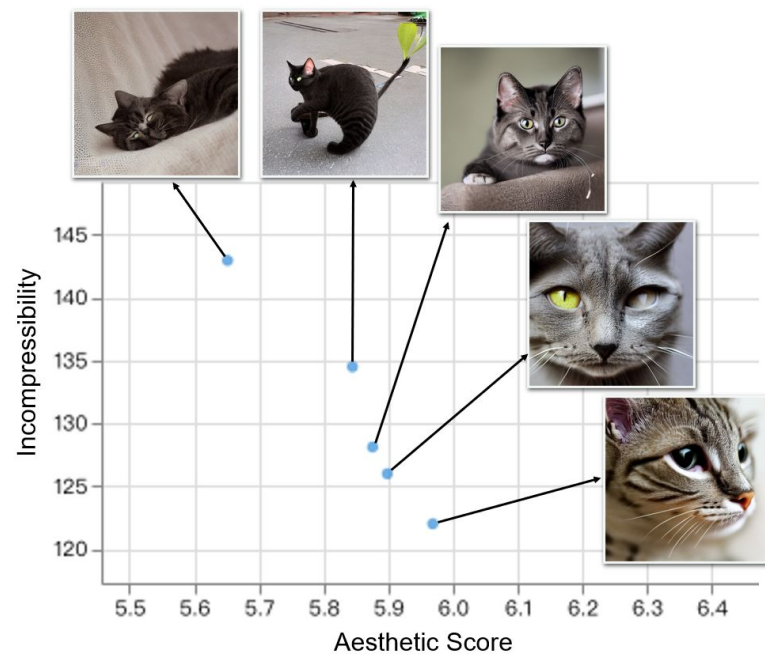
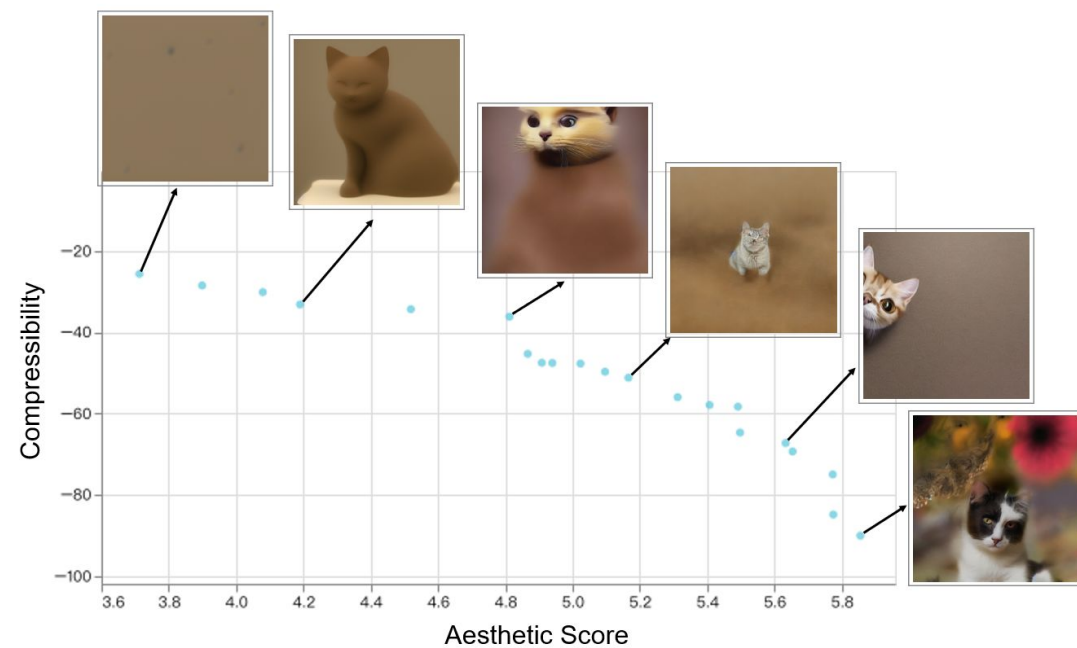


# Pareto front



# Compressibility & Aesthetic



# Incompressibility & Aesthetic





# Overtunning (single objective)

aesthetic/incompressibility with pretrained alignment



# Multi-objective

aesthetic + incompressibility with pretrained alignment



## Other results

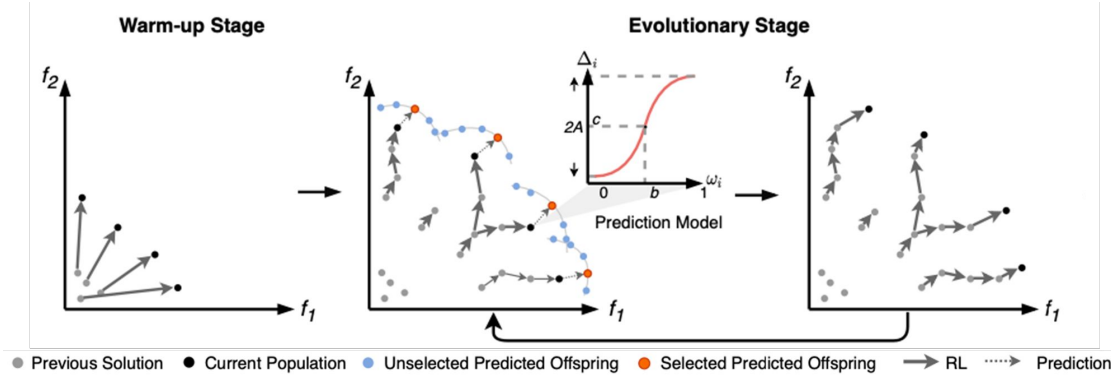


Weight  $\omega_i$  decides the ratio of different rewards.

Task set  $T = \{(\pi_i, \omega_i)\}_{i=1}^n$  selected with improvement prediction model.

MOPG  $\rightarrow$  given  $T$ , update policy  $\pi_i$  using weight  $\omega_i$  and generate  $n$  new policies.

Select new policies from the  $n$  policies.



#### ▷ Warm-up Stage

Generate task set  $T = \{(\pi_i, \omega_i)\}_{i=1}^n$  by random initial policies and evenly distributed weight vectors.

$\mathcal{P}' \leftarrow \text{MOPG}(T, m_w, \mathcal{R})$  (Section 3.2)

Update  $\mathcal{P}$  and EP with  $\mathcal{P}'$ .

#### ▷ Evolutionary Stage

**for** generation  $\leftarrow 1, 2, \dots, M$  **do**

Fit improvement prediction models  $\{\Delta^i\}$  for each policy in  $\mathcal{P}$  from data in  $\mathcal{R}$ . (Section 3.3)

$T \leftarrow \text{TaskSelection}(n, \mathcal{P}, \{\Delta^i\}, \text{EP})$  (Section 3.4)

$\mathcal{P}' \leftarrow \text{MOPG}(T, m_t, \mathcal{R})$  (Section 3.2)

Update  $\mathcal{P}$  and EP with  $\mathcal{P}'$ .

**end for**