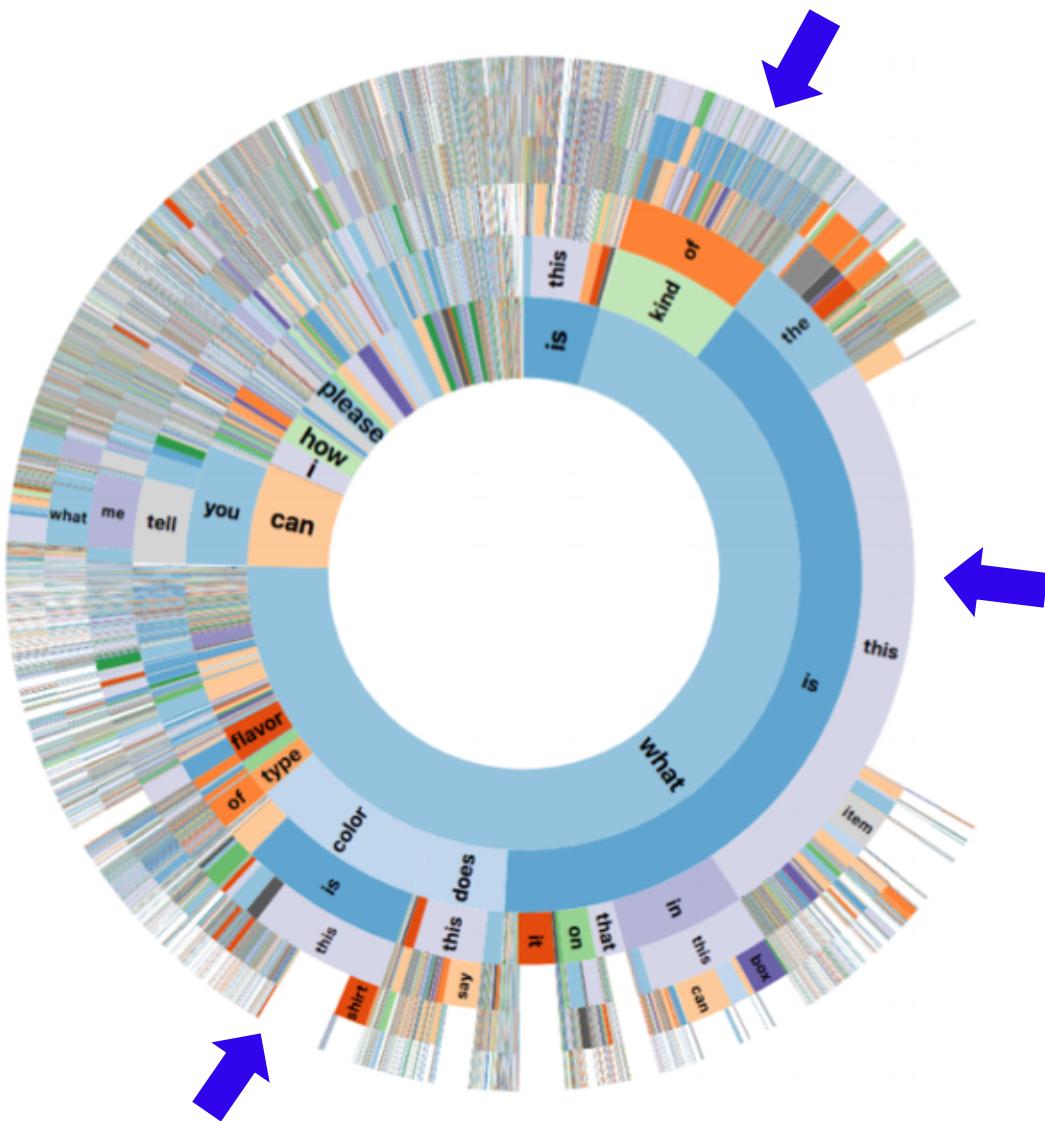


Difficult Questions



What is this?
What is this item?
What is it?
What does this say?
What kind of ...
What does this ...
What is the ...
Can you tell ...
What color is ...

Easy Questions



What is this?
What color is ...
What does this say?
What is in this can?
What is it?
What kind of ...
What is the ...
Can you tell ...
What does this ...

Difficult Questions



Q: What is this?

GT A: mtn dew can

FAIR A-STAR: soda



Q: What's in this bottle?

GT A: smoked spanish paprika

FAIR A-STAR: unanswerable



Q: Can you tell me what these buttons say?

GT A: food cancel

FAIR A-STAR: unanswerable



Q: What's on the display?

GT A: 27 3

FAIR A-STAR: unanswerable



Q: What is this?

GT A: waffles

FAIR A-STAR: food



Q: What is the vitamin type, brand and dosage?

GT A: niacin spring valley 500mg

FAIR A-STAR: unanswerable

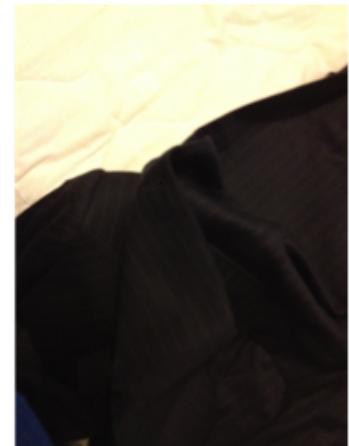
Easy Questions



Q: What kind of tea is this?
GT A: unanswerable
FAIR A-STAR : unanswerable



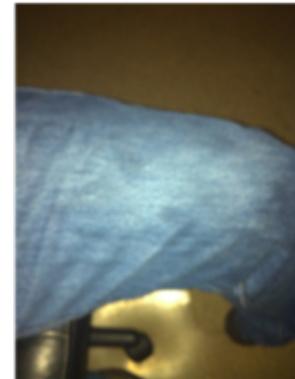
Q: What brand is this medication and what are the instructions?
GT A: unsuitable
FAIR A-STAR : unsuitable



Q: What color is this?
GT A: black
FAIR A-STAR: black



Q: how old is this camera?
GT A: unanswerable
FAIR A-STAR : unanswerable



Q: What kind of wine is this?
GT A: unanswerable
FAIR A-STAR unanswerable



Q: What color is this?
GT A: blue
FAIR A-STAR : blue

Answer Type Analysis

- FAIR A-STAR is top performer on ‘Yes/No’, ‘Other’ and ‘Overall’
- SKTBrain-SNU is top performer on ‘Unanswerable’
- PAS-D is top performer on ‘Number’

	Test				
	Yes/No	Number	Unanswerable	Other	Overall
FAIR A-STAR	72.89	25.48	86.00	38.41	56.29
SKTBrain-SNU	69.14	19.05	86.15	31.57	52.03
PAS-D	67.43	25.95	79.72	34.60	51.63
baseline (FT[6])	66.90	22.00	77.60	29.40	47.50

Red arrows and percentages indicate performance differences relative to the baseline (FT[6]):

- Yes/No: 6% increase from 66.90 to 72.89
- Number: 3.7% increase from 22.00 to 25.95
- Unanswerable: 8.6% decrease from 77.60 to 86.00
- Other: 9% increase from 29.40 to 38.41
- Overall: 8.8% increase from 47.50 to 56.29

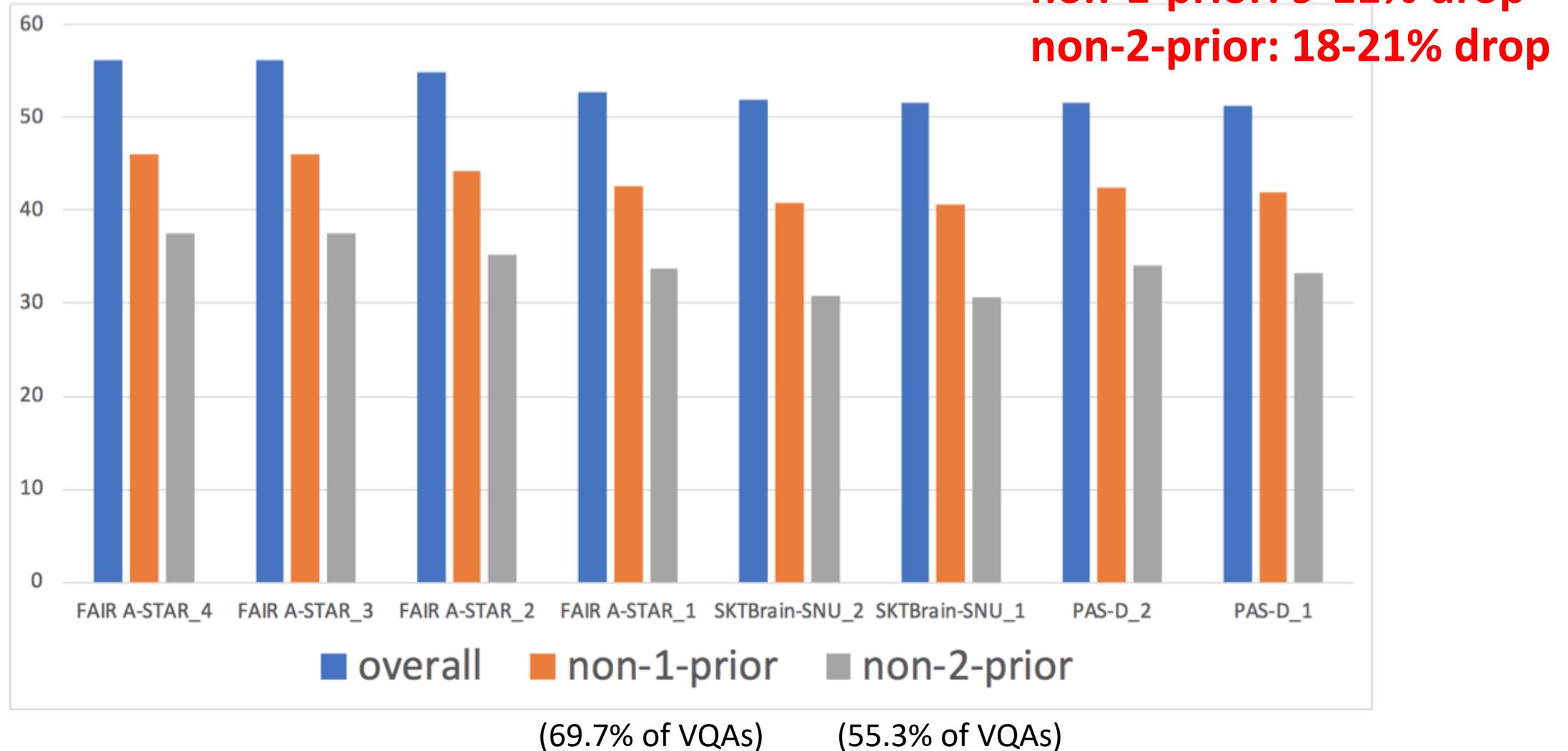
Are Models Drive by Priors?

Restrict to VQs with non-popular answers (for each question) in training:

- 1-Prior: Test answers are not the top-1 most common in training
- 2-Prior: Test answers are not the top-2 most common in training

Are Models Drive by Priors?

non-1-prior: 9-11% drop
non-2-prior: 18-21% drop



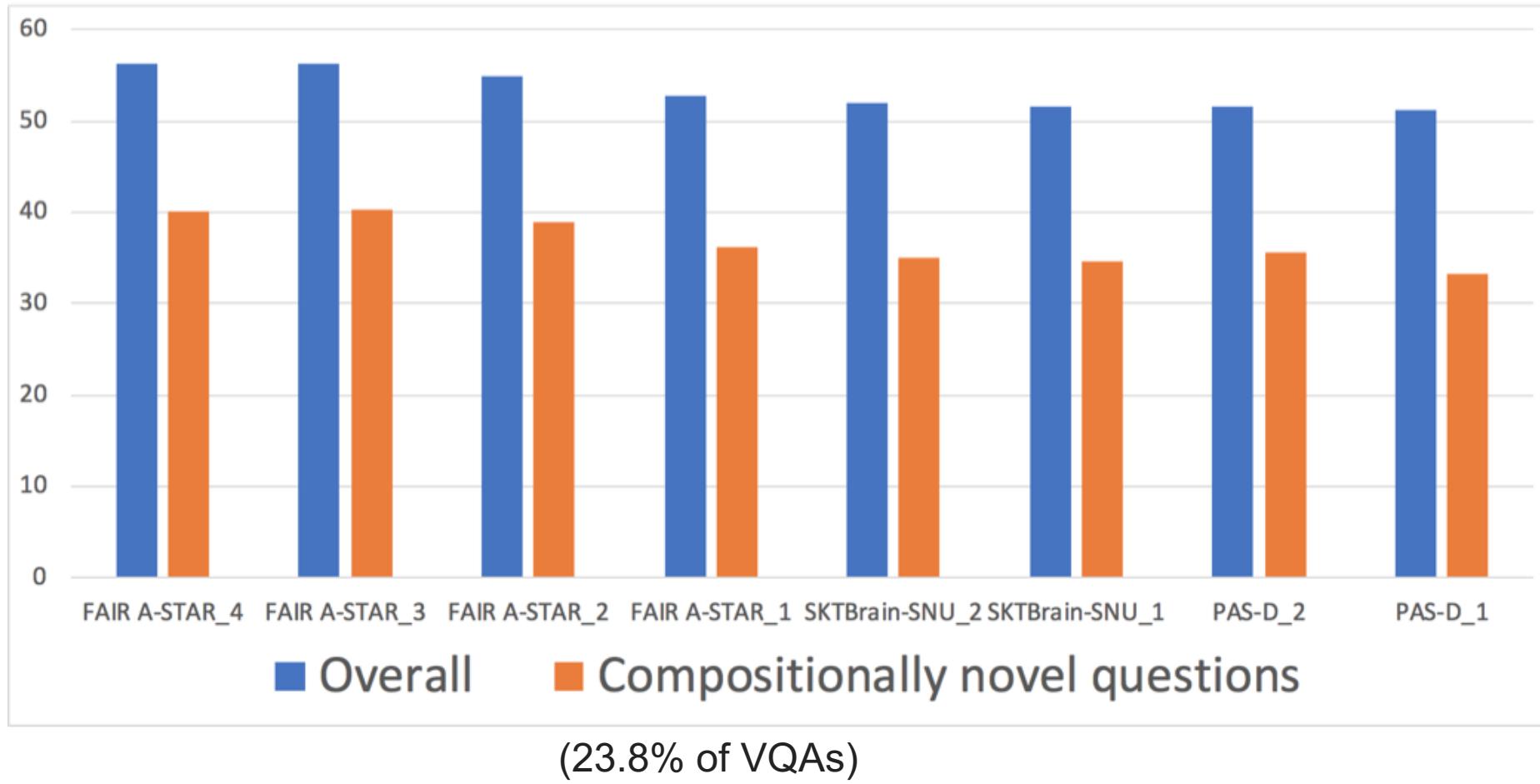
Are Models Compositional?

Restrict to VQs which are compositionally novel (excludes unanswerable VQs):

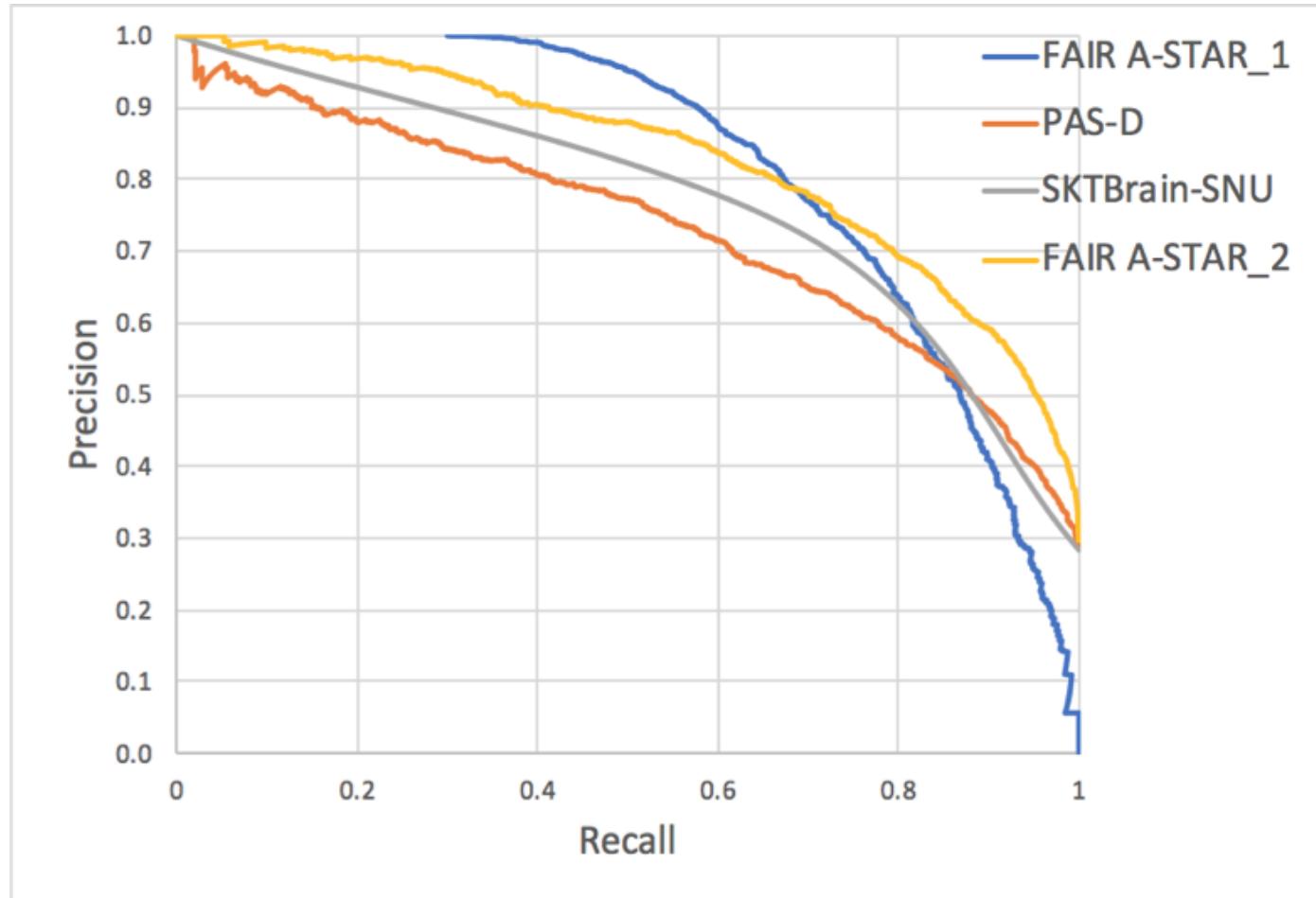
- The QA pair is not seen in training
- Constituting concepts (words) are seen in training
- 1-Prior: Test answers are not the top-1 most common in training
- 2-Prior: Test answers are not the top-2 most common in training

Are Models Compositional?

16-18% drop



VQ Answerability Performance



Test	Average Precision
FAIR A-STAR_2	82.89
FAIR A-STAR_1	82.05
PAS-D	73.11
SKTBrain-SNU	58.84
Baseline (Q+I)	71.70

Success Cases



Q: What is this?

A: mtn dew can

Answerable: 1

FAIR A-STAR: 0.9949



Q: What's in this bottle?

A: smoked spanish paprika

Answerable: 1

FAIR A-STAR: 0.9534



Q: What's on the display?

A: 27 3

Answerable: 1

FAIR A-STAR: 0.9798



Q: What brand is this medication and what are the instructions?

A: unsuitable

Answerable: 0

FAIR A-STAR: 0.0401



Q: What is this?

A: waffles

Answerable: 1

FAIR A-STAR: 0.9824



Q: What kind of tea is this?

A: unanswerable

Answerable: 0

FAIR A-STAR: 0.2734

Failure Cases



Q: What's in this can?

A: unanswerable

Answerable: 0

FAIR A-STAR: 0.7598



Q: What's in this can?

A: unanswerable

Answerable: 0

FAIR A-STAR: 0.9258



Q: Is the name of this supplement showing and what is it?

A: unanswerable

Answerable: 0

FAIR A-STAR: 0.8259



Q: What is the expiry date?

A: 07 12 11

Answerable: 1

FAIR A-STAR: 0.3338



Q: What is this?

A: unsuitable

Answerable: 0

FAIR A-STAR: 0.9543



Q: What brand is this cake?

A: giant eagle

Answerable: 1

FAIR A-STAR: 0.1276

Outline

- Dataset and Tasks
- Challenge Description
- Challenge Winners
- Analysis of Results

Next Up...

- 10:30-10:50am: Overview of challenge, winner announcements, & results analysis
- 11-11:30am: Talks by challenge winners (7 min Talk + 3 min Q&A)
 - 11-11:10: [SKTBrain-SNU](#)
 - 11:10-11:20: [PAS-D](#)
 - 11:20-11:30: [FAIR A-STAR](#)
- 11:30-12:30pm: Poster session (common poster area: room 2607; marked with workshop name)
- 12:30-1:45pm: Lunch