

VizWiz VQA: Beyond Question Answering



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: 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

Q: Who's this guy?
A: Jeffrey P. Bigham

Carnegie Mellon University
@jeffbigham



Human-
Computer
Interaction
Institute



VizWiz VQA: Beyond Question Answering



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: Who's this guy?
A: Jeffrey P. Bigham



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

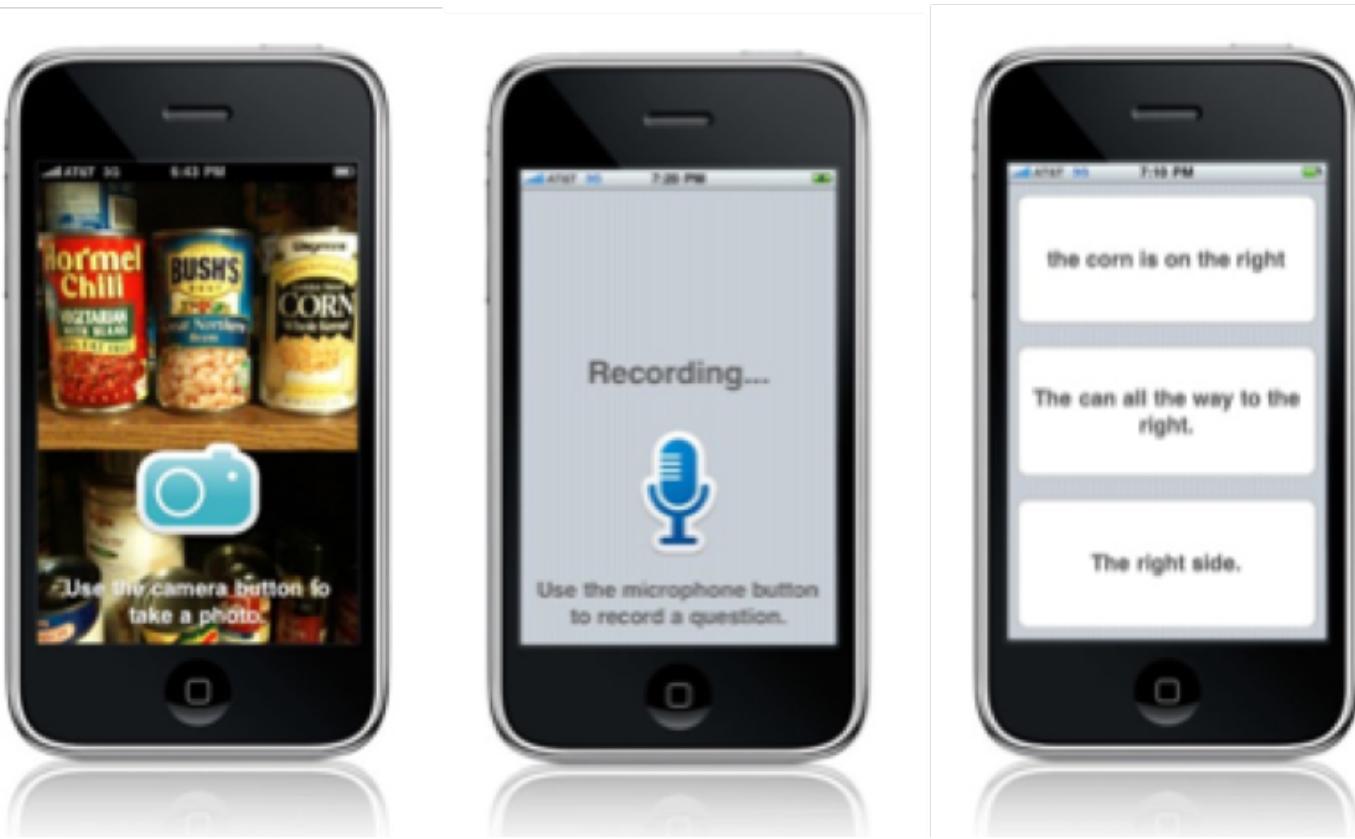
Carnegie Mellon University
@jeffbigham



Human-
Computer
Interaction
Institute



What do blind people want to know?

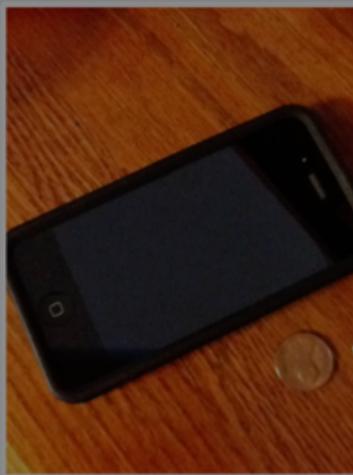


73. What is this can



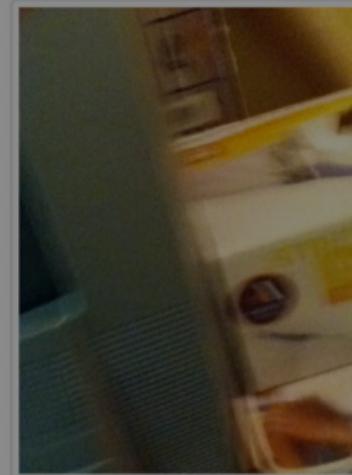
77. (no audio)

74. Hi is there an error message on this screen; and what does the screen say; or at least as much of it as you can see. Thank you



78. What's this?

75. Hello could you please tell me what's on this box? Thanks a lot

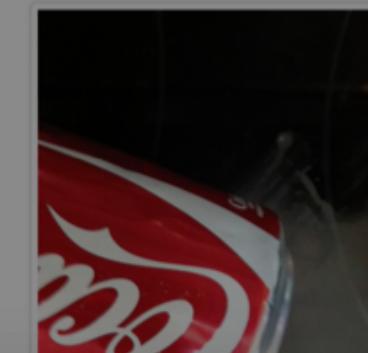
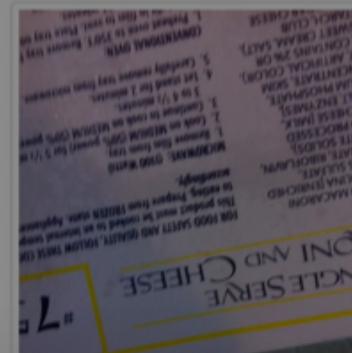
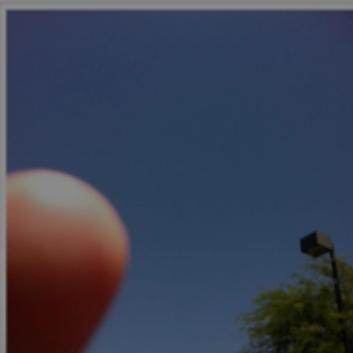
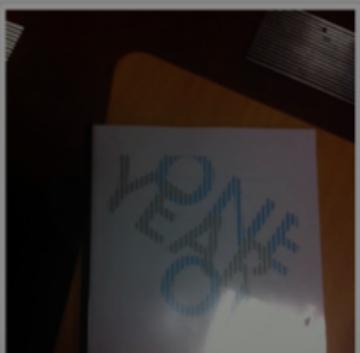


79. What is that please

76. Could you please tell me what is in this jar thank you



80. -this envelope addressed to

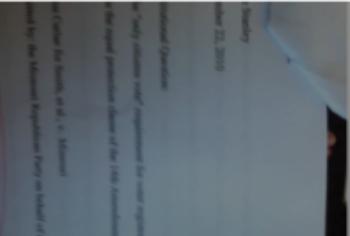




145. Can you identify this for me please



146. What color is this shirt



147. What's the title of this essay



148. What is the color



149. Is this set to 'fan'



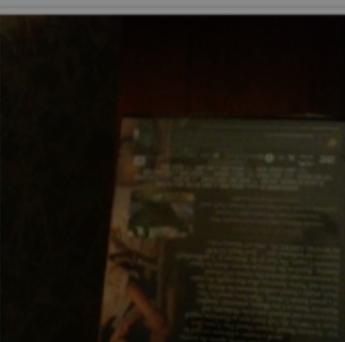
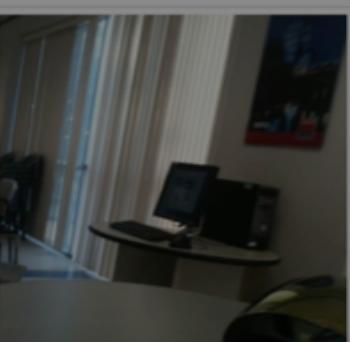
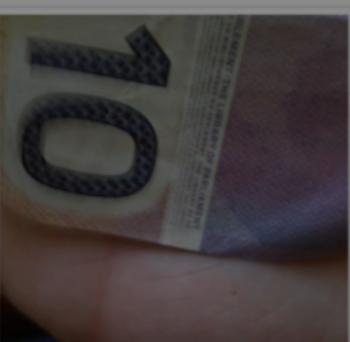
150. Is the air conditioner set to air; fan; or dehumidifier



151. What is the phone number for the Contact Centre



152. -s how to cook these





157. What is this?



158. Are you a computer?



159. What does this shirt say



160. What's on this page?



161. What is in this can.



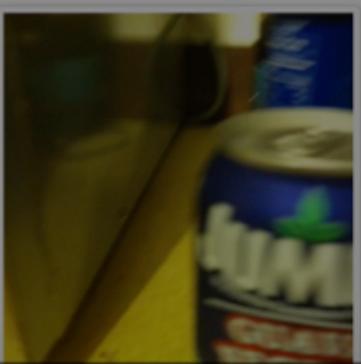
162. - his name.



163. What does this part of the sky look like

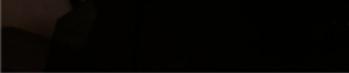


164. What's this and how long do I cook it for?

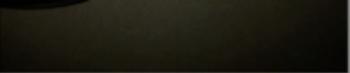




185. What's in this box



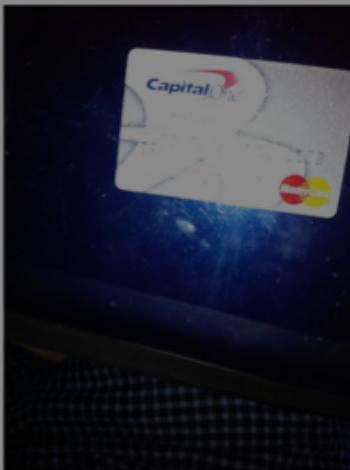
186. What is the top button on the left?



187. I'd like to know what this can is; please



188. Is this diet or regular coke? Thanks.



189. - card number.



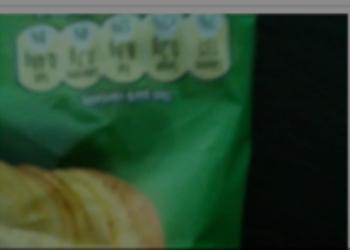
190. Okay I'm just curious what the controls are for this machine that's in front of me; I can feel that there are two rows of buttons; so can you tell me what the buttons are from left to right for each row?



191. What's the weather like outside on this photo? Thank you.



192. What is in this box?





245. Are these blue or green balls?



246. What is that?



247. What is this



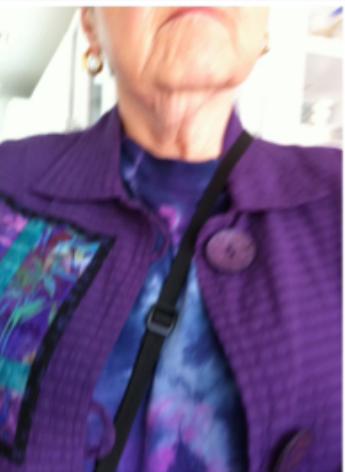
248. (no audio)



249. This is much lighter



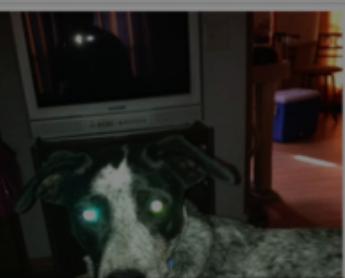
250. What's in the tin



251. Hi I would like to know if all my colors go together; I know I have some purples in my shirt; and I want to make sure I have the right pouch on my waist. Thank



252. What is this stand? Thank you.



845. (no audio)



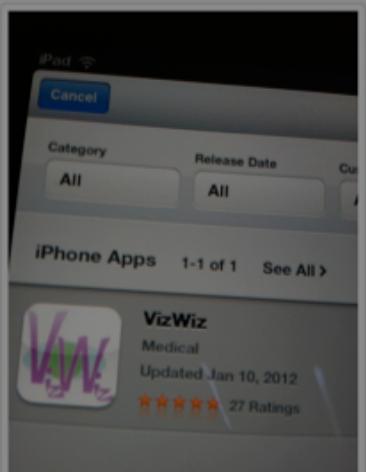
846. Please fully describe what you see in this image; thank you.



847. Yes; I would like to know what this jar contains; please. Thank you.



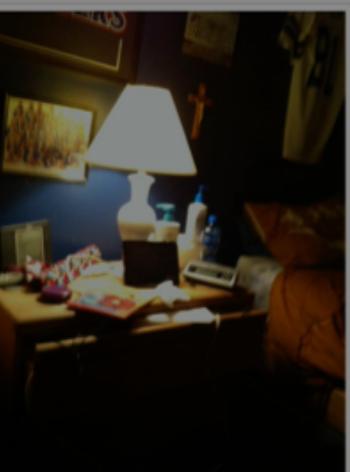
848. What is this item?



849. (no audio)



850. Can you please tell me the progress of the disk scan that's taking place?



851. Is my light on?



852. What is the name of this bottle?



It is difficult to say both speak dialects of the same language. Eight North America during the seventeenth century by gland who settled along the Atlantic coast from Maine to edge of these colonies consisted of dialects reflecting the geographic division of their former home country. English exhibits a remarkable degree of uniformity. Homogeneity resulted from the mingling of settlers from various they pushed westward; since World War II it has been due mobility of Americans. Today few people live in the which they were born; most move from one place to another jobs, marry, or retire. Nevertheless, certain regional differences are well known and readily recognizable when those of Boston, Virginia, or Texas. Vocabulary may be identifying where older speakers from rural areas have come the dragonfly is referred to in most of Virginia as snake飞, in western Pennsylvania as snake feeder, in eastern North Carolina, in New England as (devil's) darning needle, in coastal Florida as ear sewer, and so on. People also speak varieties not only according to their regional and social backgrounds but also according to the environment in which



Crowdsourcing Physical Access



Crowdsourcing Accessibility to the Physical World

Amazon Mechanical Turk - Mozilla Firefox
File Edit View History Bookmarks Tools Help
<https://www.mturk.com/mturk/accept?hitId=2XMZJ8es> Google

amazon mechanical turk Artificial Intelligence beta Account Settings | Sign Out | Help Your Account HITs Qualifications 41,186 HITs available now

All HITs | HITs Available To You | HITs Assigned To You

Search for HITs containing that pay at least

Comment and vote on an article. Easy!
Requester: Product Search Reward: \$0.03 per HIT Duration: 3 hours
Qualifications Required: HIT approval rate (%) is not less than 80

Which door is the women's restroom?

Click here to hear the original request



Finished with this HIT? Let someone else do it?
[Submit HIT](#) [Return HIT](#)

Done www.mturk.com

VizWiz Dataset

vizwiz.org/data



Check out Seeing AI!

VizWiz Challenge Jeffrey

[Overview](#) [Dataset](#) [Challenge](#) [Publications](#) [Contact Us](#)

Answering Visual Questions from Blind People

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: Who is this mail for? A: unanswerable	Q: When is the expiration date? A: unanswerable	Q: Can you please tell me what the oven temperature is set to? A: unanswerable

Jeffrey P. Bigham
@jeffbigham

Using Devices

User + AI + Crowds



186. What is the top button on the left?



20. Which knob; is it the left or the right; is for temperature control?



362. Which button says power? And have I got it set to delicate or daily?



190. Okay I'm just curious what the controls are for this machine that's in front of me; I can feel that there are two rows of buttons; so can you tell me what the buttons are from left to right for each row?



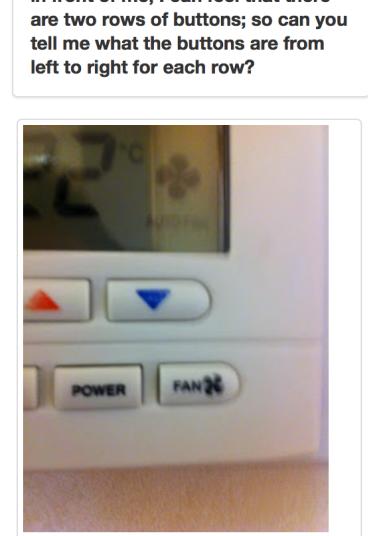
221. Could you tell me what the time remaining on this microwave is; it should be in seconds; like 33 seconds or 1 minute and 33 seconds or something like that. Thank you.



92. What is this machine and what are the buttons labeled as



233. Okay let's try this a second time; can you tell me all the buttons on this crockpot; what they are; left to right; up and down; and also -



665. (no audio)

Bigham, Jeffrey, et al. "VizWiz: nearly real-time answers to visual questions." UIST 2010.
Brady, Erin, et al. "Visual challenges in the everyday lives of blind people." CHI 2013.

X

“Take photo”



A. Initial Crowdsourced Segmenting and Labeling

Web Server & Database



Segmenting



Labeling

B. Real-Time Recognition and Control

“Weight Defrost”

Computer Vision Server

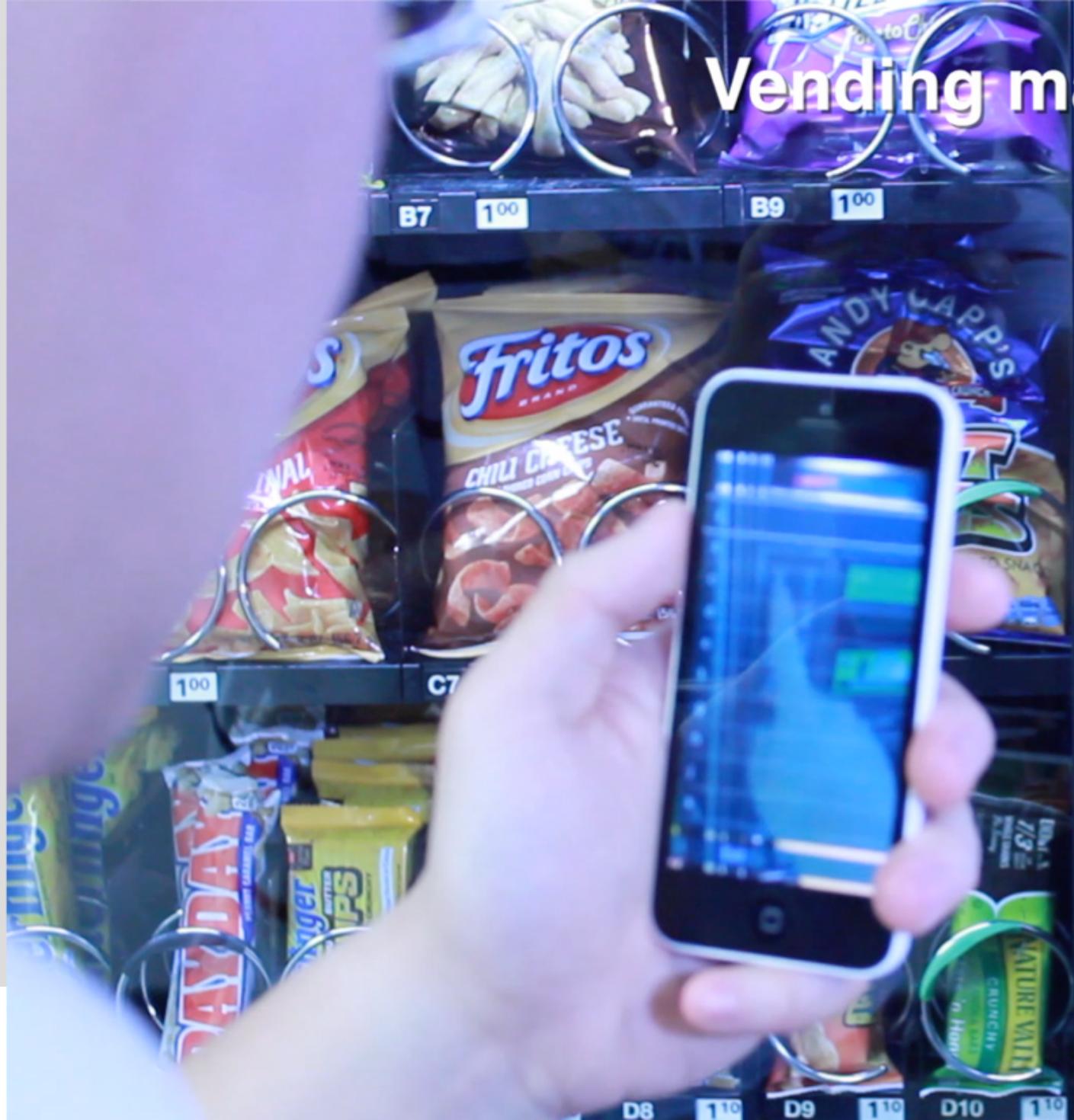


Local Client

Remote Services

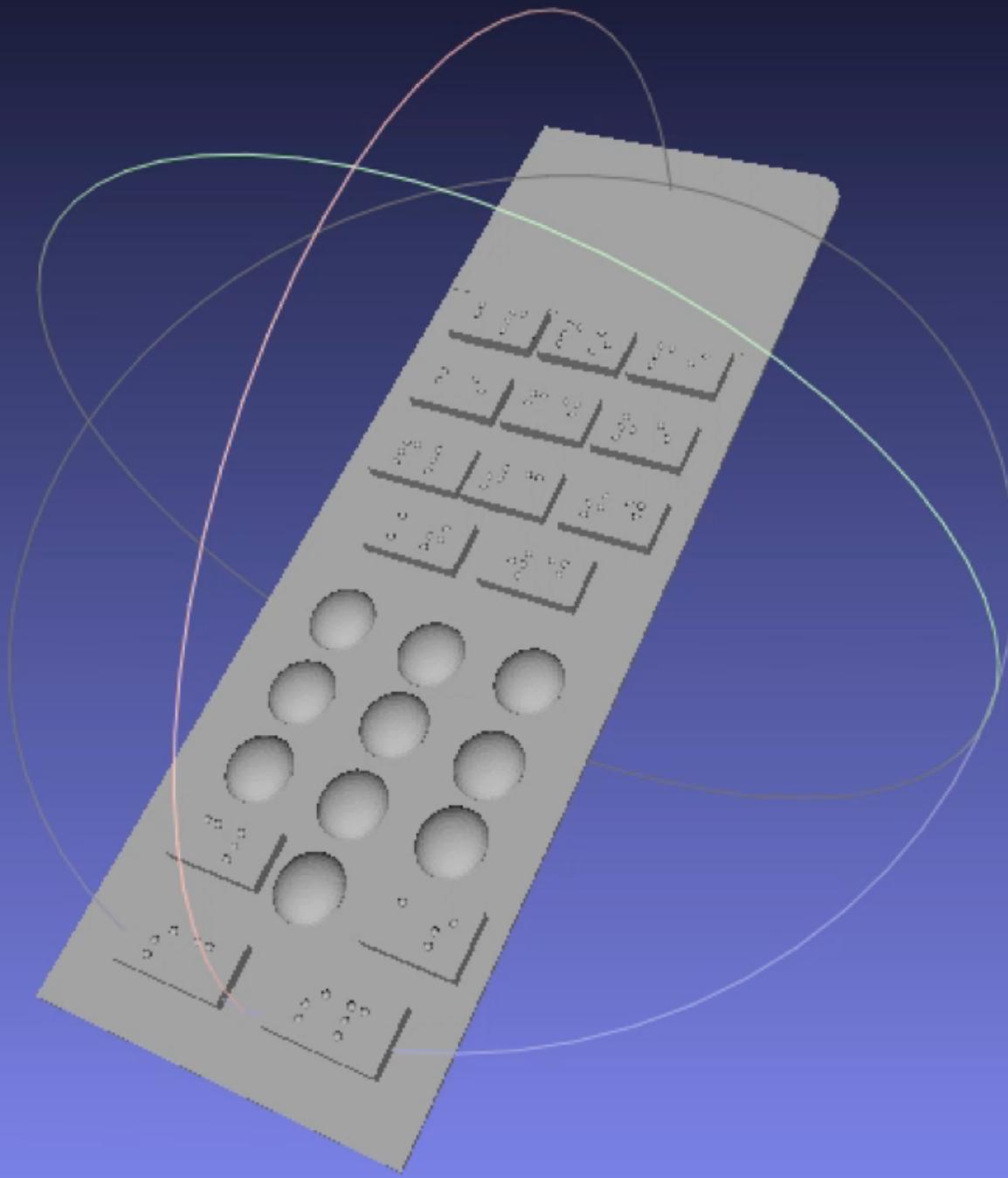
- Object Localization
- Fingertip Detection
- Information Lookup
- Providing Feedback
- Providing Guidance

Vending machine

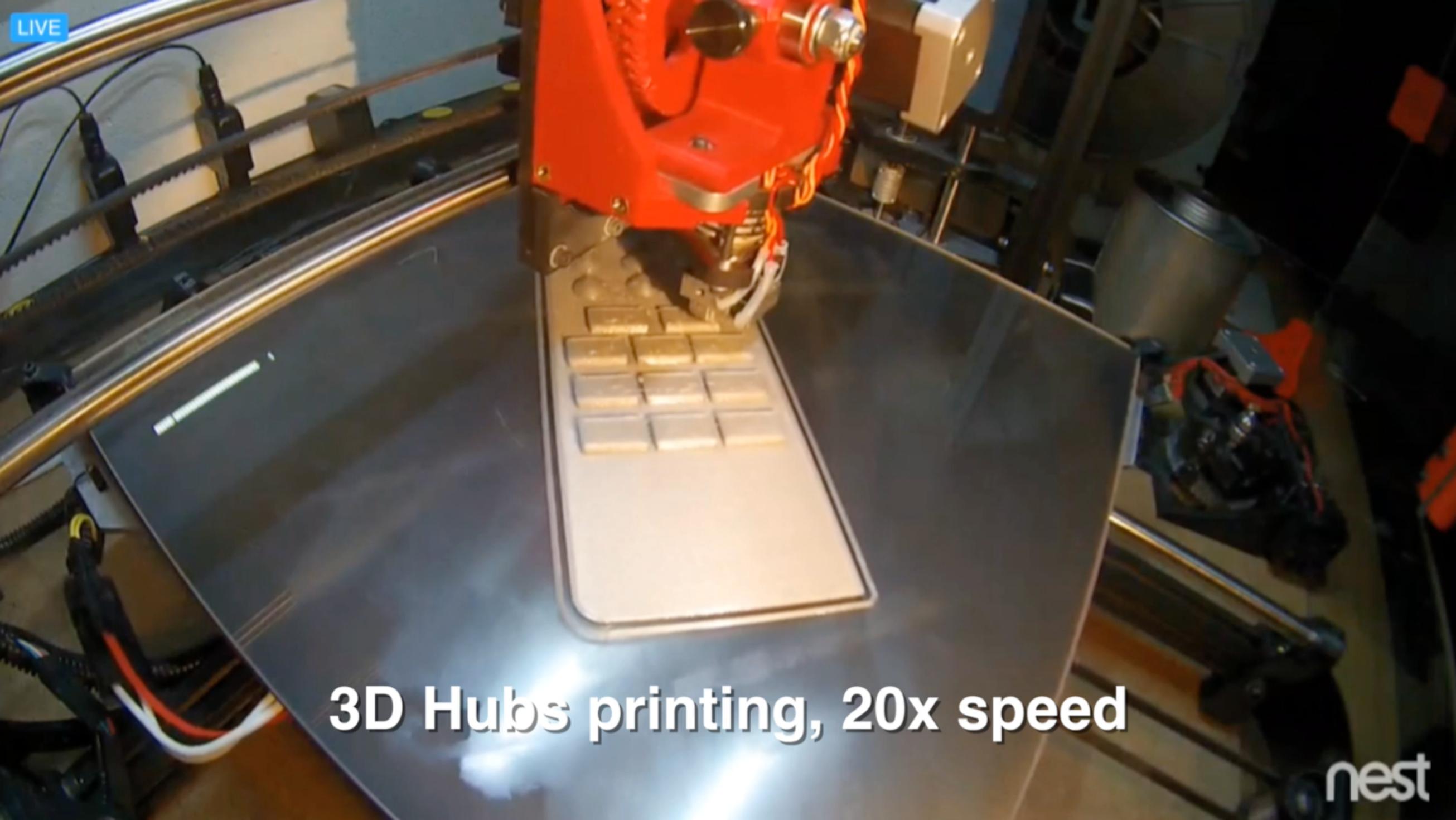




Choosing "Add Interface" in app..



LIVE

A close-up view of a 3D printer's extruder head, which is red and metallic, printing a small model of the Eiffel Tower. The printer is mounted on a black frame with visible tracks and cables. The print itself is a light-colored plastic, showing the characteristic lattice structure of the tower's legs.

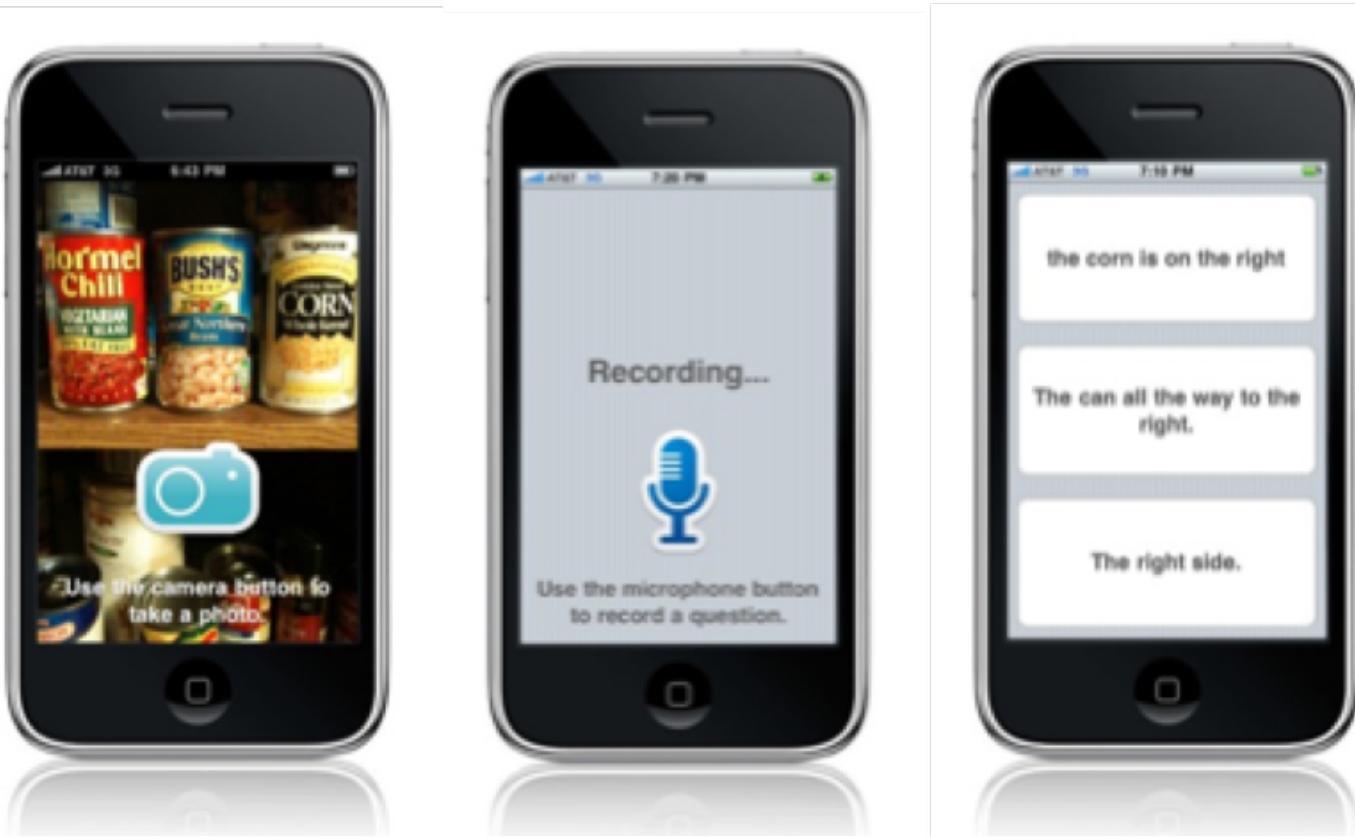
3D Hubs printing, 20x speed

nest



User aligning overlay with control panel..

What do blind people want to know?



People + Computer Vision



noeyesneeded.com



Human-
Computer
Interaction
Institute



People + Computer Vision



noeyesneeded.com



Supporting Blind Photography

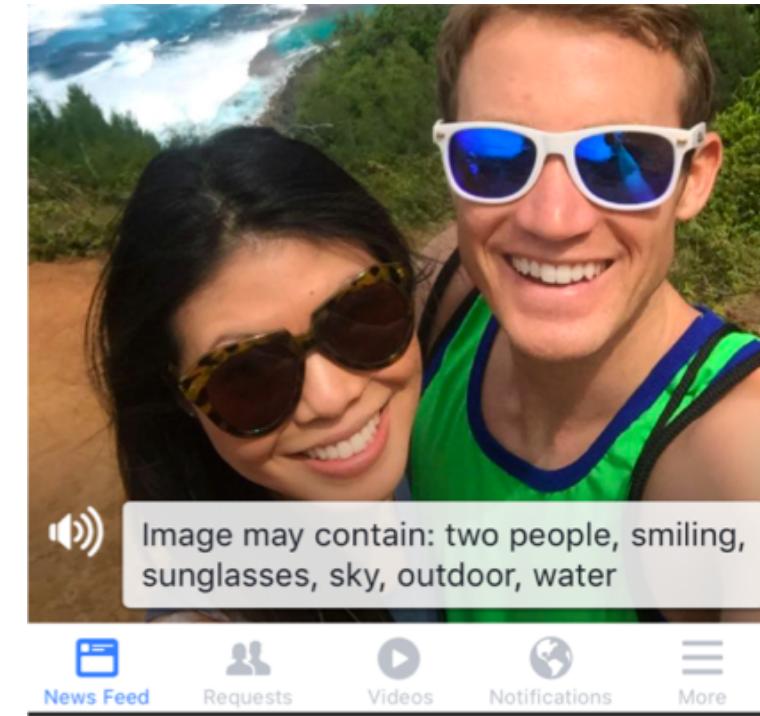
The image is a composite of two parts. The top part shows a person's hands holding a smartphone over a clear plastic container filled with various items, including a US dollar bill. The bottom part is a screenshot of a mobile application interface titled "Album". It shows a grid of four photos: "Pattern" (a colorful patterned cloth), "People" (a person's face), "Apple i Phone" (a red iPhone), and "Lenovo Computer Mouse" (a computer mouse). Below the photos are icons for Help, People, Object, Freestyle, and Album. A text overlay at the bottom of the screenshot reads "App providing feedback for aiming..".

Supporting Blind Photography. Jayant, Chandrika, Ji, Hanjie, White, Samuel, Bigham, Jeffrey P. ASSETS 2011.

People + Computer Vision



Target What People Want to Know



Wu, et al. Automatic Alt-text: Computer-generated Image Descriptions for Blind Users on a Social Network Service. CSCW 2017.



Human-
Computer
Interaction
Institute



People + Computer Vision

Help Users Understand Potential for Errors



MacLeod, Bennett, Morris, Cuttrell, *et al.* Understanding Blind People's Experiences with Computer-Generated Captions of Social Media Images. CHI 2017.



People + Computer Vision



noeyesneeded.com

- * **Amplify Users with AI Support** (blind photography)
- * **Understand User Needs** (solve real problems)
- * **Help Users Understand Potential for Errors**



Human-
Computer
Interaction
Institute



vizwiz.org/data

The screenshot shows a web browser window for the VizWiz Challenge website. The URL in the address bar is vizwiz.org/data/. The page title is "VizWiz Challenge". The user is identified as "Jeffrey". The main content area is titled "Answering Visual Questions from Blind People". It displays a grid of 12 visual questions, each with an image and a question-and-answer pair.

Image	Question	Answer
	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	

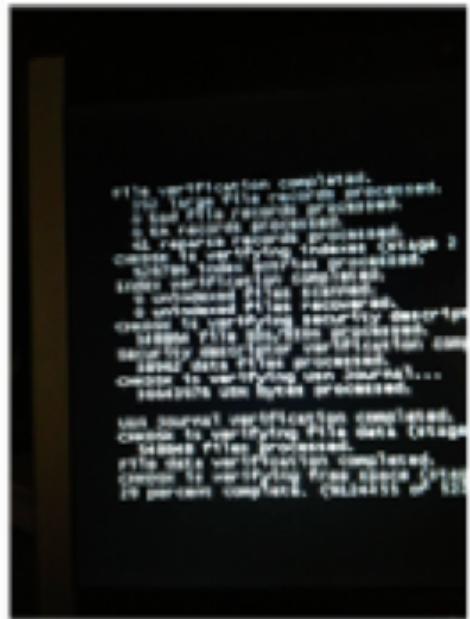
Jeffrey P. Bigham
@jeffbigham

Dataset	Which Images?	Who Asked?	How Asked?
DAQUAR [28]	NYU Depth V2 [34]	Automatically generated (templates)	_____
VQA v1.0: Abstract [7]	Abstract Scenes	Crowd workers (AMT)	Typed
VQA v1.0: Real [7]	MSCOCO [25]	Crowd workers (AMT)	Typed
Visual Madlibs [43]	MSCOCO [25]	Automatically generated (templates)	_____
FM-IQA [16]	MSCOCO [25]	Crowd workers (Baidu)	Typed
KB-VQA [40]	MSCOCO [25]	In-house participants	Typed
COCO-QA [32]	MSCOCO [25]	Automatically generated (captions)	_____
VQA v2.0: Real [17]	MSCOCO [25]	Crowd workers (AMT)	Typed
Visual7W [45]	MSCOCO [25]	Crowd workers (AMT)	Typed
CLEVR [19]	Synthetic Shapes	Automatically generated (templates)	_____
SHAPES [6]	Synthetic Shapes	Automatically generated (templates)	_____
Visual Genome [23]	MSCOCO [25] & YFCC100M [35]	Crowd workers (AMT)	Typed
FVQA [39]	MSCOCO [25] & ImageNet [14]	In-house participants	Typed
TDIUC [20]	MSCOCO [25] & YFCC100M [35]	Crowd workers (AMT), In-house participants, Automatically generated	Typed
Ours - VizWiz	Blind people use mobile phones to take a picture and ask question		Spoken

<https://www.cs.cmu.edu/~jbigham/pubs/pdfs/2018/vizwiz-dataset.pdf>

VizWiz Dataset – Answer the Question

Disk Scan



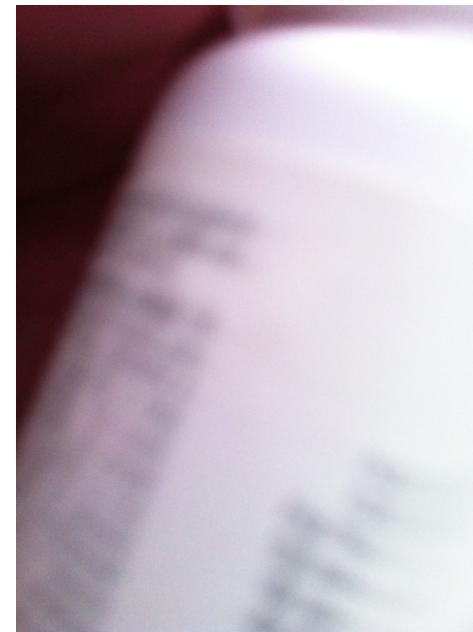
What is on my monitor?

Sweet Pepper



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

Unanswerable



What type of pills are these?

Unanswerable



What is this?

VizWiz Dataset – Answerable Question?

Answerable



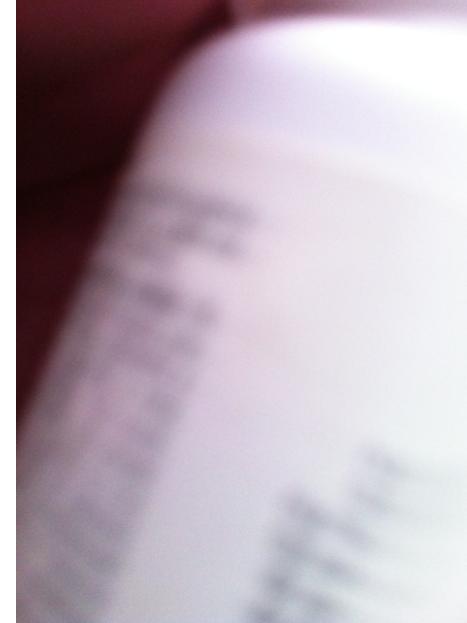
What is on my monitor?

Answerable



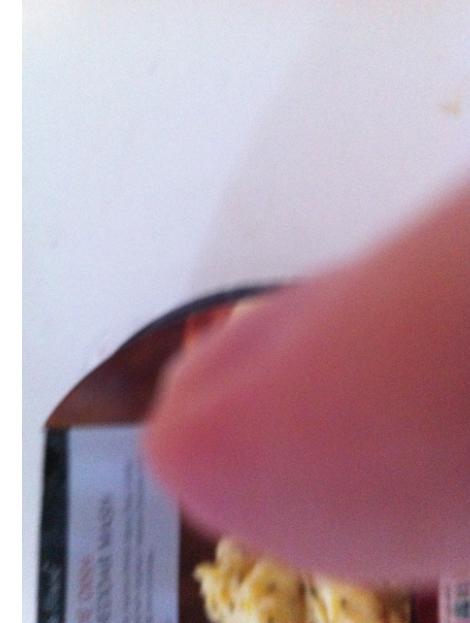
Hi there can you please tell me what flavor this is?

Unanswerable



What type of pills are these?

Unanswerable



What is this?

The End



VizWiz Dataset Team



Funded by:
Google, Microsoft, NSF, NIDILRR

Jeffrey P. Bigham
@jeffbigham

www.jeffreybigham.com



Human-
Computer
Interaction
Institute

