

WHO are we empathizing with?

Who is the person we want to understand?
What is the situation they are in?
What is their role in the situation?

People with
Visual
Impairments

Content
Creators

General
Audience

Social-
Media Users

Multilingual
Users



What do they HEAR?

What are they hearing others say?
What are they hearing from friends?
What are they hearing from colleagues?
What are they hearing second-hand?

Direct
Auditory
Experience

User
Feedback,
Comments,
Opinions

Personal
anecdotes
from friends

Trends and
advancements
in image
caption
technology

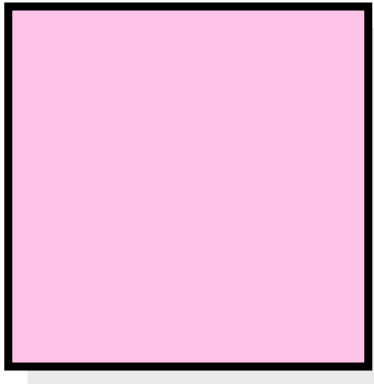
Impact on
their
productivity

GOAL

What do they THINK and FEEL?

PAINS

What are their fears,
frustrations, and anxieties?



Quality and
diversity of
text

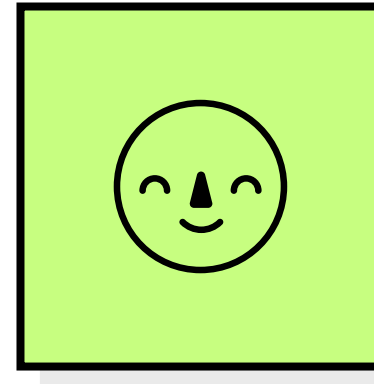
Tasks are so
synthetic

Limited in
size

Biased
towards
certain
domains

Biased
towards
genres

Biased
towards
styles



GAINS

What are their wants,
needs, hopes, and dreams?

Captions
provide on
screen

Translating an
input image
into a textual
description

Increases the
accessibility
of your
content

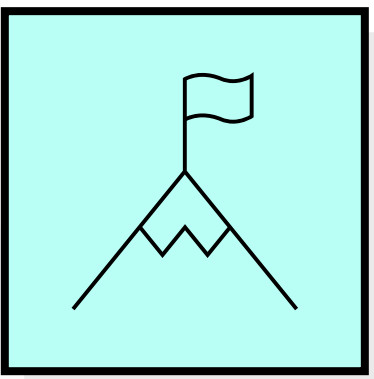
Caption
provides
readable
versions of
videos

Audio tracks to
support
viewers who
are deaf or
hard of hearing

improve
assistive
technology

Edit or improve
existing
captions to
enhance them

What do they need to DO?
What do they need to do differently?
What job(s) do they want or need to get done?
What decision(s) do they need to make?
How will we know they were successful?



Learn new
technique for
optimizing

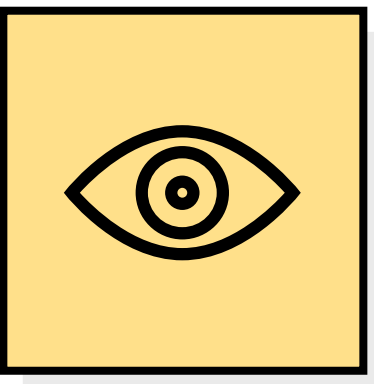
Decision
related to
most suitable
Image

Increased
User
Engagement

User
reviews

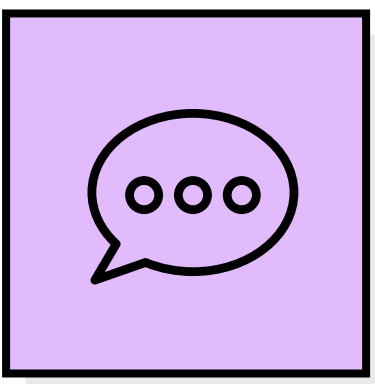
Competing
products and
platforms

Pricing
Model and
Subscription



What do they SEE?

What do they see in the marketplace?
What do they see in their immediate environment?
What do they see others saying and doing?
What are they watching and reading?



What do they SAY?

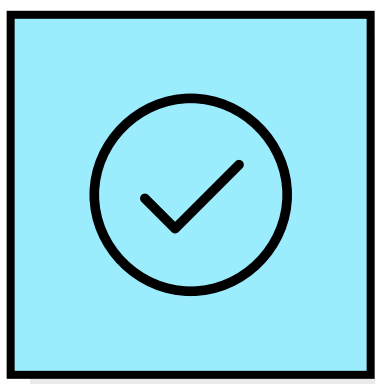
What have we heard them say?
What can we imagine them saying?

User
Feedback
Reviews

Positive
comments
Negative
feedback

Anticipating
their
demands

Imaging
User
Satisfaction



What do they DO?

What do they do today?
What behavior have we observed?
What can we imagine them doing?

Assist with
content
creation

Specific
actions when
adding to
images

Sharing
images with
captions on
various
platforms

Imagining users
using image
captions in new
and creative
ways as a tech

The specific
actions they
take when
reading captions
to the image

Analyzing user
engagement
with
generated
captioning