

When to Add Human Narration to Photo-Sharing Social Media

Lawrence Kim*, Abena Boadi-Agyemang*, Alexa Siu, John Tang



Stanford
University



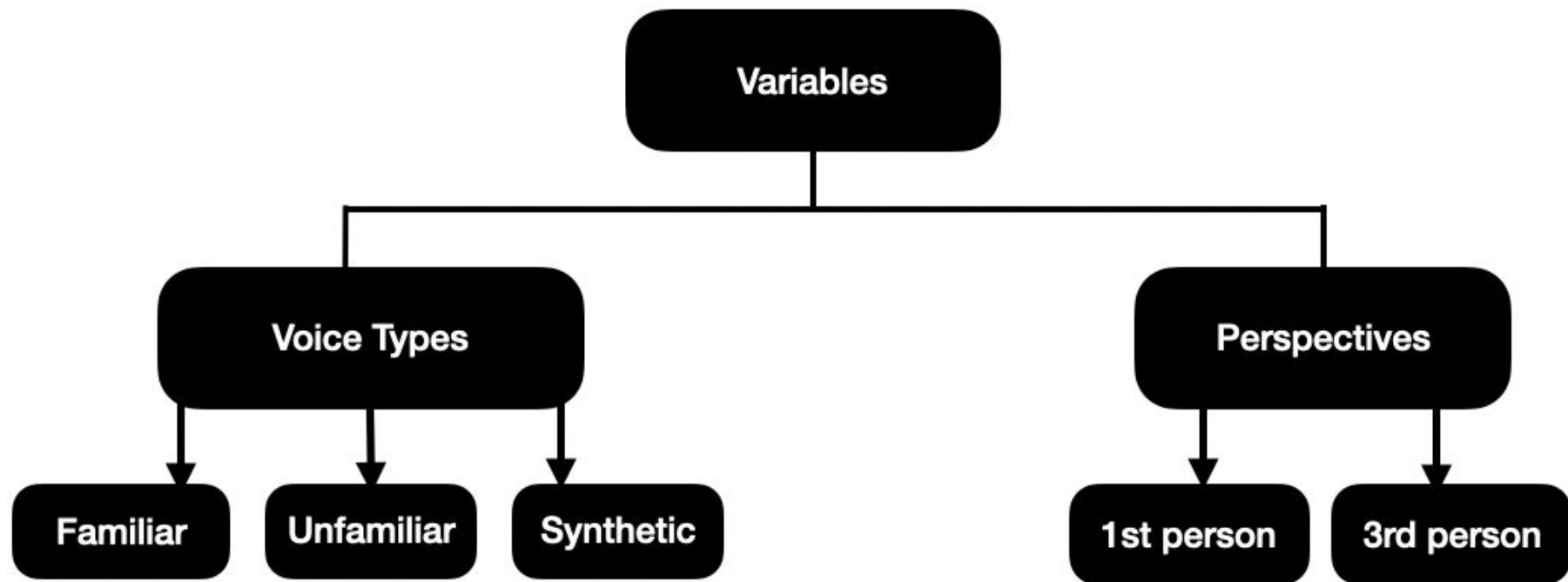
Microsoft

Introduction



- Photo-based social media is inaccessible.
- **Automatic alt text is vague and imprecise.**
- **Human narration offers rich description.**

Method



Method: Example



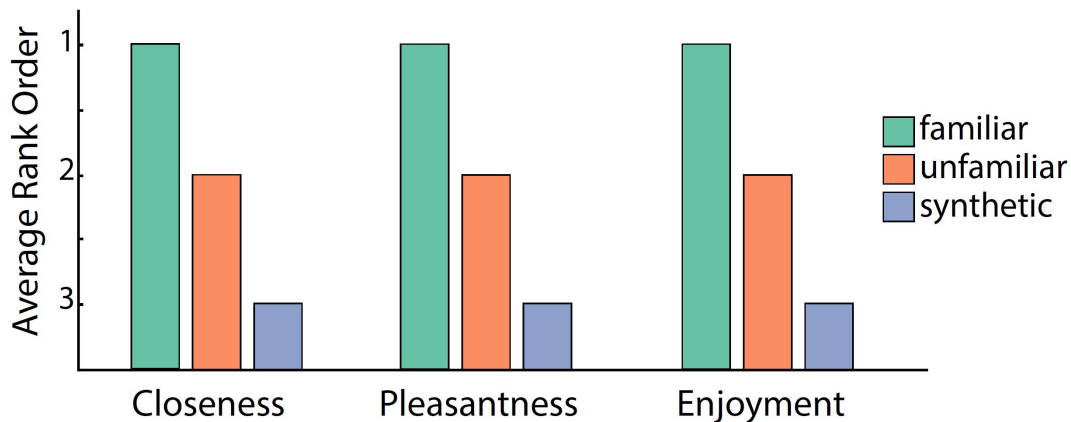
**1st person perspective:
“I’M LOOKING”**

VS



**3rd person perspective:
“SHE’S LOOKING”**

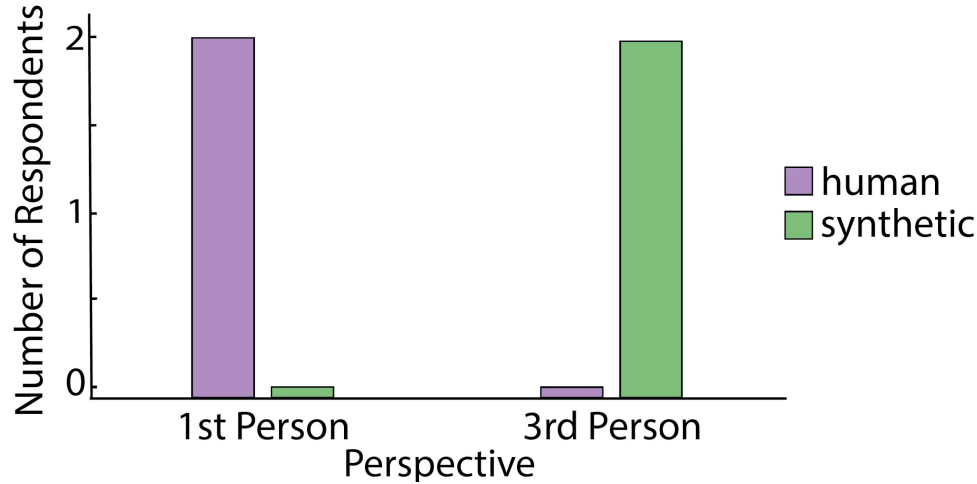
Prelim. Results - Voice Type



“... easier to make the connection with the content of the post when the narrating voice was someone I knew”

*“I always use [a] screen reader so I’m used to it ...
I always prefer a human voice over the synthesized voice.”*

Prelim. Results - Voice Type x Perspective



- **Human voice** is appropriate for content in **first person perspective** because it helps *establish a connection* between viewer and content creator.
- **Synthetic voice** is appropriate for content in **third person perspective** as it helps *create unbiased mental images*.

When to Add Human Narration to Photo-Sharing Social Media

Takeaways:

- In general, people **prefer familiar human voice** over unfamiliar or synthetic voices.
- **Context matters:**
 - Users prefer **human voice** for content described in **first person perspective**, but prefer **synthesized voice** for content in **third person perspective**.