

Strategy: The People Behind the Data

Fundamentals of the strategy

What is the goal of this strategy?

- To **humanize** data using **creativity** and **imagination**
- To apply **critical thinking** in order to **identify biases** in methods / data / interpretation.

When to use it?

When using /
working with
data

Why use it?

Humans → **Data** ← Humans
↓
Humans

IDEO's Design Principles

1. Data is **not truth**
2. **Don't presume** the desirability of AI
3. **Respect** privacy and the collective good
4. Unintended consequences of AI are **opportunities** for design

The 4 stages procedure

Stage 01: Search for the dataset documentation, source materials, source data.

"The more source material is, the better it is"

Try to find any information on:

- The procedure, methods, tools (e.g.: questionnaire) , etc. that were used to collect the data
- The data as it was recorded (source)
- The processing that was performed (e.g. how variables were computed, etc.)

Stage 02: Select a few inspiring questions/variables/columns from the dataset or its documentation.

"Get inspired"

Select a few questions/variables/columns that:

- Can help you imagine characteristics of people represented in the data
- Can lead to odd, extreme or inconsistent values in the dataset

Stage 03: Select a few rows from the dataset, read the data, imagine the people behind the data, their profile and their stories.

"Use your creativity and imagination"

Option 1 - Select rows at random

OR

Option 2 - Select rows based on specific characteristics (e.g. gender, race, score...)

For each row, write the story of the person represented by the data:

- What could explain the value they get on this attribute, or the score they obtain on this scale?
- What is their character?
- What is their past?
- How about their family?

Stage 04: Write down your conclusions in terms of ethical impacts/risks.

"Mix empathy with critical and analytical thinking"

Answer these questions:

- What have you learned about the data based on your exploration?
- Which potential harmful impacts could using this data generate?
- What would be your next steps: would you use these data? What other possibilities would you have?