



Understanding Generative AI Capabilities in Everyday Image Editing Tasks



Mohammad Reza Taesiri³, Logan Bolton¹, Brandon Collins¹, Viet Dac Lai², Franck Dernoncourt², Trung Bui², Anh Totti Nguyen¹

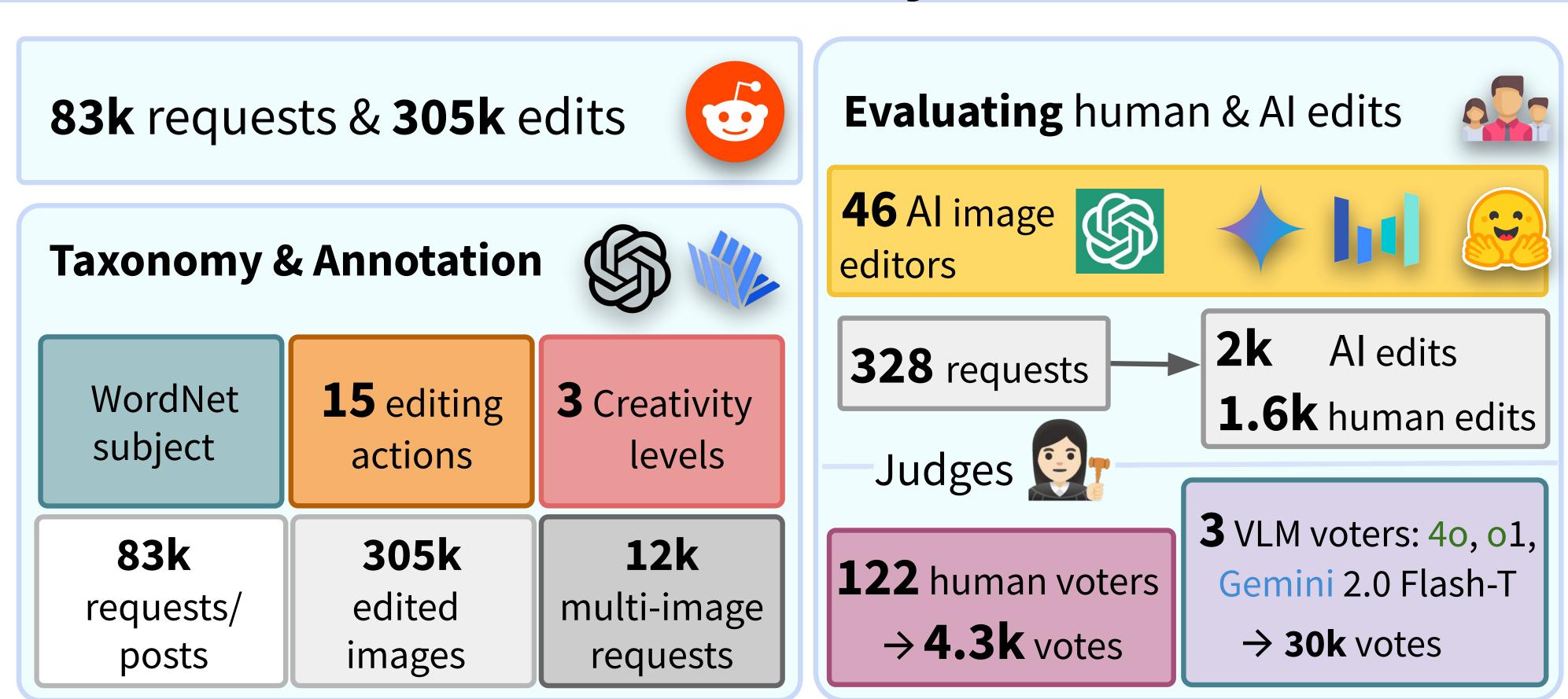
Motivation

- Image editing models (Gemini, GPT-4o, SeedEdit, Instruct-pix2pix) are typically benchmarked/trained on *synthetic datasets*
- How well do models perform on **real-world tasks** that consumers ACTUALLY want?

Data Collection

- Collected 10 years of historical data from /r/PhotoshopRequest
 - 83k edit requests & 305k human-edited images
- 1,654 Al-created edits for a random subset of edit requests

Summary



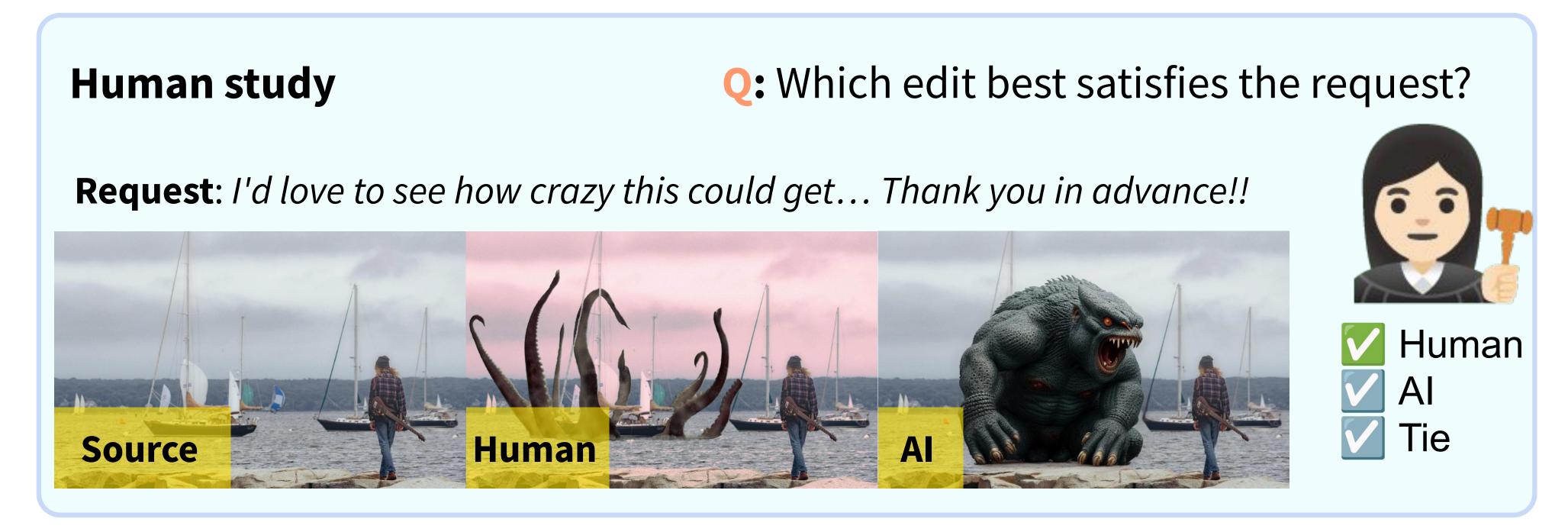


Image editing models can complete ~1/3 of all real-world editing requests

Experiments

Task: Given a source image and edit request, determine whether the human edit or the AI edit is better

Data: 1,312 human edits and 1,654 Al edits

Judge	AI Wins	Human Wins	Tie	Count
2 Human	25.8	66.0	8.2	4,359
© 01	47.5	51.6	0.9	10,352
GPT-40	42.1	52.4	5.5	10,313
♦ Gemini-2.0-Flash-Thinking	52.5	41.9	5.5	10,354

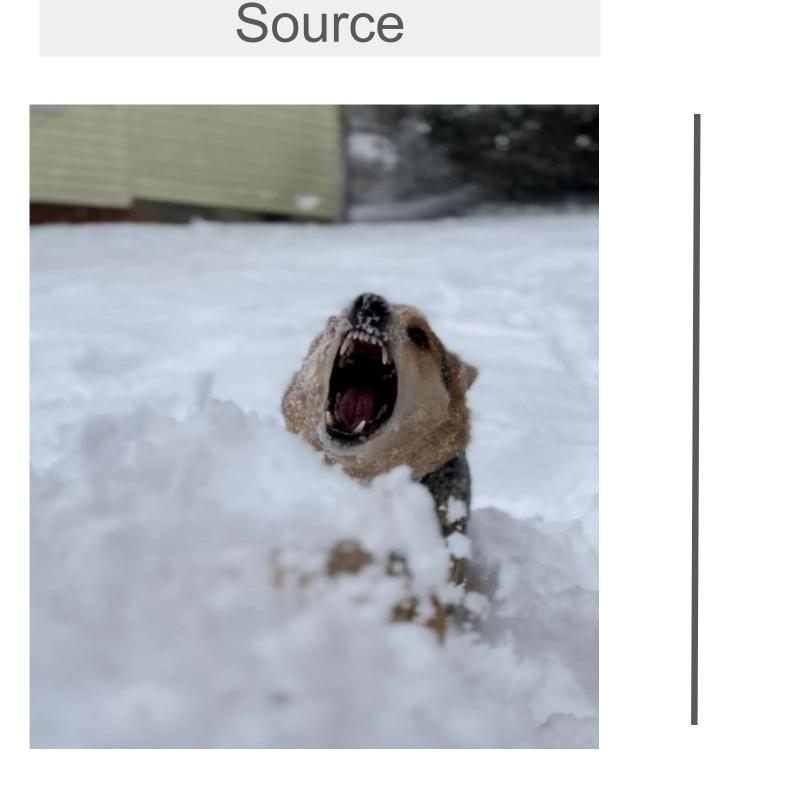
Evaluations

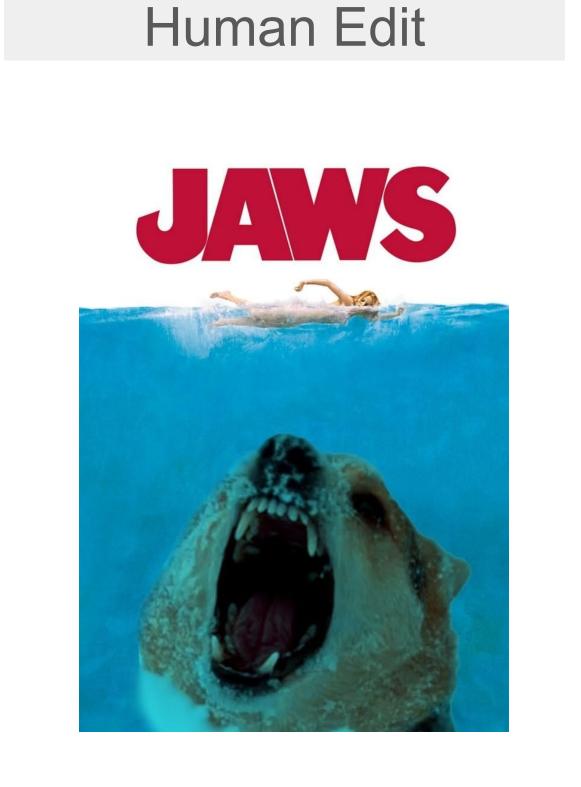
- . Human as a Judge
- i. Collected 4,359 votes from 122 unique users
- 2. VLM as a Judge
 - i. VLM judges have high disagreement with human voters
- 3. LAION Aesthetic Score
 - i. Metric that correlates with "aesthetic" rating

Main Findings

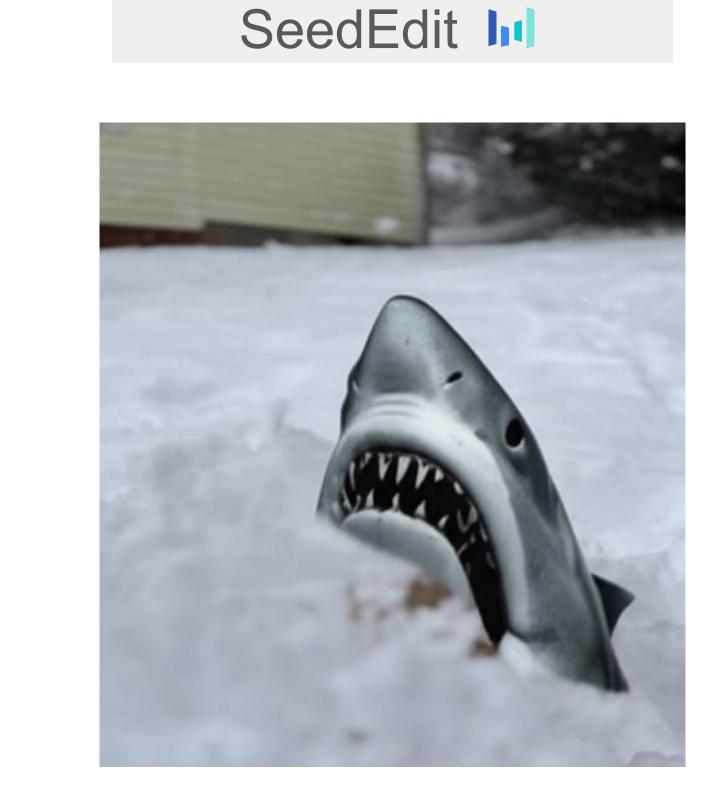
- 1. Automated metrics are poor predictors of human preference
 - i. Human voters prefer human edits 66% of the time
- ii. o1 selects edits from GPT-4o 83.9% of the time (heavily biased)
- 2. Al-generated edits struggle with detail-oriented tasks
 - i. Text, faces, etc.
- 3. Al-generated edits perform best at high-creativity requests
- 4. Generalist autoregressive models (Gemini, GPT-4o) are equally good or better than specialized image editing models.

Source Human Edit Gemini SeedEdit In SeedE

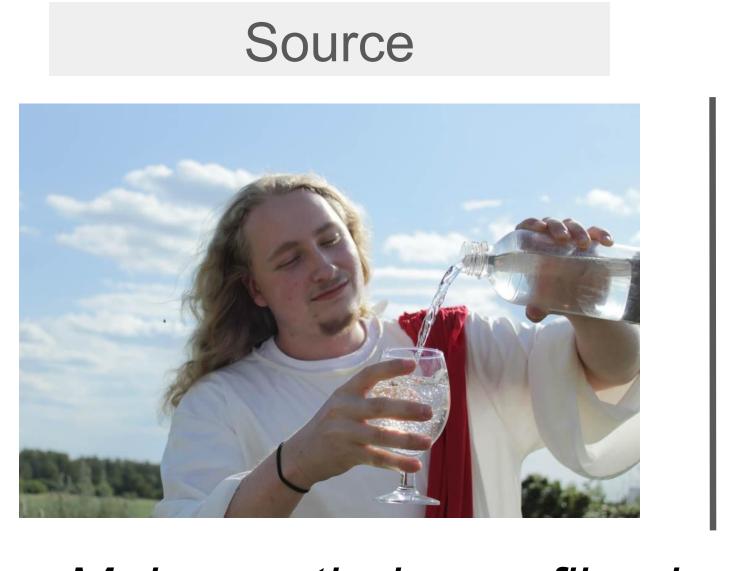




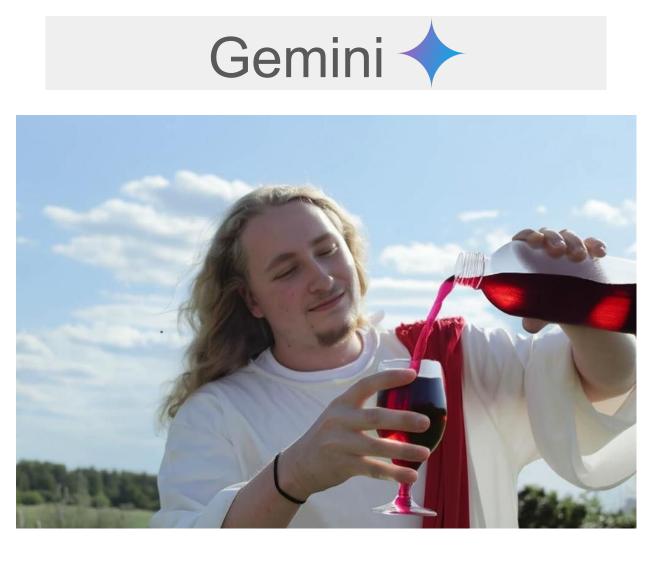




Edit the image to replace the dog's head with a shark's head from the movie 'Jaws', maintaining the same expression and positioning.









Make my tinder profile about Jesus perfect! Please make the water stream and the water in the glass red like wine!

