Social Science is Necessary for Operationalizing Socially Responsible Foundation Models
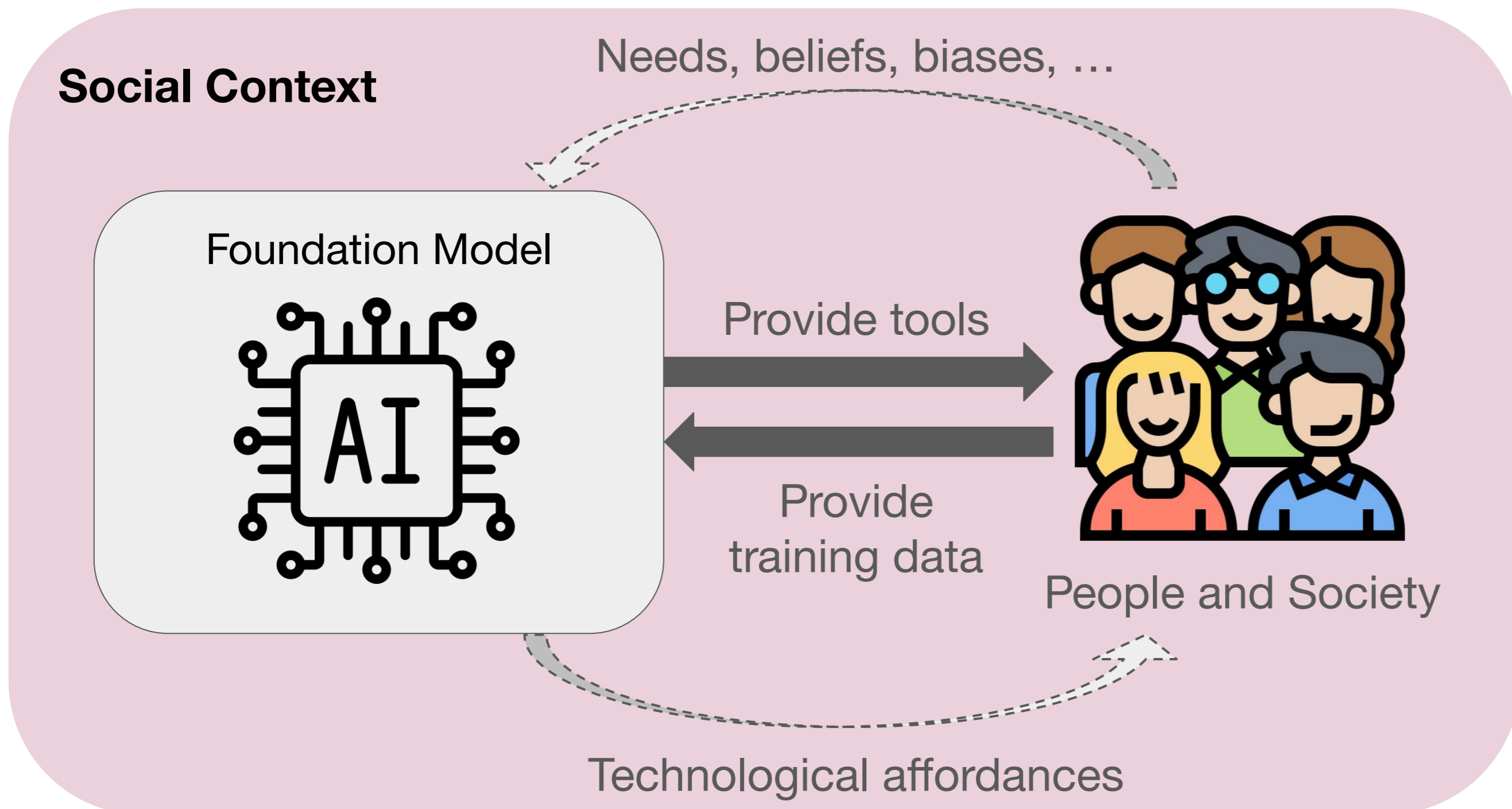
Adam Davies[1], Elisa Nguyen[2], Michael Simeone[3], Erik Johnston[3], Martin Gubri[4]

[1] University of Illinois Urbana-Champaign, [2] University of Tübingen,
[3] Arizona State University, [4] Parameter Lab

ICLR

Paper Link

Social Context

Needs, beliefs, biases, …

Foundation Model

AI

Provide tools

Provide training data
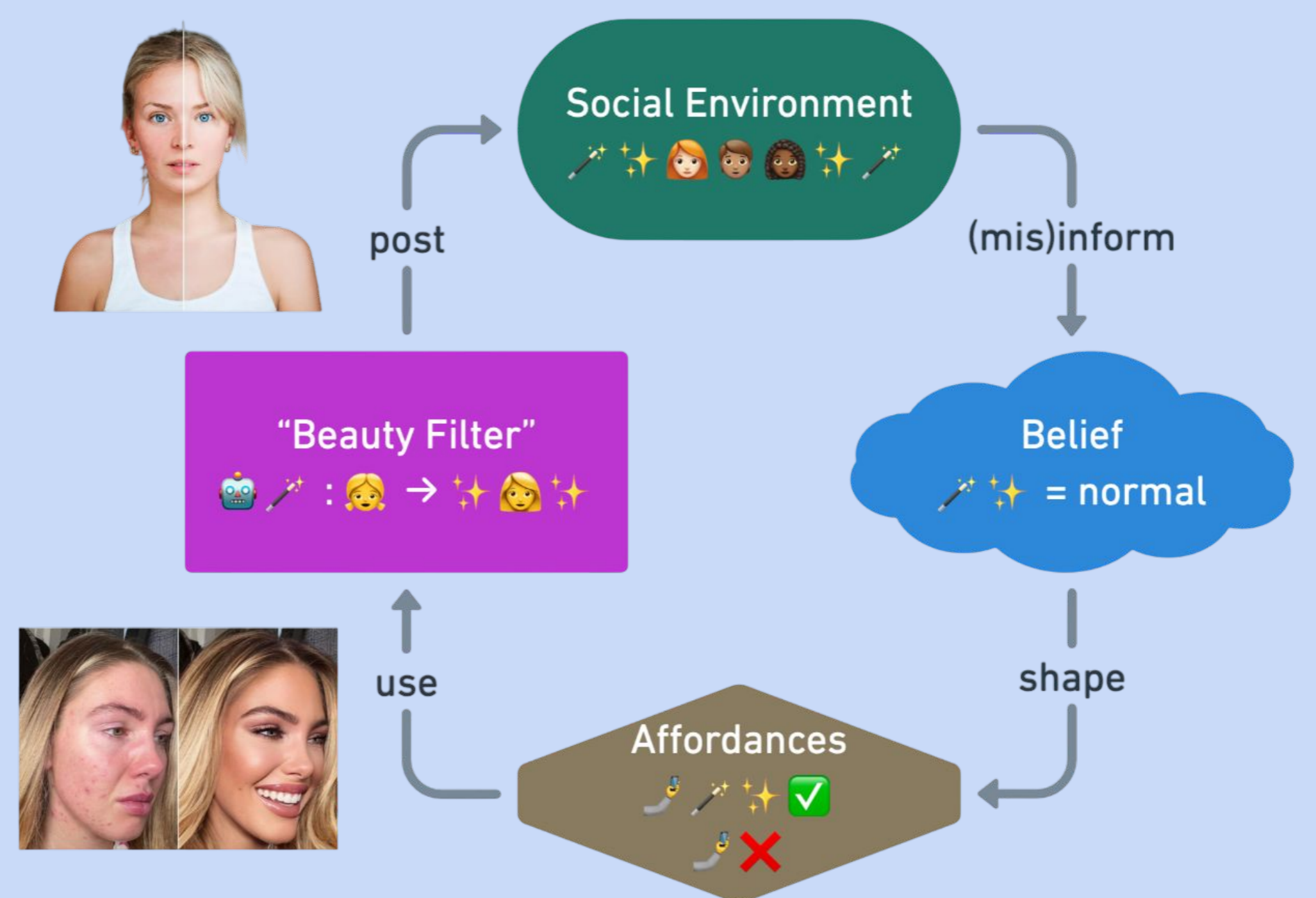
People and Society

Technological affordances

Foundation models are (embedded in) *sociotechnical systems*

- Enable, encourage, constrain actions
- Reciprocally shape social environment

Consider *beauty filters*:

- FM-enabled "filters" modify selfies to align with socially-constructed notion of "beautiful"
- "Beautified" selfies proliferate → must use filters to look "normal"
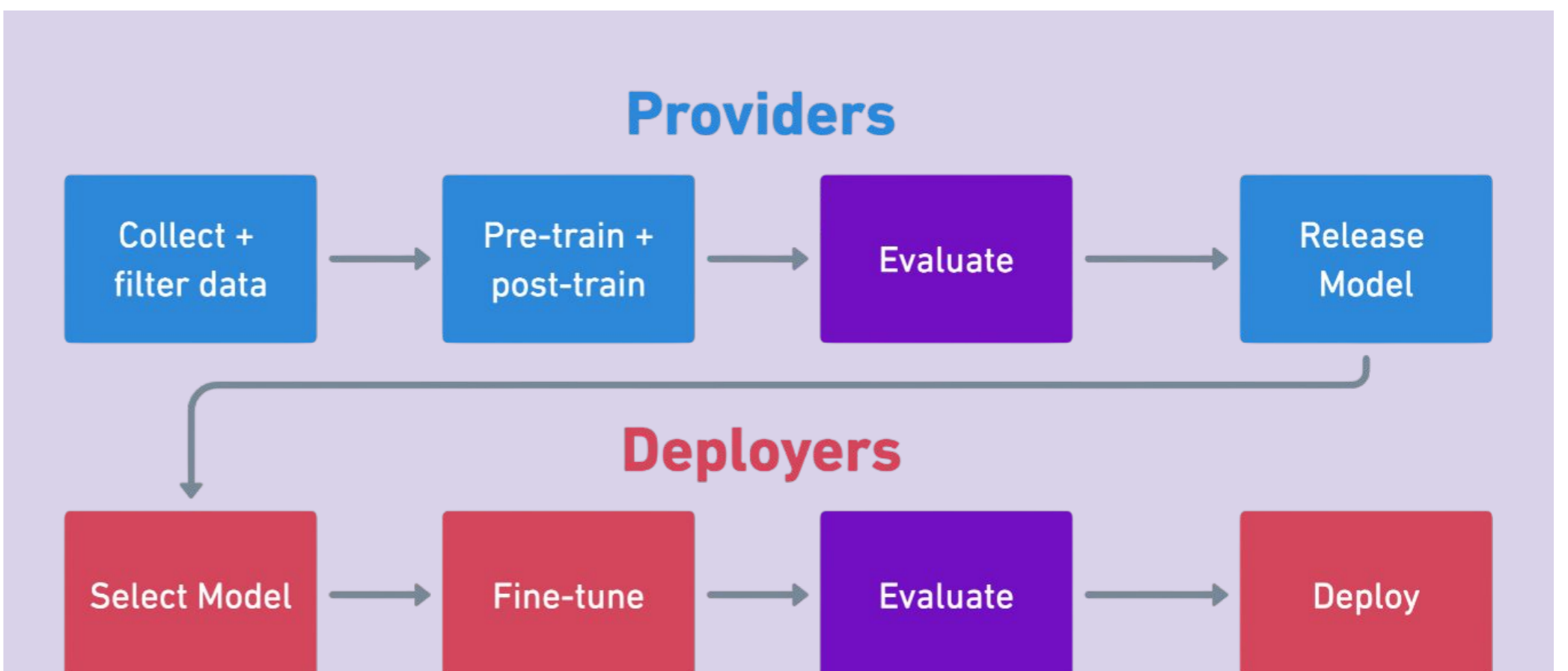- Reinforces unrealistic beauty standards, fatphobia, eating disorders, etc.

Social Environment

post

(mis)inform

"Beauty Filter"

Belief
= normal

use

shape

Affordances

How to *understand, anticipate,* and *mitigate social harms*?

➢ Need <u>AI + social science collabs</u> to…

1. **Study** sociotechnical systems
2. **Design** technical interventions
3. **Anticipate** social impacts

Incentives shape (collective) action:

- For *tech firms,* responsibility yields <u>financial stability</u> + <u>long-term gains</u>
- For *researchers,* need greater consideration of <u>interdisciplinary work</u> for <u>career advancement</u>

Providers

Collect + filter data → Pre-train + post-train → Evaluate → Release Model

Deployers

Select Model → Fine-tune → Evaluate → Deploy

Duty of care:

- Model **providers** responsible for <u>proactively considering sociotechnical systems</u> in *{data collection, training, evaluation}*
- Model **deployers** responsible for <u>addressing application-specific social impacts</u> in *{model selection, fine-tuning, deployment}*