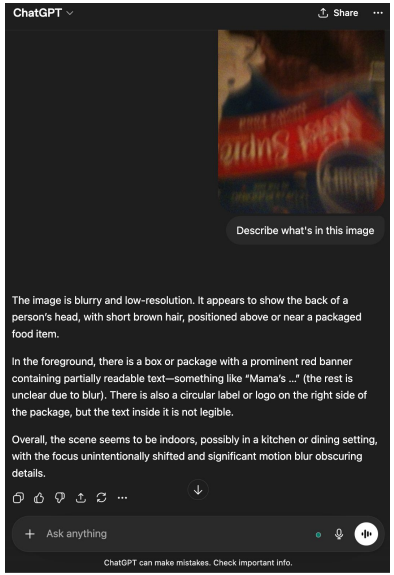
“It's trained by non-disabled people”: Evaluating How Image Quality Affects Product Captioning with Vision-Language Models

Kapil Garg, Xinru Tang, Jimin Heo, Dwayne R. Morgan, Darren Gergle,
Erik B. Sudderth, and Anne Marie Piper

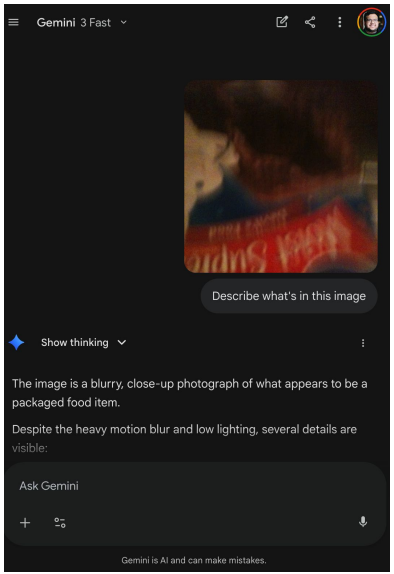
UC Irvine Donald Bren School of Information & Computer Sciences
Northwestern University
NSF U.S. National Science Foundation
CHI'26

VLMs Increasingly Power AI Tools for Visual Accessibility

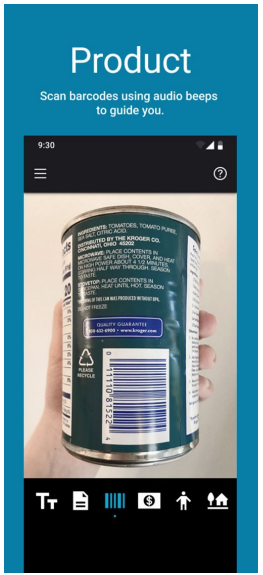
ChatGPT



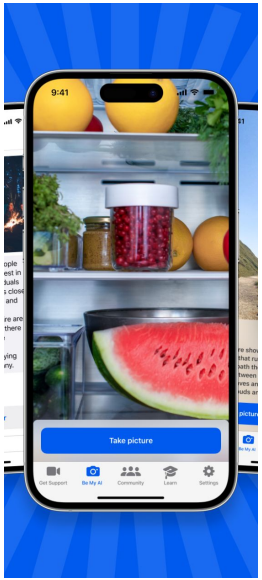
Gemini



SeeingAI



Be My AI



General Purpose VLM-Based Tools

Accessibility-Specific VLM-Based Tools

Pillsbury Moist Supreme Cake Mix, Devil's Food Cake Flavor

GPT 4.1

Box of **frozen Stouffer's Lasagna with Meat & Sauce, Party Size variety**. The packaging is primarily red and blue with white lettering.

Gemini 2.5 Flash

A rectangular box of **Purina Fancy Feast dry cat food**. The top of the box is a red banner with white lettering for the **brand name, which is partially cut off but shows "Fancy Feast."** The bottom of the box is dark blue with a small circular logo on the bottom right.

Research Gaps

How do blind and low-vision (BLV) people decide to use AI for product identification, and what challenges occur?

BLV participants only noticed half of the identification errors with a object recognition tool
[Hong and Kacorri, 2024]

Dealing with errors by progressively verifying information with other AI tools and people
[Tang et al., 2025; Adnin and Das, 2024; Alharbi et al., 2024; Gonzalez et al., 2024]

Gap: less about trade-offs or preferences for AI versus human assistance, and what information they are seeking

How robust are vision-language models (VLMs) to image quality issues (e.g., blur, framing, rotation) for product identification?

Image quality also seems to play a role, observed in recent interview studies
[Zhao et al., 2018; Alharbi et al., 2024]

Gap: systematic way to assess how sensitive state-of-the-art VLMs are to image quality

Our Paper's Contributions

1. Details BLV people's decision-making process for using AI tools to identify products and the associated challenges, based on a survey

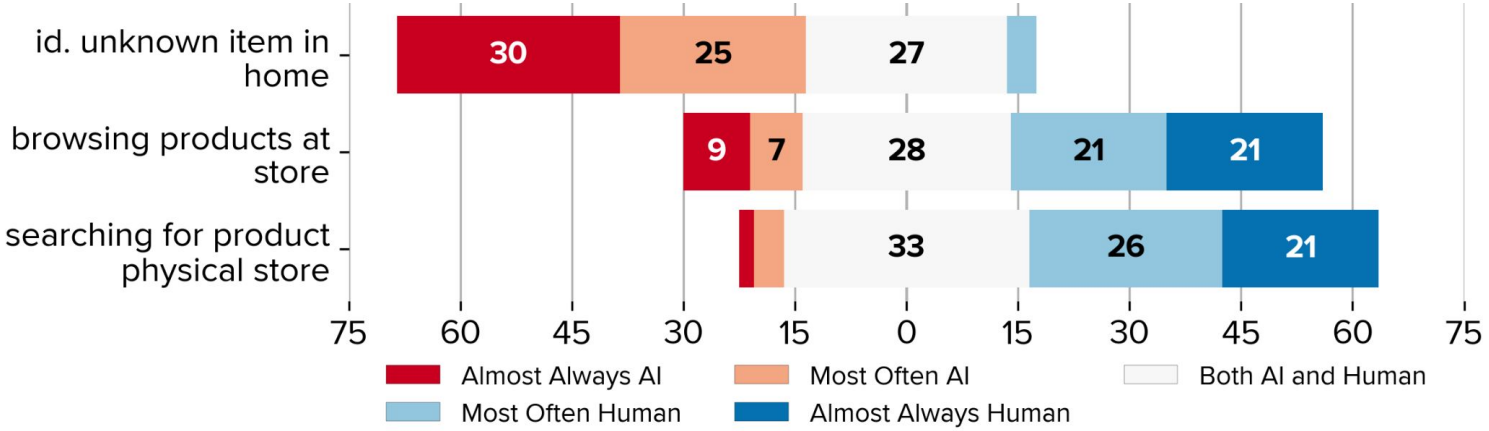
2. Develop an annotated dataset to benchmark the performance of four top-performing VLMs that enable existing AI tools

3. We discuss disability-centric methods for VLM development across data curation, evaluation, model training, and end-user AI tools

Study 1: How Do BLV People Decide to Use AI for Product Identification, and What Challenges Occur?

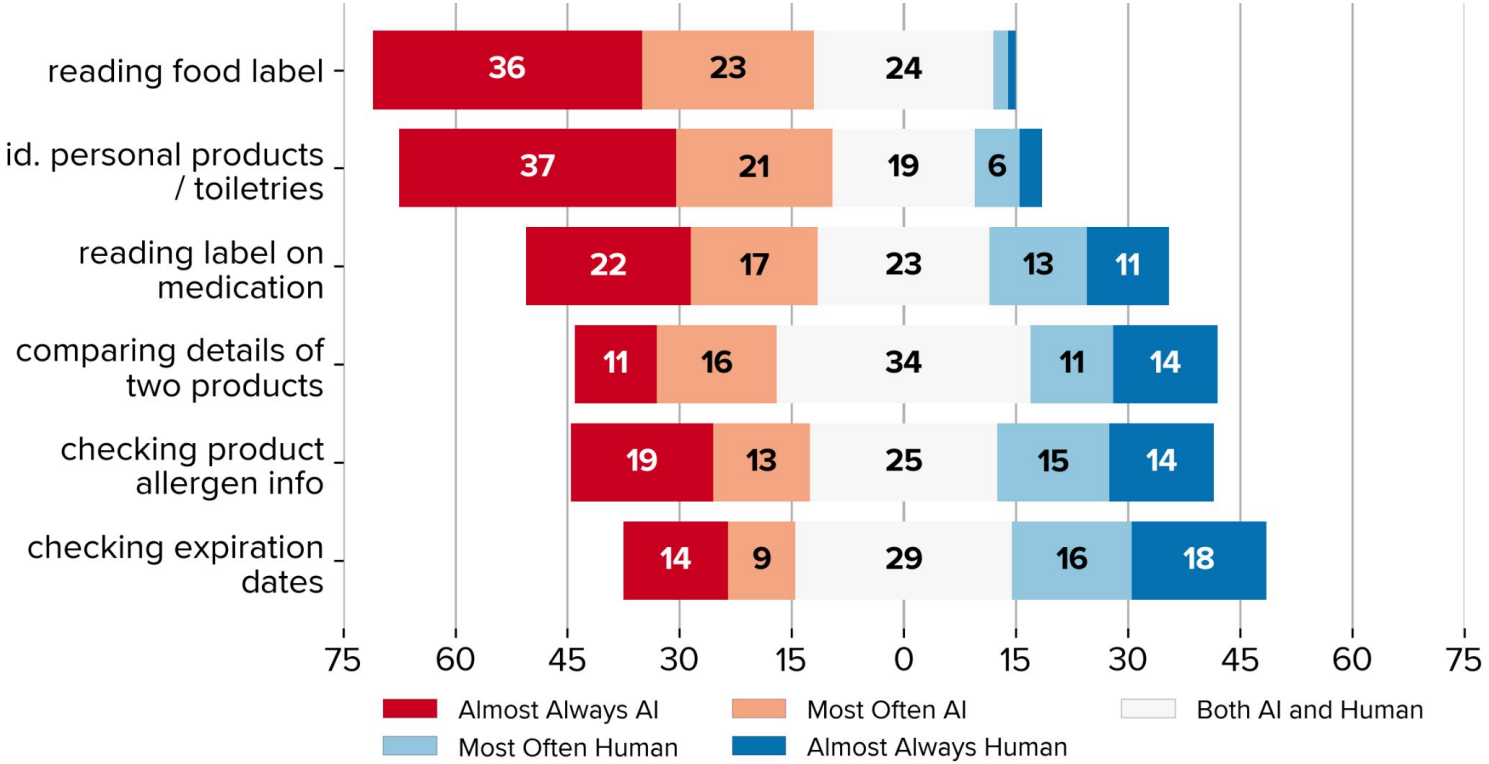
- Survey of 86 BLV people who used at least 1 AI tool for image captioning
  - Recruited via email lists, the National Federation of the Blind, and American Foundation for the Blind
- Likert scale and open-ended questions about their preferences, experiences, and challenges
- Analysis
  - Descriptive statistics
  - Excerpts of quotes

Study 1: AI in the Home, Human Assistance in Stores

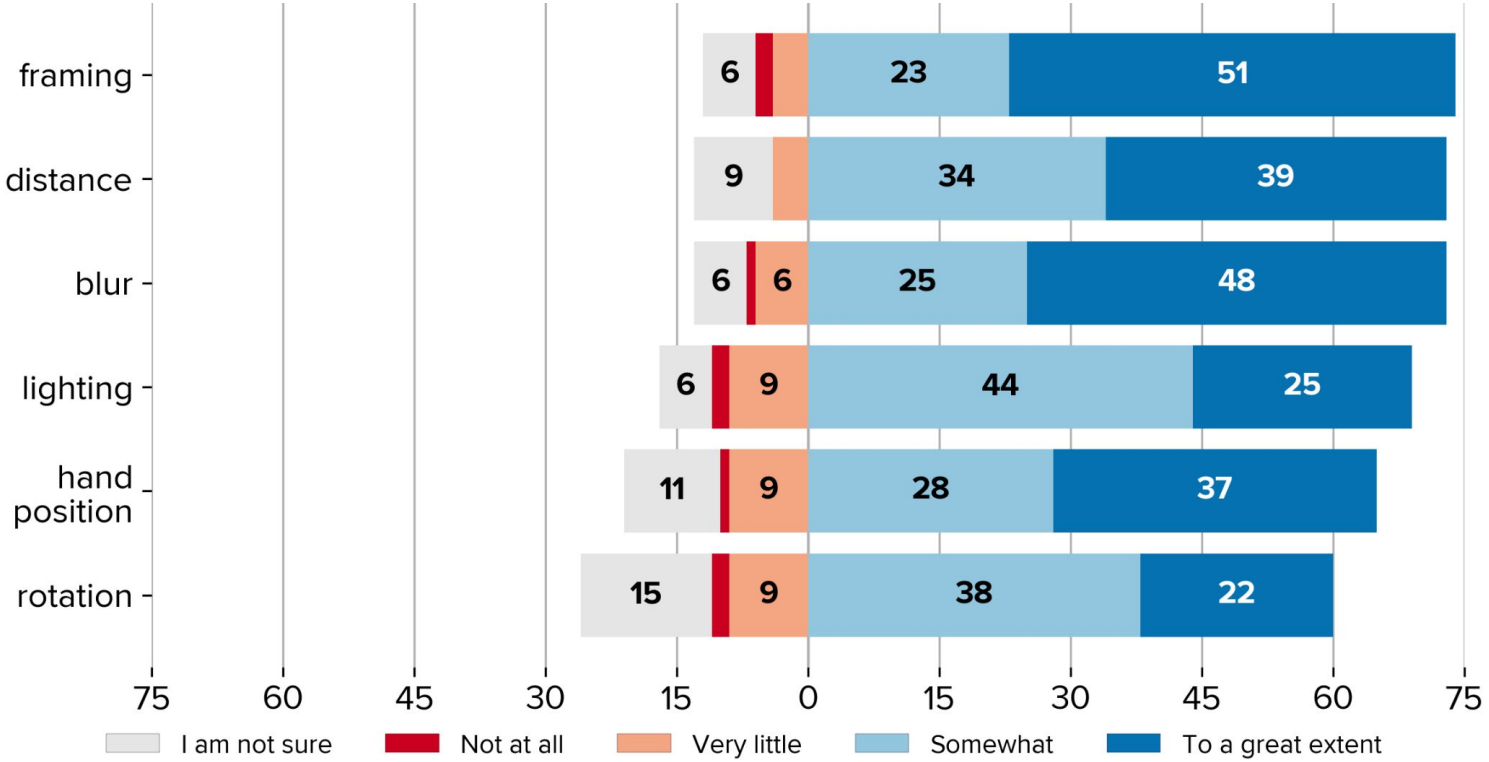


"...a human can often infer or already know where to go. Would take longer with just AI"

Study 1: AI for Higher-Level Information, Humans for Detailed Information



Study 1: Impact of Image Quality on Captions



Study 2: How Robust Are VLMs to Image Quality Issues for Product Identification?

Method

- Dataset (subset of VizWiz): 729 high-quality images with no issues, 1,130 low-quality images with at least 1 image quality issue
- Annotated with:
  - Product: generic term for product (e.g., cereal, soup, meal)
  - Brand: detectable brand information (e.g., Betty Crocker, Kraft)
  - Variety: details about type, flavor, or variety (e.g., peanut, low-sodium)
- Captions from: GPT-4.1 (OpenAI), Gemini 2.5 Flash (Google), Llama 3.2 90B (Meta), and Molmo 72B (AllenAI)
- Manually coded accuracy by four researchers, allowing for small misspellings and term variation, but strict on key annotation details

Study 2: How Robust Are VLMs to Image Quality Issues for Product Identification?

Analysis

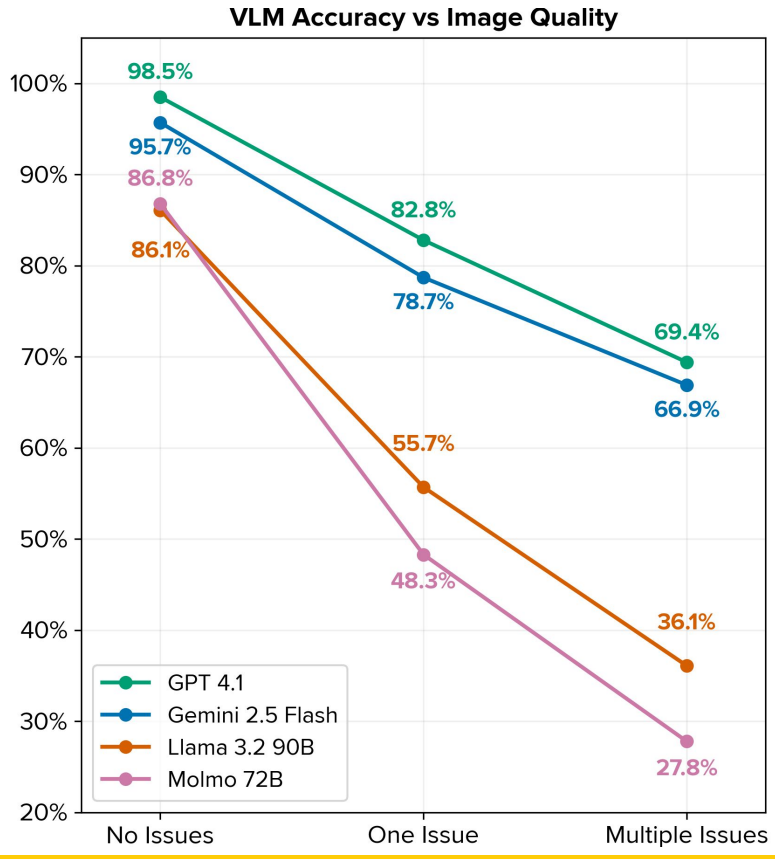
- Descriptive statistics to determine overall accuracy for different image quality issues
- Inferential analysis via logistic regression for general patterns and VLM-specific patterns
  - Dependent variable (binary): whether VLM correctly identified product
  - Independent variables (binary): blur, framing, and rotation + interaction effects
  - *Full paper include additional image attributes (product has text panels or curved labels) and other interaction effects (e.g., quality issue by VLM)*

Study 2: Accuracy Declines Sharply as Image Quality Worsens

High-Quality (729 Images)

Single-Issue (460 images)

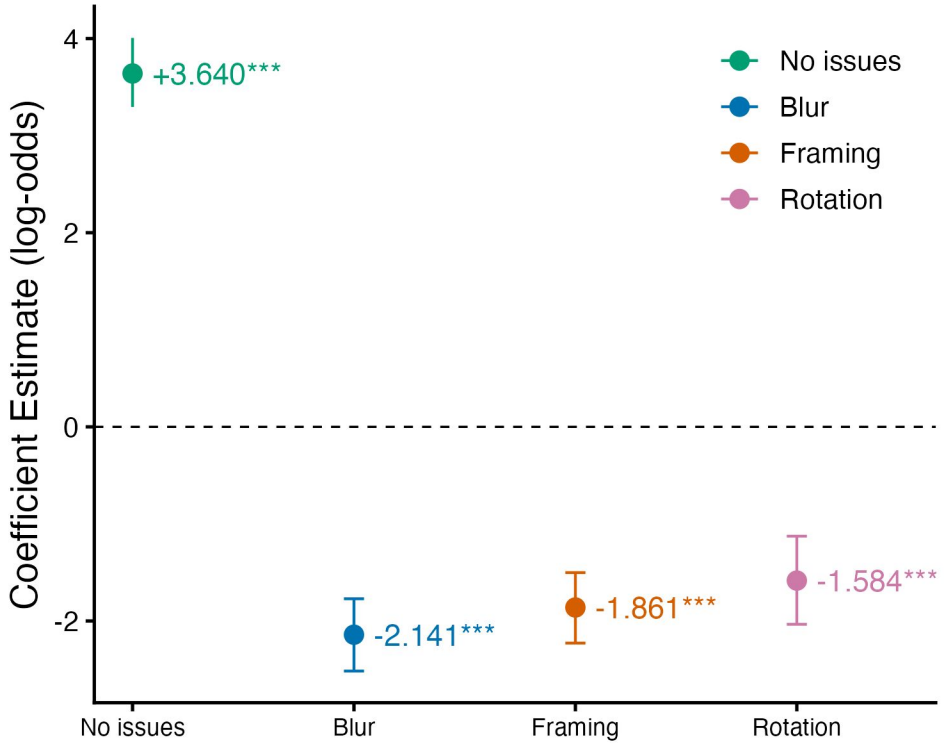
Multiple-Issues (670 images)



Study 2: Accuracy Declines Sharply as Image Quality Worsens

	High-Quality (729 Images)	Single-Issue (460 images)	Multiple-Issues (670 images)
			
GPT 4.1 Accuracy	98.5%	82.8%	69.4%
Average Incorrect	~3 in 200 Images	~17 in 100 Images	~3 in 10 Images

Study 2: Image Quality Issues Differentially Affect VLM Accuracy



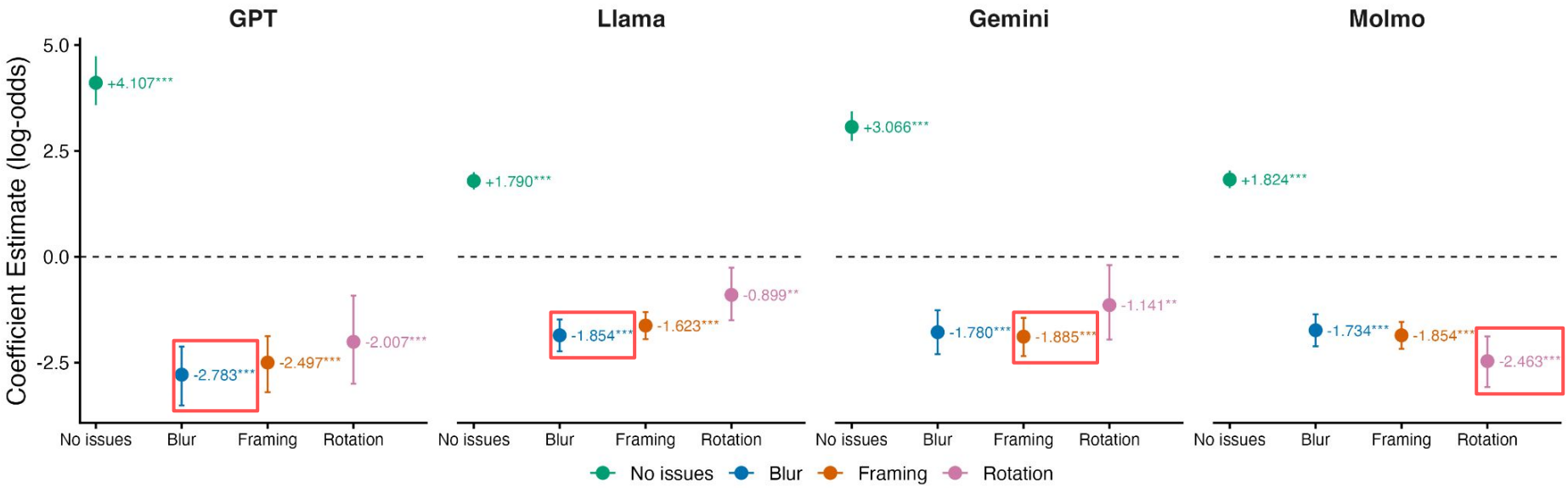
Co-Occurring Quality Issues

Blur and Rotation

Blur, Framing, and Rotation

Accuracy decreases, but not as much as each effect individually

Study 2: Each VLM is Susceptible to Different Quality Issues



To improve specific models, we could prioritize the quality issue they are most affected by.

Study 2: VLMs Can Struggle to Fill-In Information

Bottle of Dr. Pepper, Diet

GPT-4.1

Plastic **bottle** of **Dr Pepper** soda featuring a promotional label with the Marvel character Thor. The **bottle** is dark-colored with a white label displaying Thor holding his hammer.

Missing:

- variety: diet

Molmo 72B

Coca-Cola bottle. Brown plastic **bottle** with white label featuring Thor from Marvel. Thor shown holding his hammer with lightning. "Win powerful prizes" text visible. Partially visible through plastic bag.

Incorrect:

- brand: dr. pepper

Missing:

- variety: diet

Discussion: Disability-Centric Model Evaluation

Extends prior work on disability-centric model evaluation

[Gadiraju et al., 2023; Park et al., 2025; Kapur and Kreiss, 2024; Alharbi et al., 2024; Tang et al, 2025]

1. Datasets should treat image quality variability as a central design consideration, versus dataset where high-quality images are predominant (e.g., in ImageNet and MS COCO)

2. Significant human labor is necessary for high-quality dataset creation, and may need to move beyond crowdworkers or synthetic data generation

3. Better metrics that measure correctness, not similarity to reference
   [Kapur and Kreiss, 2024]

Discussion: Improving VLMs Across the Pipeline

1. **Data Curation:** Models need more post-training with low-quality images
   [quality agnostic learning, Yu et al., 2023]

2. **Training Strategies:** Training objectives that better capture important information to BLV people (e.g., Cap F1 for precision and recall on concepts)
   [Deitke et al., 2025]

3. **Inference time:** Improving caption quality during inference without additional burden (i.e., retaking a photo) through improved image input or abstention on incorrect information
   [Inpainting for cropped images, Agarwal et al., 2026]

“It's trained by non-disabled people”: Evaluating How Image Quality Affects Product Captioning with Vision-Language Models

Kapil Garg, Xinru Tang, Jimin Heo, Dwayne R. Morgan, Darren Gergle, Erik B. Sudderth, and Anne Marie Piper

UC Irvine Donald Bren School of Information & Computer Sciences

Northwestern University

U.S. National Science Foundation

CHI'26

Paper + Dataset

Study 1: how do BLV people decide to use AI for product identification, and what challenges occur?

Based on a survey of 86 BLV people who use frequently used AI tools (e.g., ChatGPT, SeeingAI, BeMyAI):

55% would most often or always use AI at home

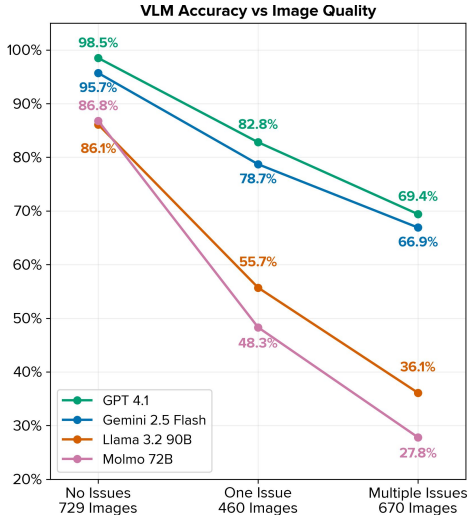
Preferred human assistance in stores for browsing (49%) or finding products (55%), because of reliability and speed

Preferred human assistance for important information (allergen info, 33%; expiration dates, 39%) versus AI for general identification (68%)

69% or more felt that image quality affected caption quality somewhat or to a great extent

Study 2: how robust are VLMs to image quality issues (e.g., blur, framing, rotation) for product identification?

On our dataset of 1,859 product images from BLV people, accuracy drops sharply as quality worsens:



Example:

Pillsbury Moist Supreme Cake Mix, Devil's Food Cake Flavor

GPT 4.1

Box of **frozen Stouffer's Lasagna with Meat & Sauce, Party Size variety**. [...]

Gemini 2.5 Flash

A rectangular box of **Purina Fancy Feast dry cat food** [...] with white lettering for the **brand name, which is partially cut off but shows “Fancy Feast.”** [...]