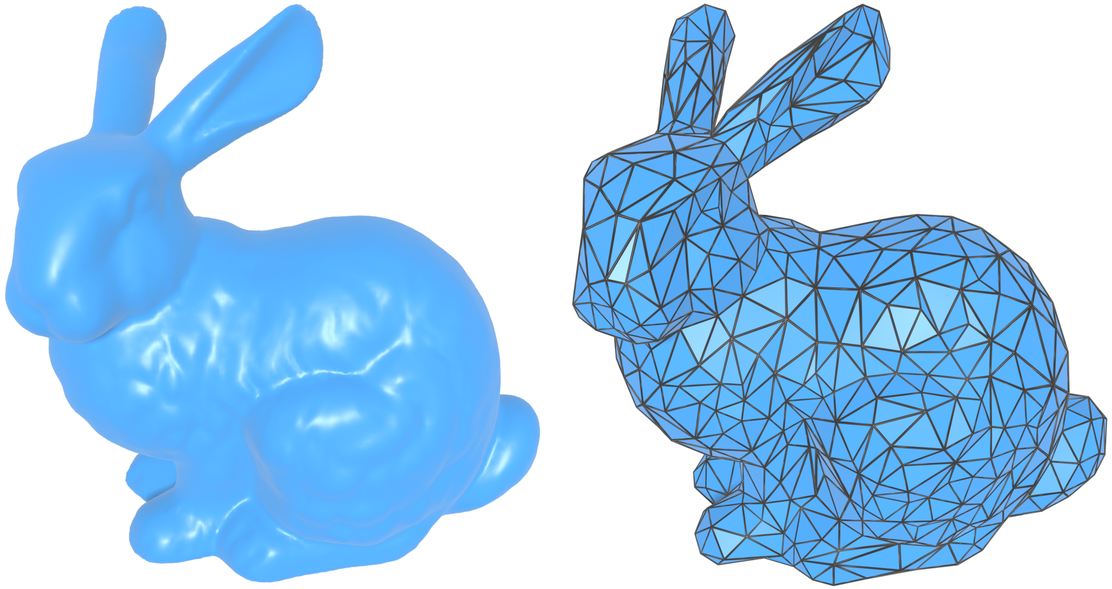
The Human in the *Infinite Loop*

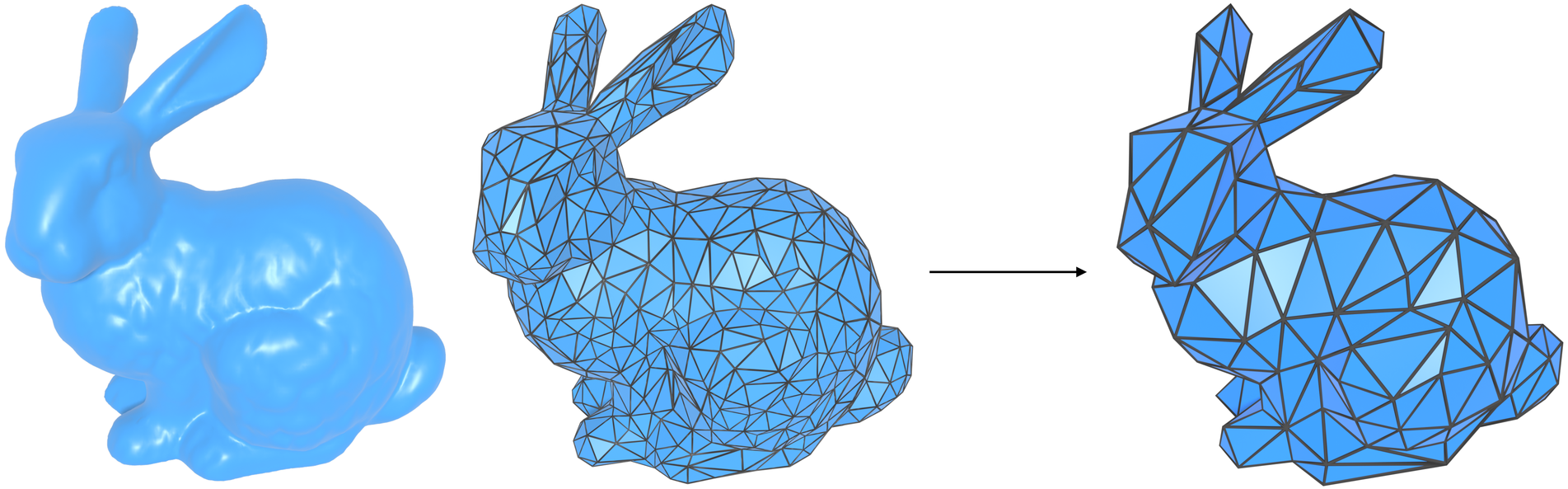
A Case Study on Revealing and Explaining Human-AI Interaction Loop Failures

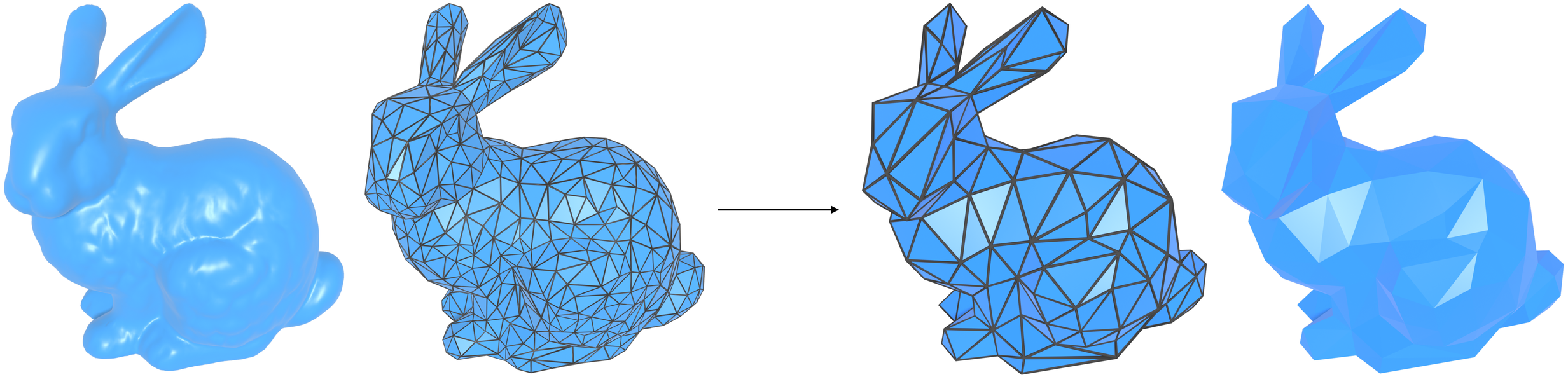
Changkun Ou, Daniel Buschek, Sven Mayer, Andreas Butz

Mensch und Computer 2022
SE04: Artificial Intelligence

September 6, 2022
Darmstadt, Germany





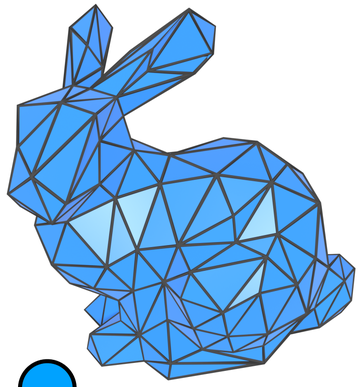


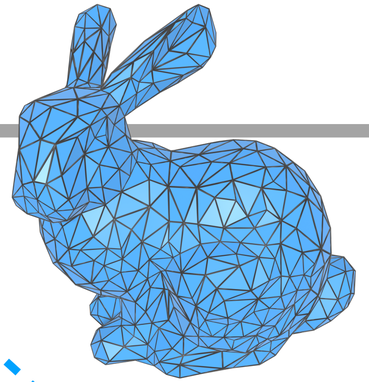
Mesh Space

Poor Visual Quality

Perfect Visual Quality

High Rendering Cost





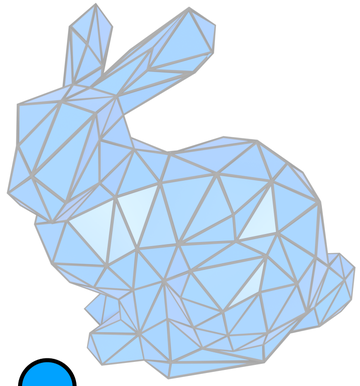
Mesh Space

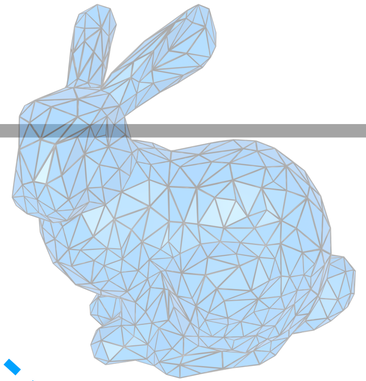
Poor Visual Quality

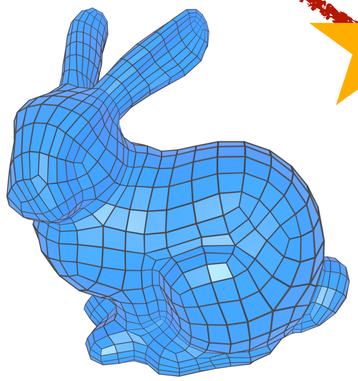
Perfect Visual Quality

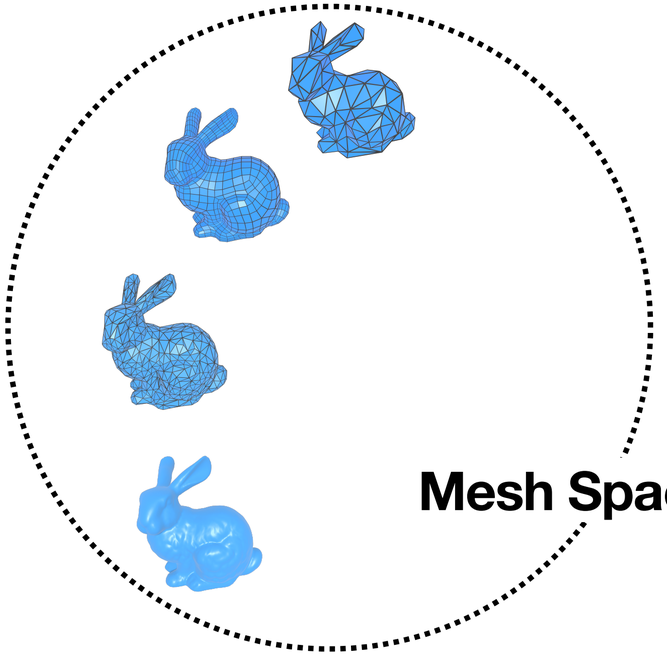
High Rendering Cost

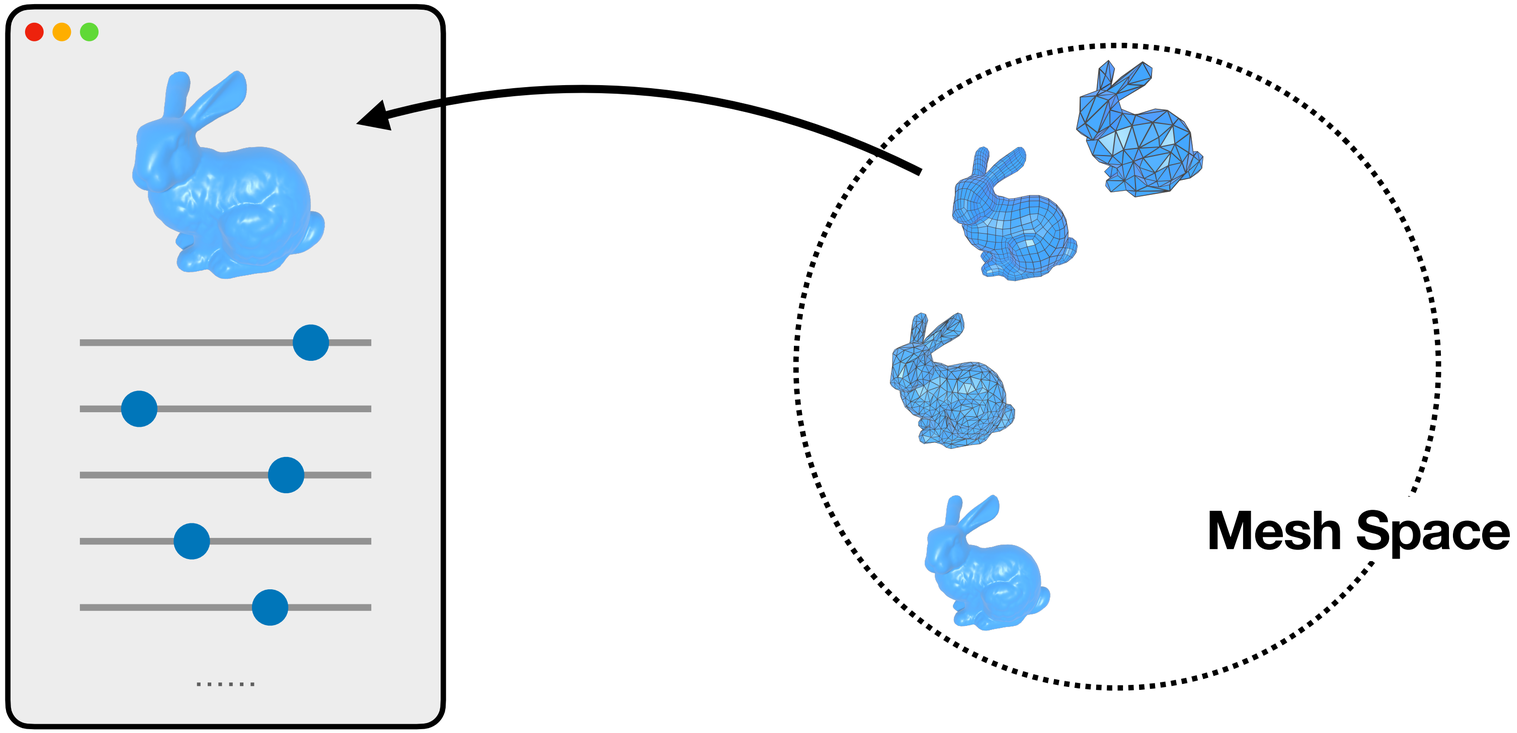
Trade-off

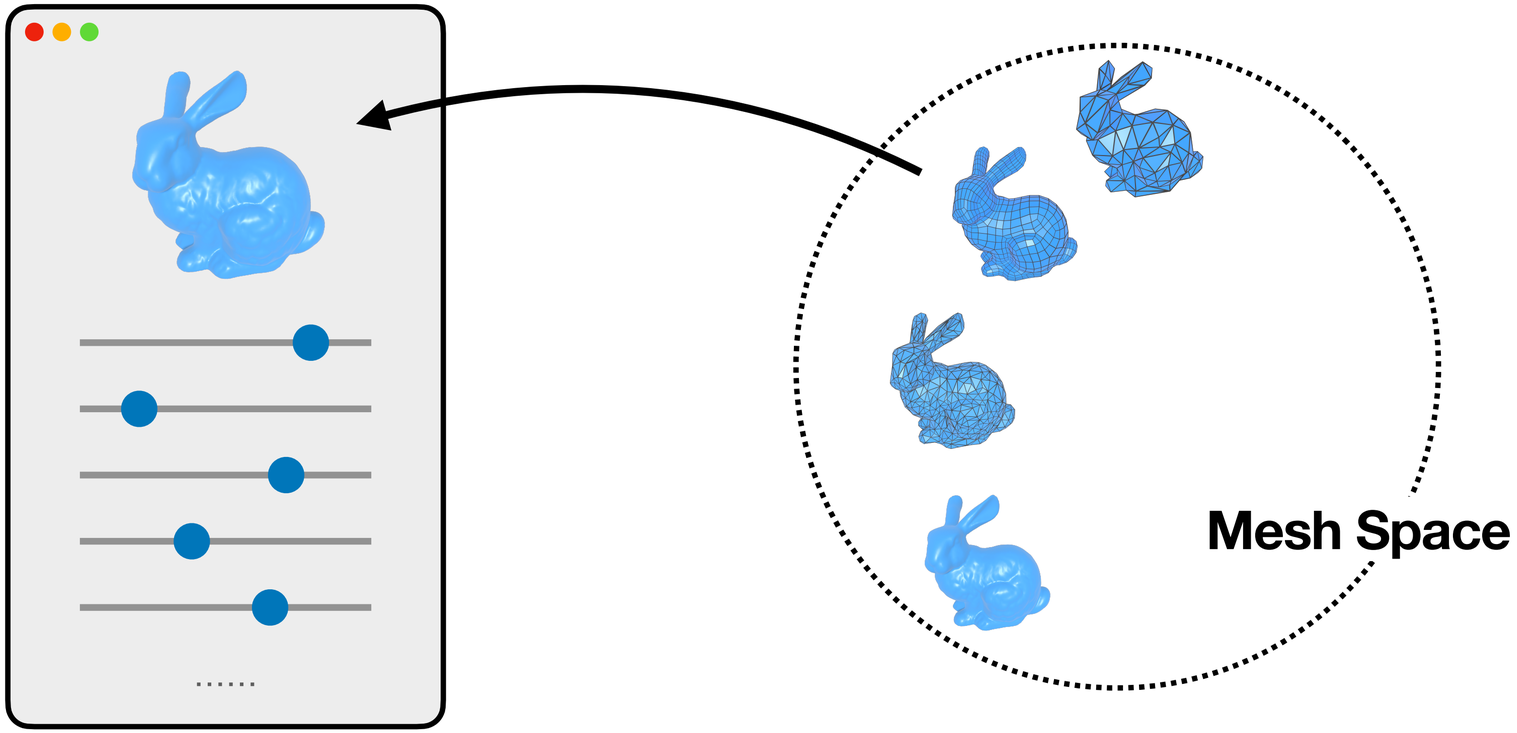






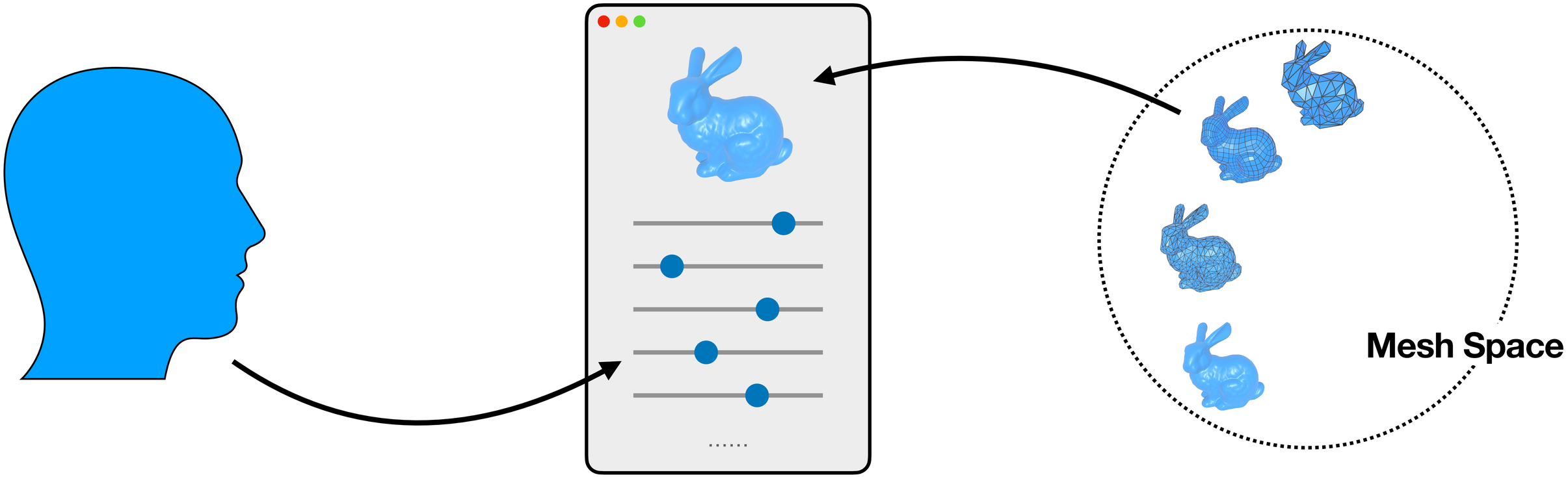

Mesh Space


Mesh Space



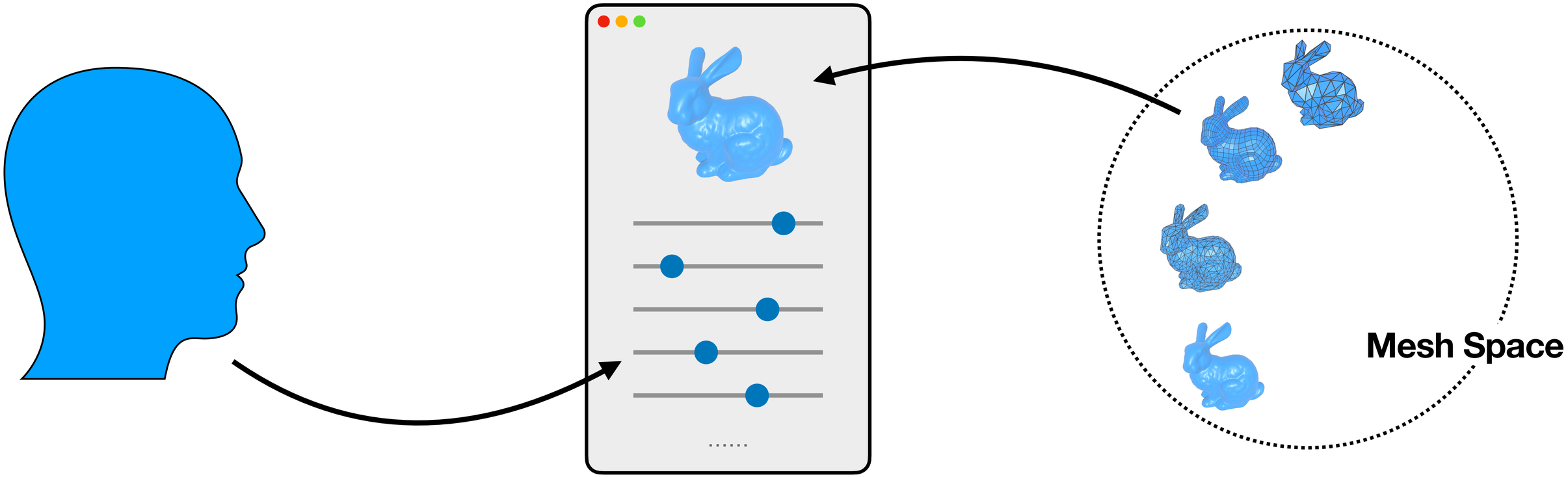
How to get the best outcome using this interface?

Bring A Human into the Loop



How to get the best outcome using this interface?

Bring A Human into the Loop



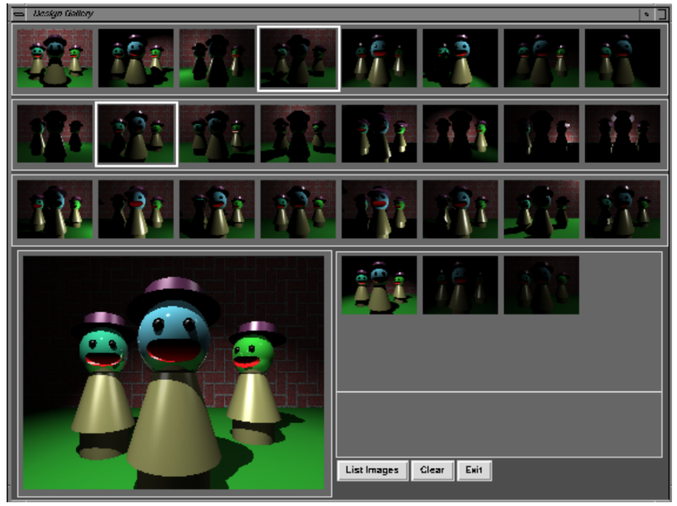
“I have no idea, but let’s do some exploration and accumulate few experiences”

Bayesian Optimization

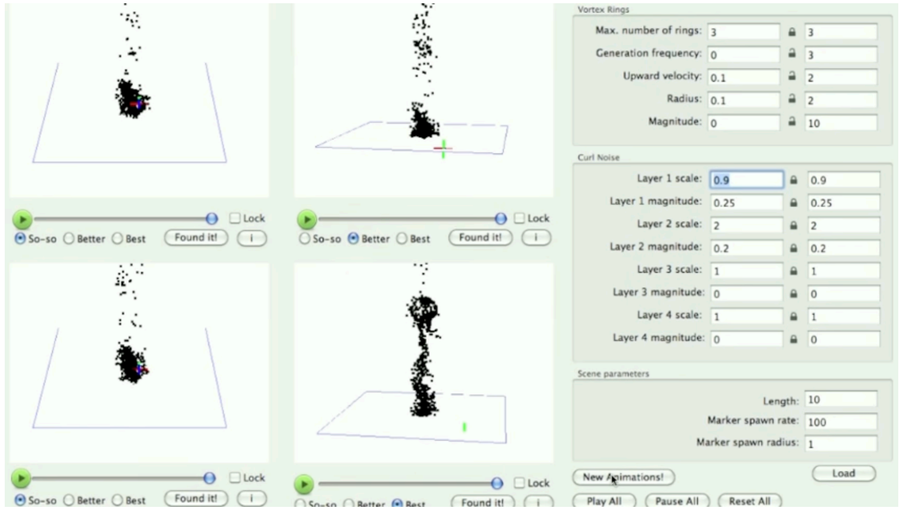
$$p^* = \text{argmax}_{p \in \mathcal{P}} h(M(p))$$

Bayesian Optimization

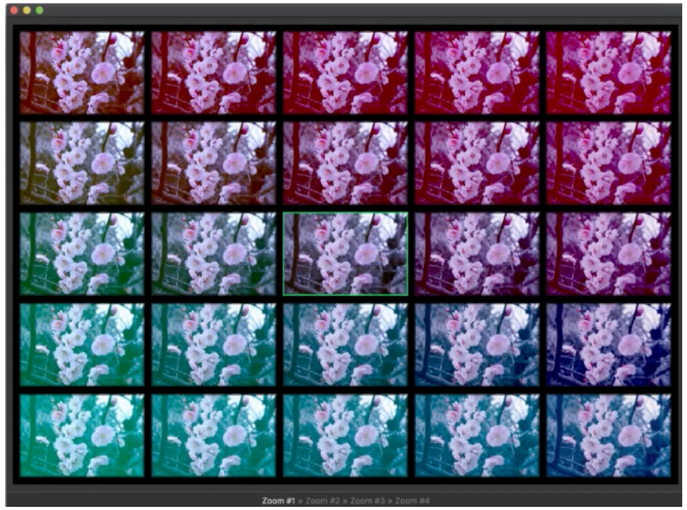
$$p^* = \text{argmax}_{p \in \mathcal{P}} h(M(p))$$



[Marks et al. 1997]

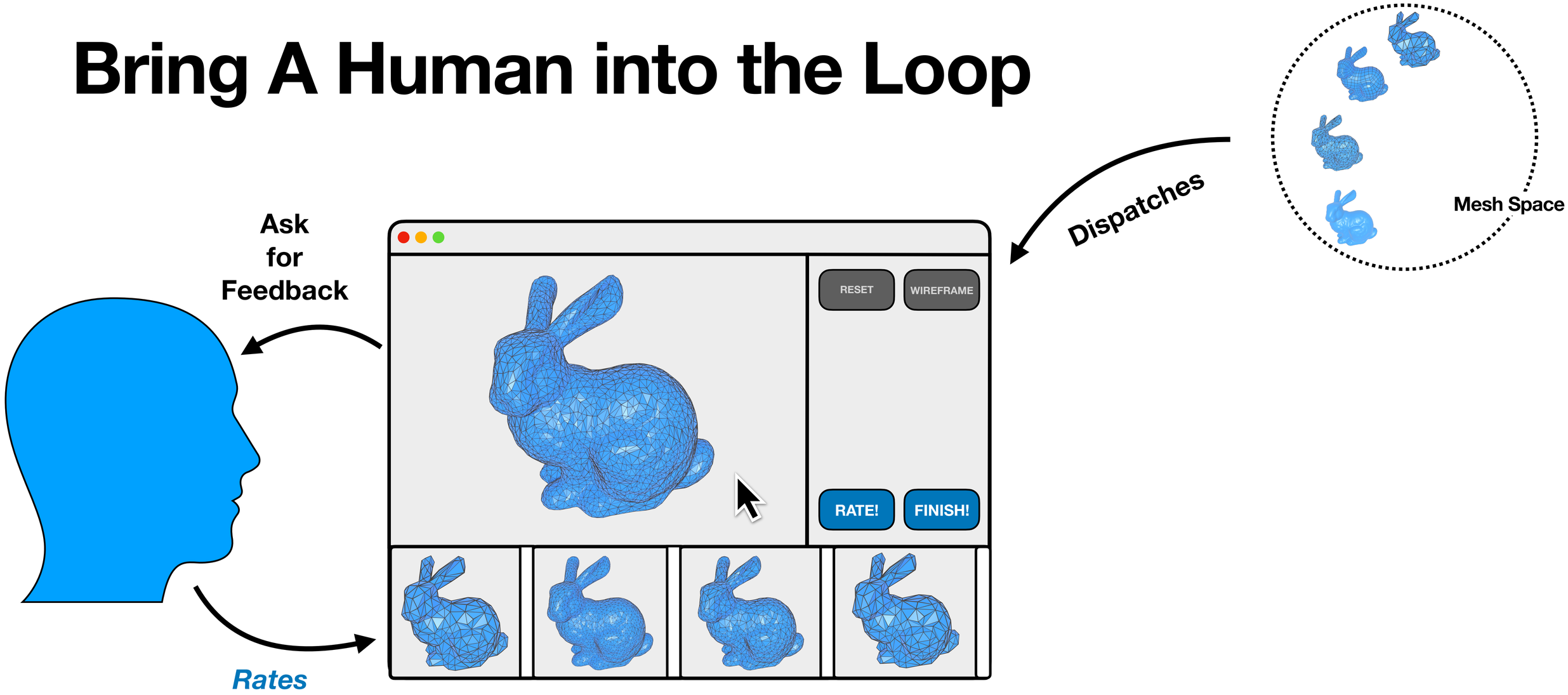


[Brochu et al. 2007]

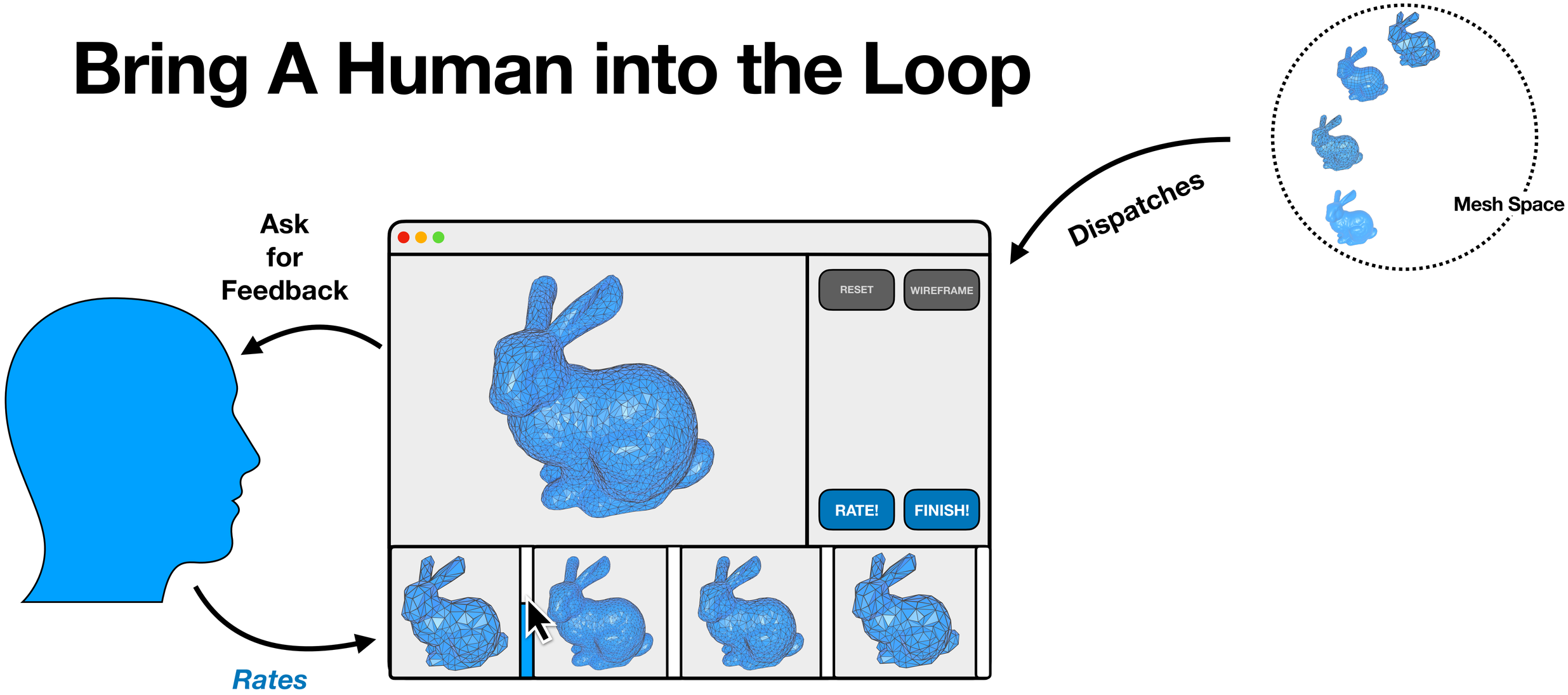


[Koyama et al. 2020]

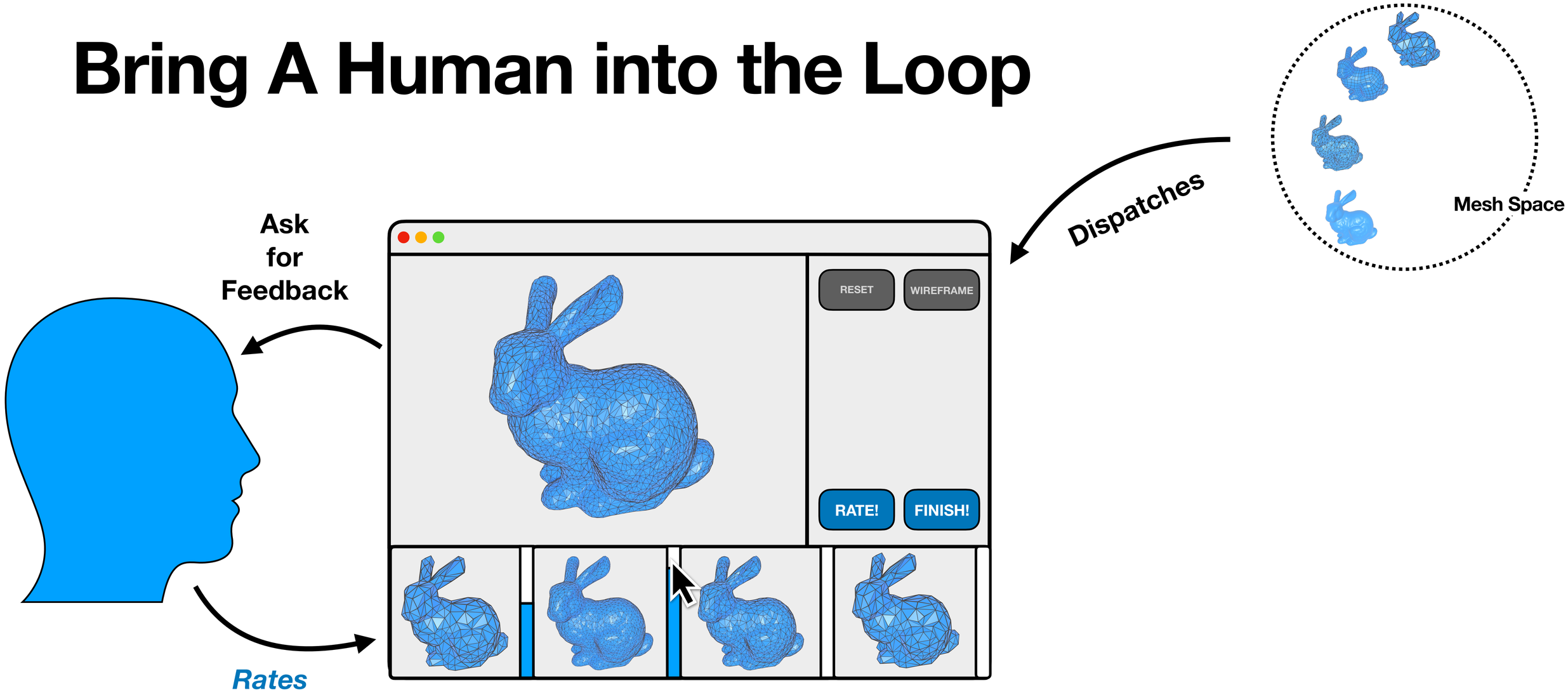
Bring A Human into the Loop



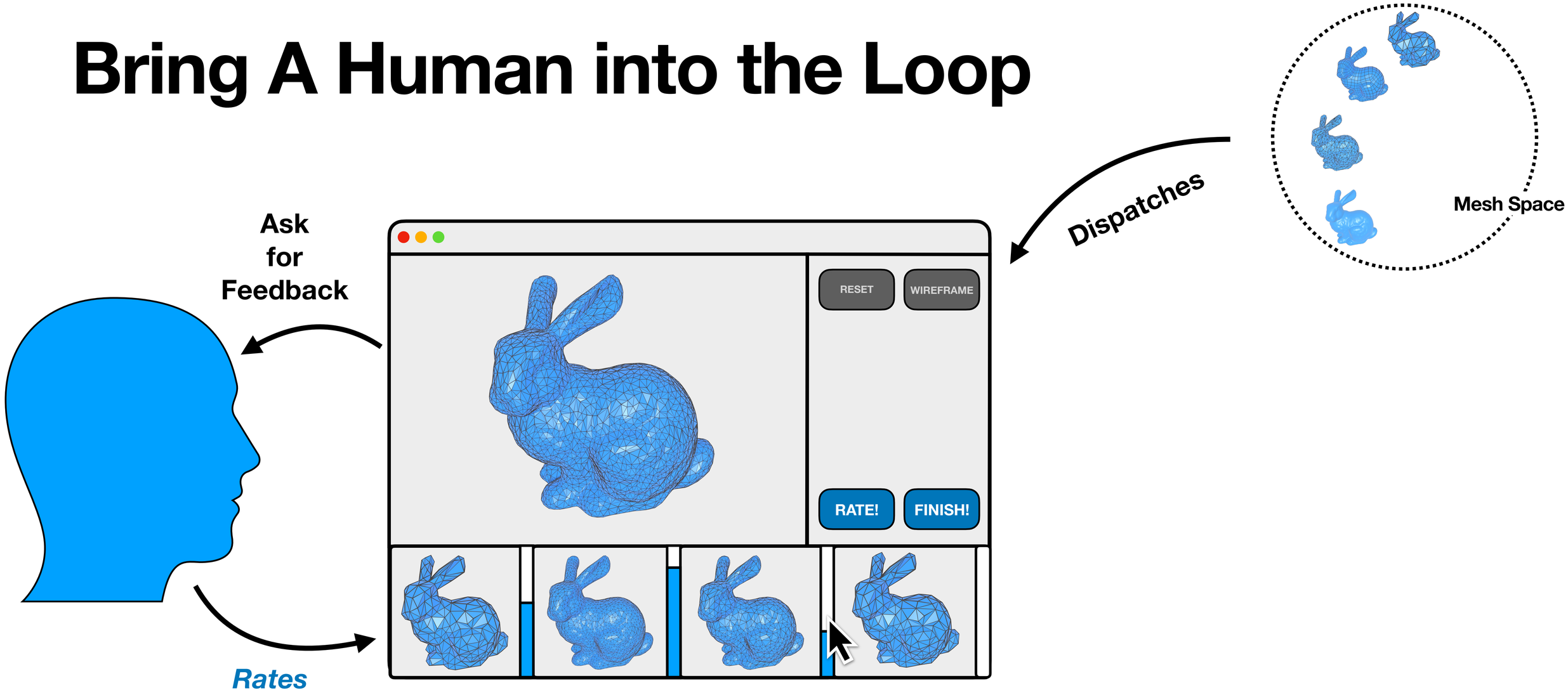
Bring A Human into the Loop



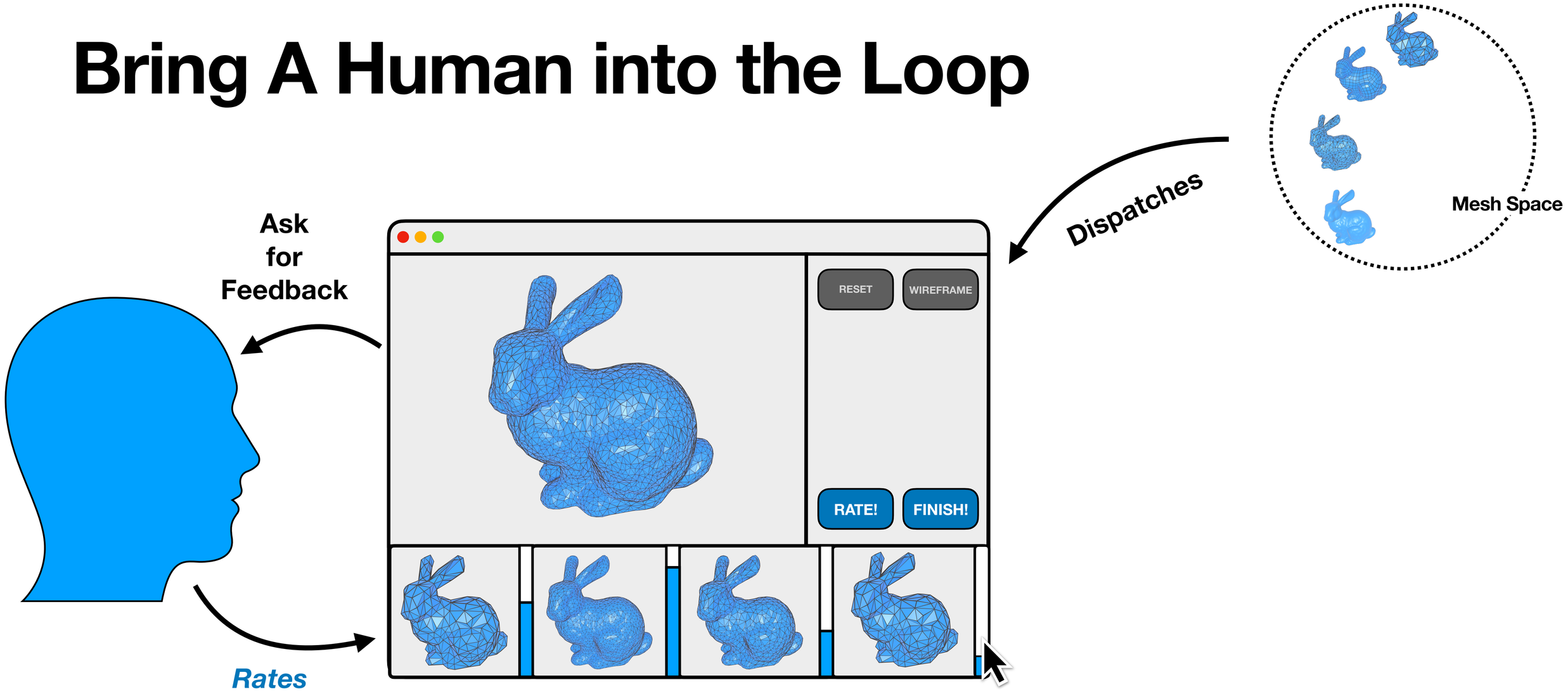
Bring A Human into the Loop



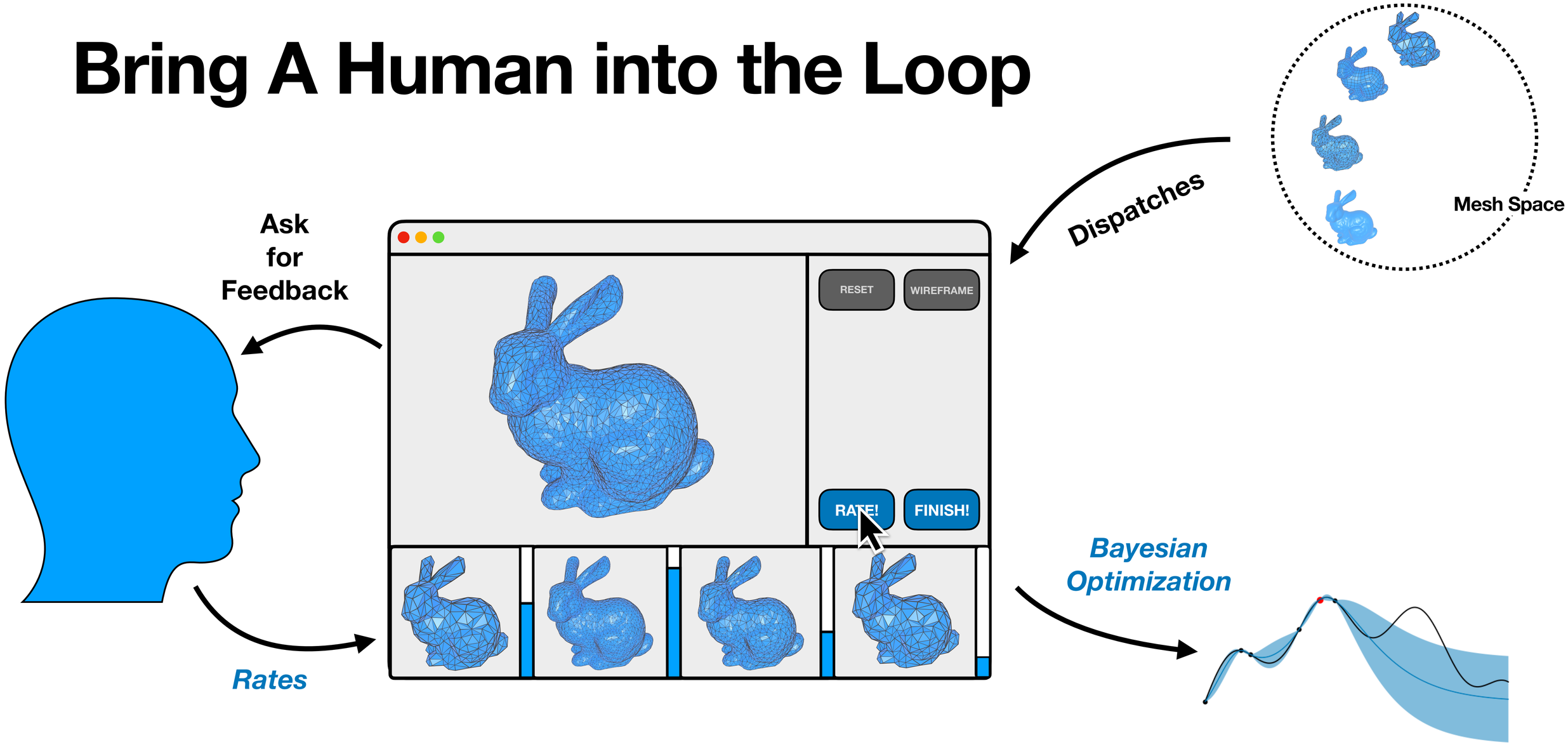
Bring A Human into the Loop



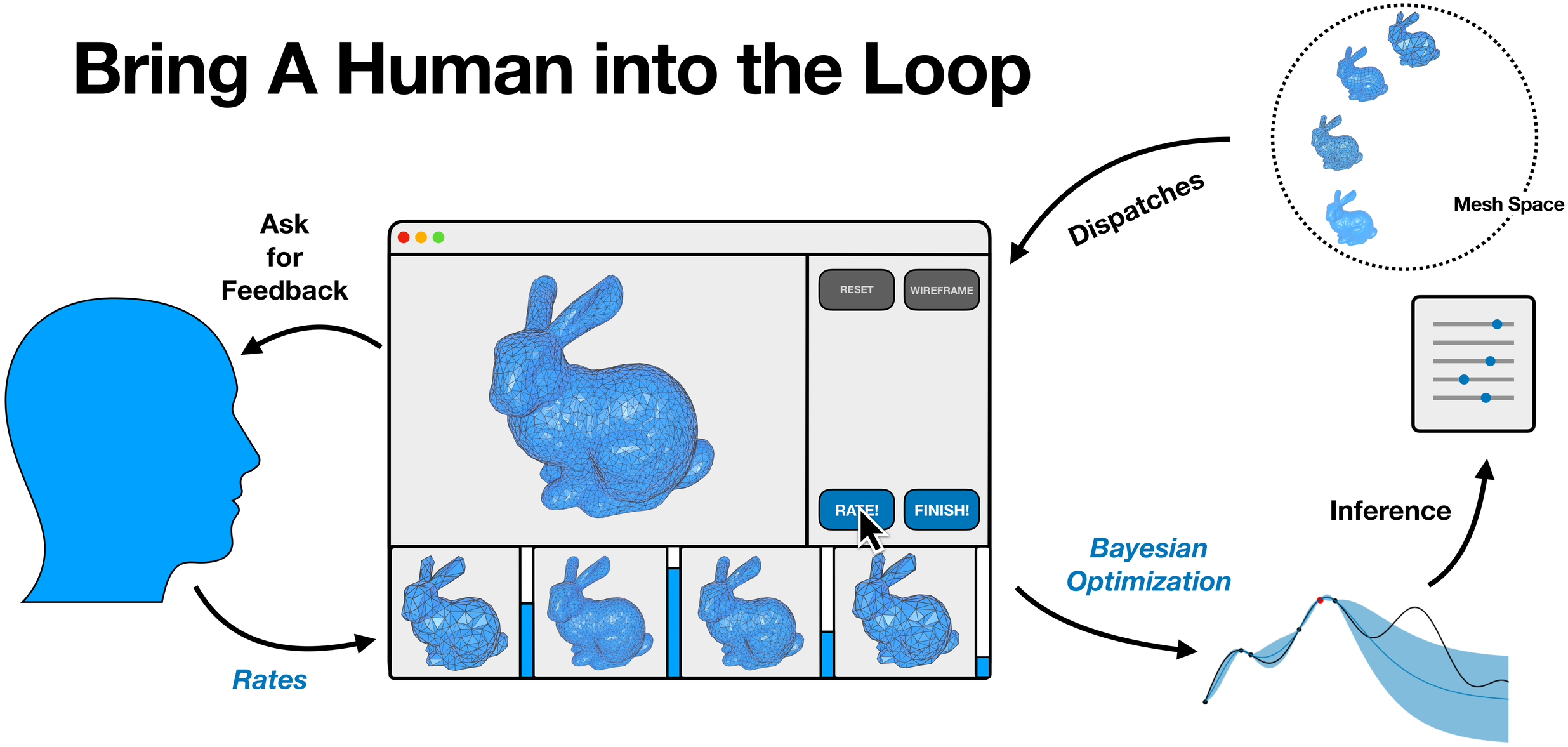
Bring A Human into the Loop



Bring A Human into the Loop



Bring A Human into the Loop

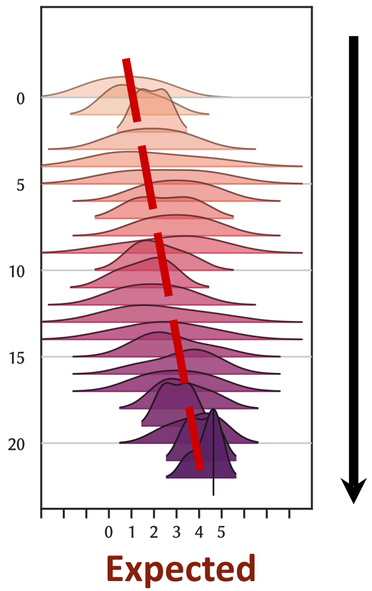


Experiments

- Field study
  - 3 months of usage by 2 technical artists
- Lab study
  - 20 participants, each per 90 minutes

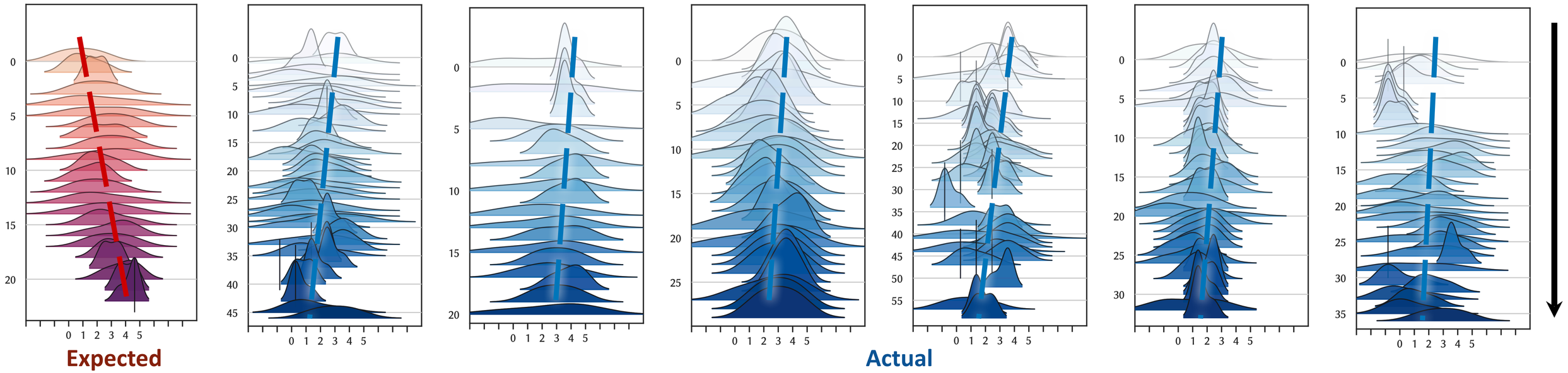
Rating Process: The Intuition

- If the AI (i.e., Bayesian optimizer) successfully optimized the outcome, the overall ratings should towards more to the “right” (Higher ratings).



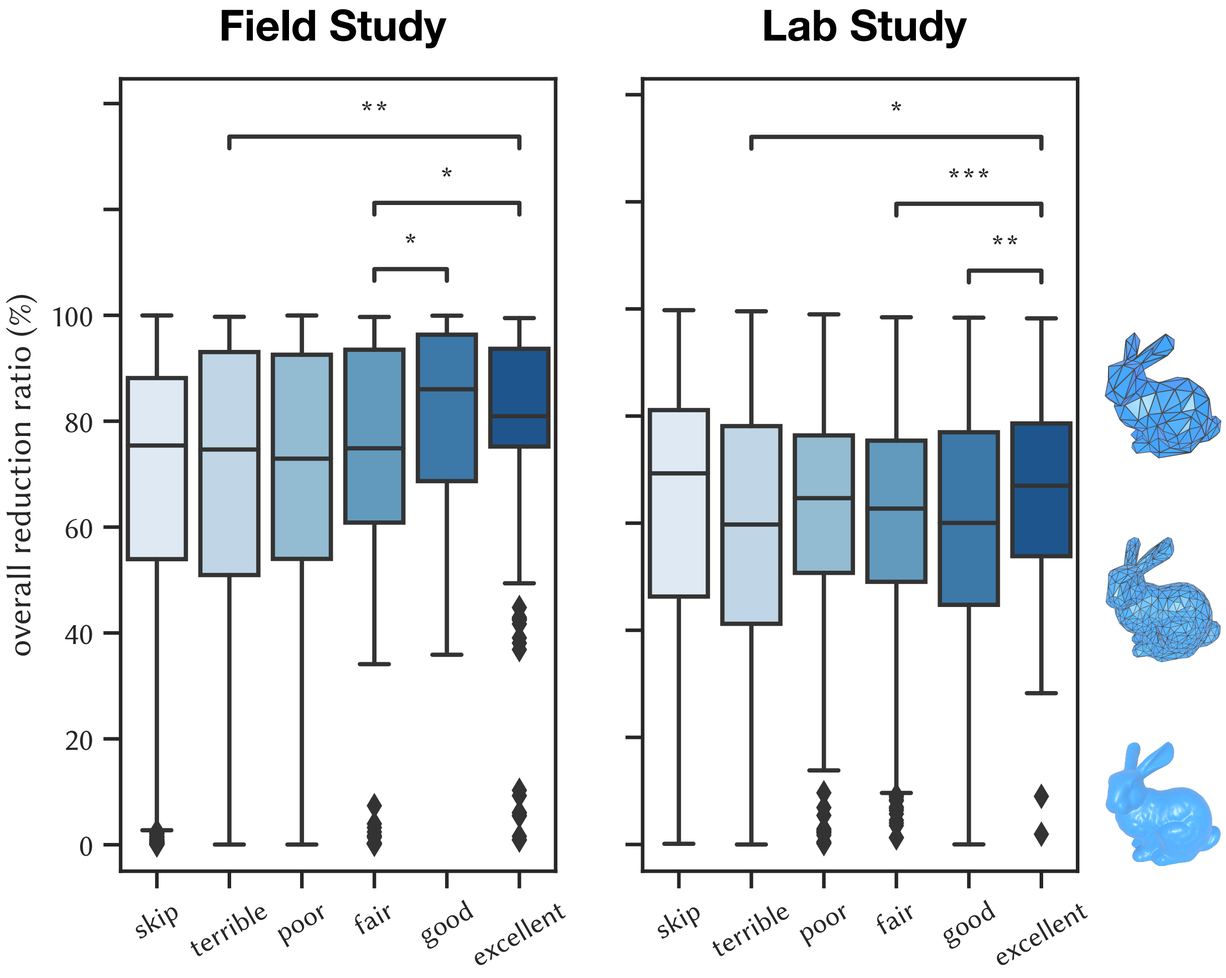
Rating Process: Observations

- Large amount of mismatch between expected and actual ratings
- Either stationary or non-stationary decreasing (Augmented Dickey-Fuller & Mann-Kendall)



Did the Optimization Work?

- Partially.
  - Objectively higher reduced models were rated higher (Mann-Whitney U)
  - Subjective satisfaction:
    - Field 11.9%, Lab 48.5%



Pitfalls: Human Side

Pitfalls: Human Side

- Heuristic bias
  - “This is similar to …” (But actually quite different)

Pitfalls: Human Side

- Heuristic bias
  - “This is similar to …” (But actually quite different)
- Loss aversion
  - “I’ve seen better results before, but the results are getting worse and worse”

Pitfalls: Human Side

- Heuristic bias
  - “This is similar to …” (But actually quite different)
- Loss aversion
  - “I’ve seen better results before, but the results are getting worse and worse”
- Diminishing returns
  - “I can’t see any differences”

Pitfalls: Machine Side

AI algorithms often assume:

Pitfalls: Machine Side

AI algorithms often assume:

- Stable (latent) preference assumption
  - “I’ve changed my mind”
  - “X is better in A, B, C but Y is better in D, E, F”

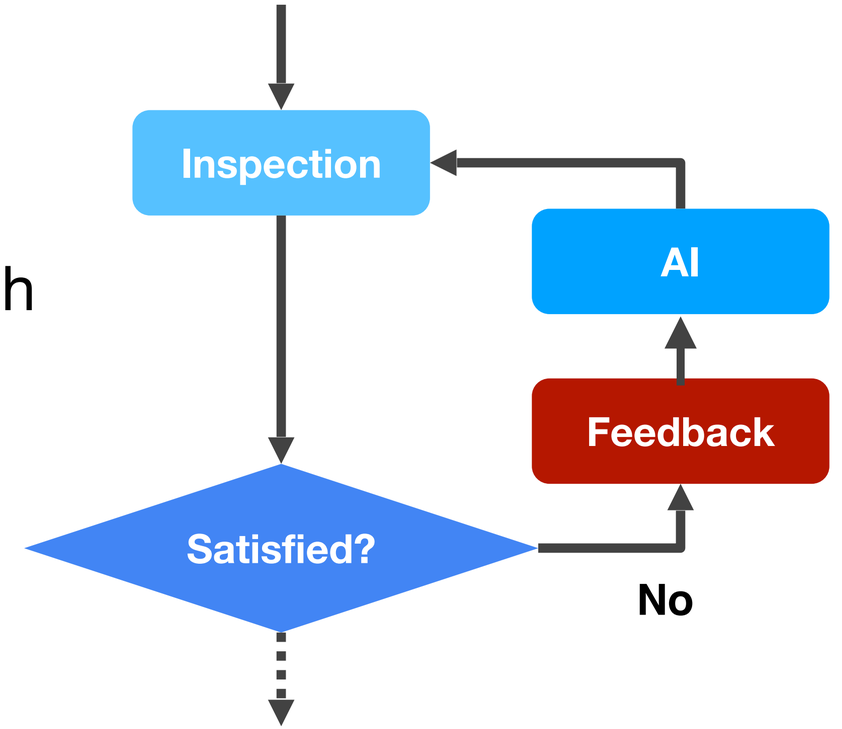
Pitfalls: Machine Side

AI algorithms often assume:

- Stable (latent) preference assumption
  - “I’ve changed my mind”
  - “X is better in A, B, C but Y is better in D, E, F”
- Complete preference assumption
  - “I don’t know”

Reflections

- A human-AI interaction loop may be successful if humans can provide feedback that respects the underlying algorithm assumptions
- In the worst case, a human may never be satisfied with the results and be kept in an *infinite optimization loop*
- Thankfully, there are potential countermeasures that could be evaluated in the future



- Software SDKs, datasets, scripts, etc. are open sourced:

  https://changkun.de/s/infloop

The Human in the *Infinite Loop*

A Case Study on Revealing and Explaining Human-AI Interaction Loop Failures

Changkun Ou, Daniel Buschek, Sven Mayer, and Andreas Butz

Mensch und Computer 2022
SE04: Artificial Intelligence

September 6, 2022
Darmstadt, Germany