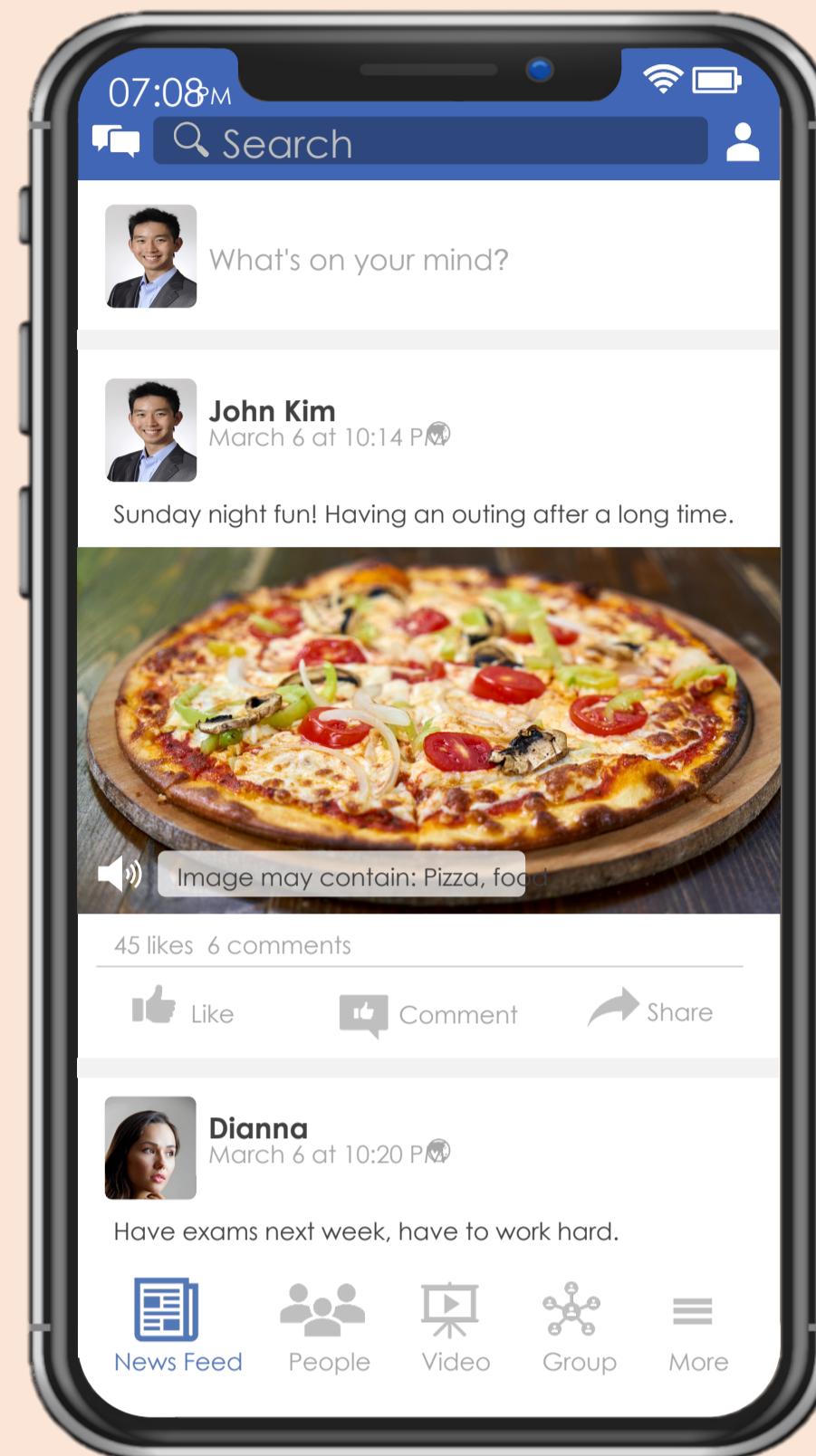


# Computer Vision and Conflicting Values: Describing People with Automated Alt Text

Margot Hanley, Solon Barocas, Karen Levy, Shiri Azenkot, and Helen Nissenbaum

## Background

- In 2016, Facebook introduced **automatic alt text**, a tool which employs computer vision to make images accessible to their blind and low vision users
- Facebook itself **acknowledged normative concerns around describing images of people**, especially in light of controversies involving people being labeled in offensive and demeaning ways by commercial computer vision systems.



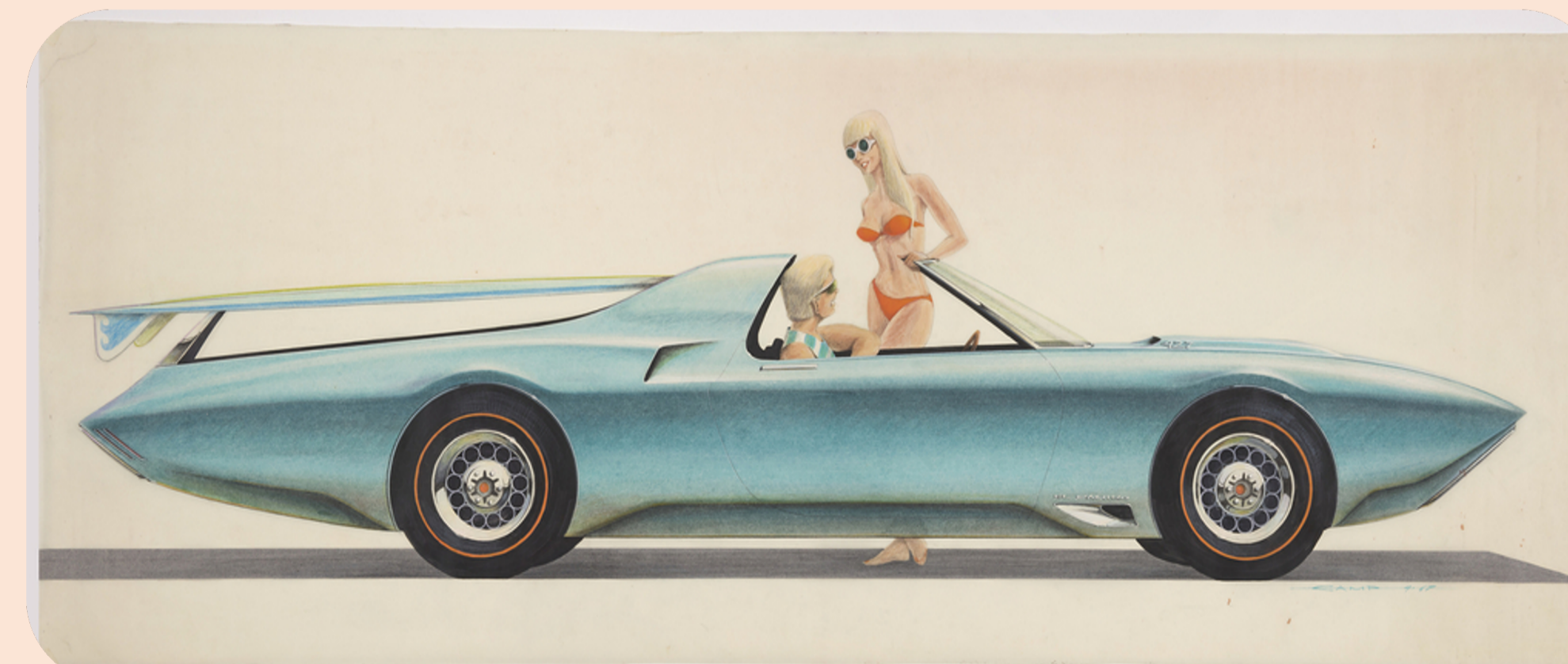
## Our research

- We investigate Facebook's **policy regarding the use of identity categories** in alt text, comparing it to the **policies and practices of museums**
- Review **all publicly available statements** from Facebook and the **formal guidelines** from museums

### Data sources:



## Policy In Practice



Alt text: "Pale blue convertible car with a surfboard on back; driver is a blonde man talking to a blonde woman in a red bikini."

AAT: "Image may contain: two people, smiling, sunglasses, sky, outdoor"



	Facebook	Cooper Hewitt	Museum Contemporary Art Chicago
<b>Race/Ethnicity/Skin Tone</b>	"By default, Facebook will only suggest a tag for a photo if it is 80 percent confident that it knows what it's looking at. But in sensitive cases — including ones involving race [...] it will require a much higher level of confidence before offering a suggestion. When it isn't confident, Facebook simply won't suggest a description. [38]"	"When describing the skin tone of a person use non-ethnic terms such as "light-skinned" or "dark-skinned" when clearly visible. Because of its widespread use, we recommend the emoji terms for skin tone as follows: 🏻 Light Skin Tone, 🏼 Medium-Light Skin Tone, 🏽 Medium Skin Tone, 🏾 Medium-Dark Skin Tone, 🏿 Dark Skin Tone. Also, where skin tone is obvious, one can use more specific terms such as black and white, or where known and verified, ethnic identity can be included with the visual information: Asian, African, Latinx/o/a (also see gender), etc."	"Demographic: race. This is in development, but for the time being identify clearly visible visual appearance when it is important to the understanding of the content. Default to "light-skinned" and "dark-skinned," when clearly visible. Where obvious and known, use more definite terms; e.g. black, Latino, Asian, etc."
<b>Gender</b>	"...we decided to leave out gender-related concepts such as woman/man, girl/boy, as gender identification is more complex, personal, and culture-dependent than what may appear visually on photos." [59]"	"No assumptions should be made about the gender of a person represented. Although, where gender is clearly performed and/or verifiable, it should be described. When unknown, a person should be described using "they, them" and "person" and their physicality expressed through the description of their features, which inadvertently tend to indicate masculine or feminine characteristics. The use of masculine and feminine are problematic and should be avoided unless necessary for describing the performance of gender."	"Demographics gender: Where necessary for understanding content gender may be described, but no assumptions should be made. Our default should be "person" except where gender is clearly evident and verifiable."
<b>Age</b>	"We ended up with a list of 97 concepts [...] including people (e.g., people count, smiling, child, baby)" [59]"	"Describe the age of represented people in an image using terminology such as baby, toddler, child, youth, teen, young person, adult, older person."	"Use terms that indicate age: baby, toddler, child, youth, teen, young adult, adult, older person."
<b>Disability</b>		"Not only [...] prominent features or physical stature, but also physical disabilities."	
<b>Physical Features</b>	"The current list of concepts covers a wide range of things that can appear in photos, such as people's appearance (e.g., baby, eyeglasses, beard, smiling, jewelry)" [13]"	"When particular features are immediately noticeable, or mutually agreed upon salient features of a known person are visually present, they should be described."	"need to create reference list for [...] hair color"
<b>Identity</b>	"And since people who use Facebook often share photos of friends and family, our AAT descriptions used facial recognition models that identified people (as long as those people gave explicit opt-in consent)." [1]"	"When describing an image of a recognizable person, identify them by name, but also describe their physical attributes. If an individual is not a public figure, and the context does not imply the importance of who is represented, it may not be appropriate to identify the individual."	"Feel free to identify clearly recognizable figures, e.g. Jesus, Bozo the Clown, Madonna, Anne Kaplan, and Sammy Davis Jr. Remember not to repeat caption information."

## Sources of Uncertainty

- Technical:** Uncertainties related to technical accuracy
- Ontological/Epistemological:** Uncertainties related to the ontology and epistemology of the categories
- Saliency:** Uncertainties related to social context and salience

## Strategies

- Directly Observable Features**
  - Only describes **directly observable physical features**
  - Makes no inferences** about identity categories
  - Defers judgement** to the person with visual impairment
- Facial Recognition**
  - Includes the **name of the image subject**
  - Allows person with visual impairment to **bring own knowledge to bear**
  - May **justify providing less or no description** of that person's appearance