

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?

Generates textual
descriptions or
captions for
images
automatically

The deaf and hard
of hearing, those
with cognitive and
learning behaviors
found to improve
understanding



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?

People that speak
English as a
second language
and have been
thought to improve
literacy rates

Who cannot make
a conceptual
understanding of
the image they
can understand
by caption

Large text can
be changed
into a single
sentence

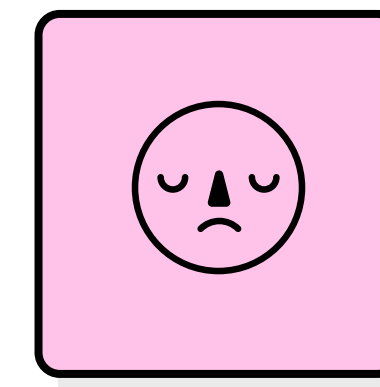
Convey
additional
information
that is not
present in the
image itself

GOAL

What do they THINK and FEEL?

PAINS

What are their fears,
frustrations, and anxieties?



One of the thing for
image captioning
and retrieval is the
quality and diversity
of the data used to
train and evaluate
the models

Most of the exiting
datasets for these
tasks are either
synthetic, Limited in
size and opposite
to genREs ,styles

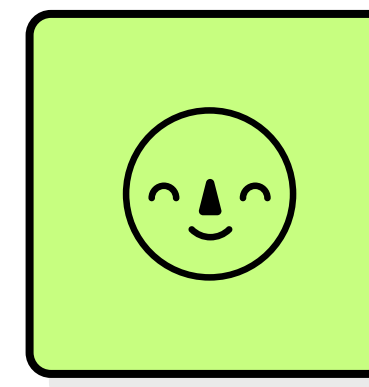
What other thoughts and feelings might influence their behavior?

Increased presence
of on screen text
which can make it
more difficult to take
the message of your
footage

Suppose we don't
have a input set
we can't context
the caption of
image

GAINS

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



Computer will
be aware of
the
approaching
objects

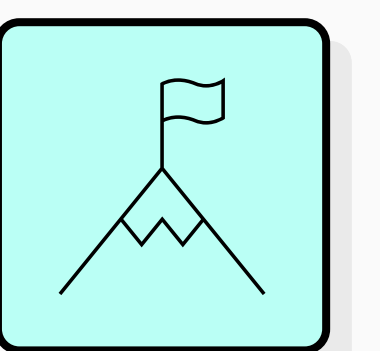
It will not spread
light in regions
occupied by an
approaching
object

In the process of
generating
caption the data
that has input
shape like a 2D
Matrix

Write the
information clearly
so the viewer does
not confused the
intent of the
caption and image

What do they need to DO?

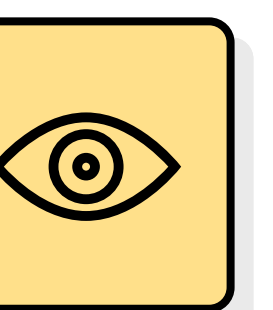
What do they need to do differently?
What job(s) do they want or need to get done?
What decision(s) do they need to make?
How will we know they were successful?



By the CNN and
LSTM model and
build a working
model of image
caption generator
by input data
set's

Image
descriptions read
out visually
impaired to get
better sense of
surroundings

Image caption
should never
be the same
as its
alternative text



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?



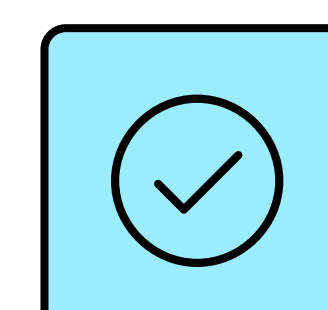
What do they SAY?

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

Clearly
identifies the
subject of
the picture

Image captions
should be
succint and
informative

Image captioning
is the process of
text generation by
recognizing
image



What do they DO?

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