Towards a Unified Multimodal Foundation

Evaluating · Aligning · Generating

Yasi Zhang

Department of Statistics and Data Science, UCLA

Advisors: Prof. Ying Nian Wu & Prof. Oscar Leong

April 2026

Outline

About Me (CV)

A Virtuous Cycle — Research Vision

PART I

Evaluating: Automated, Scalable, Fine-Grained Evaluation

EdiVal-Agent (ICLR 2026) ▪ *MT-EditFlow (in submission)*

PART II

Aligning: Regression-Aware RL for LLM-as-a-Judge

REAL (ICML 2026) ▪ *EBAMA (ECCV 2024)*

PART III

Generating: Learning under Imperfect Observations

RSD (ICLR 2026) ▪ *Flow Priors (NeurIPS 2024)*

About Me

Yasi Zhang (She/They)

yasminzhang.github.io

my resume

Education

Ph.D., Statistics & Data Science, UCLA (2022–present)

B.S., Data Science & Big Data Tech., Fudan University

Industry Experience

Student Researcher, ByteDance Seed (Spring 2026)

Student Researcher, Google Research (Spring&Summer 2025)

Applied Scientist Intern, Amazon AWS AI Labs (Summer 2024)

Research Interests

Generative AI, Multimodality, Reinforcement Learning.

Why Multimodal Foundation Models?

Human intelligence is inherently multimodal.

1. **Language** is only one channel—it cannot express spatial layout, visual texture, temporal dynamics, or physical intuition.
2. **Humans** perceive the world through vision, hearing, touch, and spatial reasoning; cognition seamlessly integrates all these modalities.
3. A truly **intelligent system** must do the same.

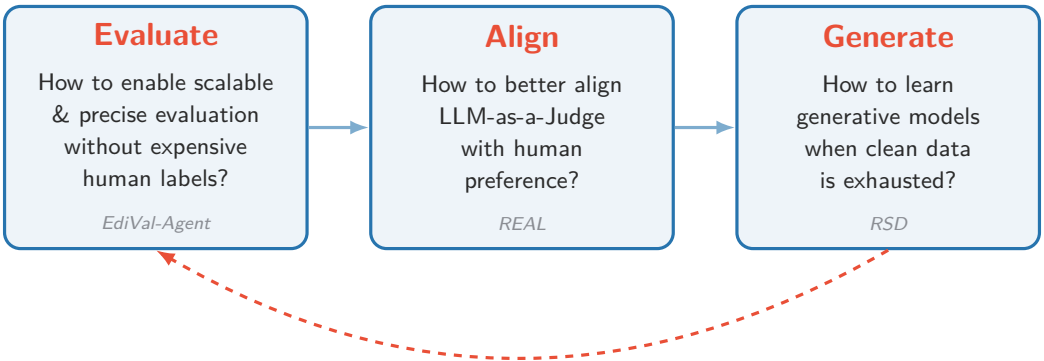
The gap between ambition and reality

Despite rapid progress, current multimodal models remain brittle:

- × **Evaluation Bottleneck**: no scalable way to measure quality without expensive human labels.
- × **Alignment Gap**: automated judges are misaligned with human preference.
- × **Data Scarcity**: clean multimodal data is finite, yet models are data-hungry.

A Virtuous Cycle

My research addresses these three gaps through a virtuous cycle of



PART I

Evaluating

How to enable scalable & precise evaluation
without expensive human labels?

EdiVal-Agent: An Object-Centric Framework for
Automated, Fine-Grained Evaluation of Multi-Turn Editing. ICLR 2026.

MT-EditFlow: Reinforcement Learning for
Multi-Turn Image Editing with Flow Matching. In submission.

EdiVal-Agent — Motivation

Problem: Instruction-based image editing has advanced rapidly, yet scalable and precise evaluation remains a bottleneck.

Why Current Evaluation Protocols Fail?

Reference-Based Metrics (L1, L2, CLIP, etc.)

- × Require ground-truth edited images — hard to collect high-quality ones.
- × The space of acceptable edits is large; a single reference captures only one realization.
- × References are often synthesized by existing editors, importing their biases into evaluation.

Zero-Shot VLM Judges

- × **Instruction Following:** prone to hallucinations in object existence, attributes & spatial/numerical reasoning.
- × **Content Consistency:** limited sensitivity to pixel-level changes; miss subtle localized edits.
- × **Visual Quality:** pretrained on natural images — miscalibrated for synthetic artifacts (e.g. extra fingers).

Why Multi-Turn Image Editing?

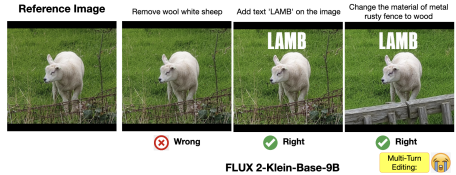
1. Multi-turn is the realistic setting.

Users iteratively refine images based on a model's own previous outputs — not single isolated edits.

2. Multi-turn is where current models struggle:

- **All-or-Nothing:** one failed turn compromises the entire sequence.
- **Error Propagation:** mistakes compound and derail subsequent turns.
- **Exposure Bias:** models conditioned on their own outputs may degrade quickly.

3. EdiVal-Agent is the first evaluation framework for multi-turn editing.

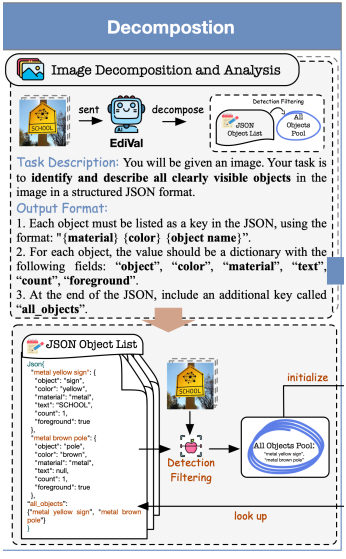


All-or-Nothing Requirement

Error Propagation

Exposure Bias

EdiVal-Agent — Stage 1: Decomposition



Goal

Parse each image into structured, object-level descriptions to enable symbolic reasoning in later stages.

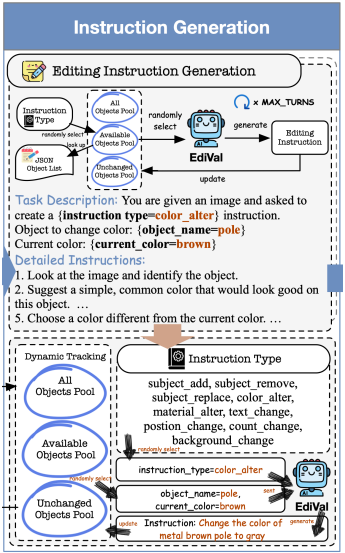
Method

- A VLM (GPT-4o) extracts per-object JSON:
  fields: `object`, `color`, `material`, `text`, `count`, `foreground`
- Naming convention: `"{material} {color} {object}"`
- Grounding-DINO validates detections & provides bounding boxes to avoid hallucinations.
- Reliable objects stored in the **All Objects Pool**.

Example

"metal yellow sign": {sign, yellow, metal, text=SCHOOL, foreground=true},
"metal brown pole": {pole, brown, metal, text=null, foreground=true}

EdiVal-Agent — Stage 2: Instruction Generation



Goal

Produce diverse, multi-turn editing instructions grounded in the current scene state.

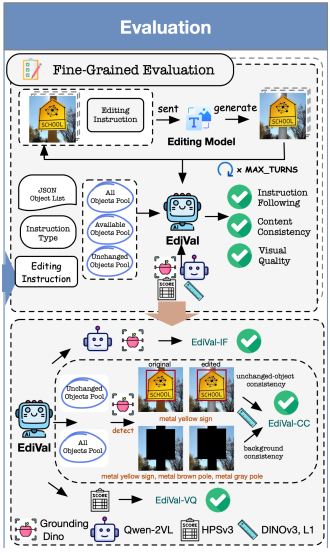
Method

- Maintain three evolving pools at each turn t: $\mathcal{P}_t^{\text{all}}$ (all objects), $\mathcal{P}_t^{\text{unch}}$ (unedited), $\mathcal{P}_t^{\text{avail}}$ (editable)
- 9 instruction types across 6 categories: subject, attribute, text, relational, counting, global
- Per turn: sample unused type → select objects from $\mathcal{P}_t^{\text{avail}}$ → emit instruction via GPT-4o → update pools.

Example (3-turn chain)

T1: "Change the color of metal brown pole to gray."
T2: "Remove the metal yellow sign."
T3: "Change the background to a library."

EdiVal-Agent — Stage 3: Evaluation



Goal

Assess each edit along three orthogonal axes, using the structured scene representation from Stages 1–2.

EdiVal-IF (Instruction Following)

- **Symbolic** (add/remove/...): open-vocab detector + geometric & logical checks.
- **Semantic** (color/material/...): VLM on detector-guided crops.

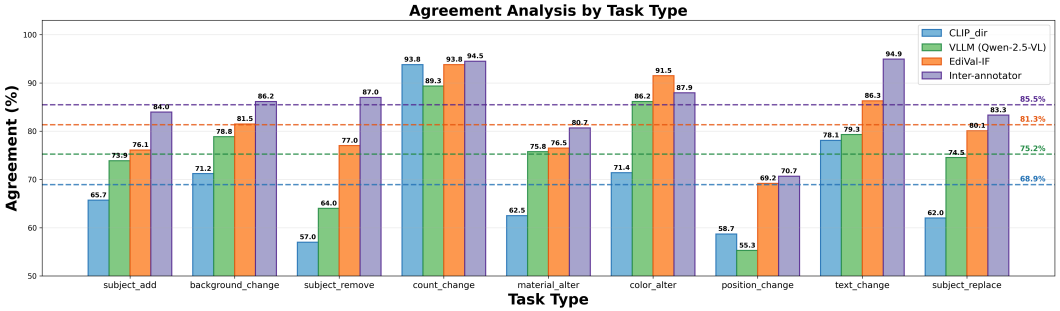
EdiVal-CC (Content Consistency)

- DINO similarity of unchanged objects ($\mathcal{P}_t^{\text{unch}}$) & background ($\Omega \setminus \mathcal{P}_t^{\text{all}}$) between I_0 and I_t.

EdiVal-VQ (Visual Quality)

- Human Preference Score v3 (HPSv3); reported separately as aesthetic preference is task-dependent.

EdiVal-Agent — Alignment with Human Judgments



Key finding

EdiVal-IF achieves higher human agreement accuracy than prior automated metrics.

Compared baselines

- CLIP-based: CLIP-directional distance
- VLM zero-shot judges: Qwen2-VL
- Upper bound: inter-annotator agreement

Why EdiVal-IF wins

- Object-centric decomposition avoids holistic averaging.
- Symbolic checks catch exact failures that VLMs hallucinate past.
- Crop-guided VLM queries reduce spatial confusion.

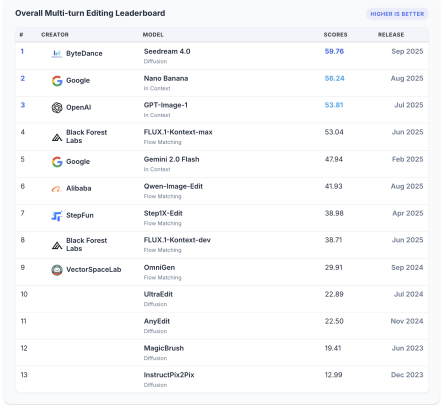
EdiVal Bench — Benchmark Overview

Benchmark at a glance

- **9 instruction types** across 6 categories (subject, attribute, text, relational, counting, global).
- **16 SOTA editing models** spanning in-context, flow-matching, and diffusion paradigms.
- Multi-turn sequences of 3 turns per image, exposing compounding effects invisible to single-turn benchmarks.

Why it matters

First large-scale benchmark that evaluates editing models in the *realistic multi-turn* setting with *fine-grained, object-level* metrics.



Overall Multi-turn Editing Leaderboard — HIGHER IS BETTER

#	CREATOR	MODEL	SCORES	RELEASE
1	ByteDance	Seedream 4.0 Diffusion	59.76	Sep 2025
2	Google	Nano Banana In Context	56.24	Aug 2025
3	OpenAI	GPT-Image-1 In Context	53.81	Jul 2025
4	Black Forest Labs	FLUX.1-Kontext-max Flow Matching	53.04	Jun 2025
5	Google	Gemini 2.0 Flash In Context	47.94	Feb 2025
6	Alibaba	Qwen-Image-Edit Flow Matching	41.93	Aug 2025
7	StepFun	Step1X-Edit Flow Matching	38.98	Apr 2025
8	Black Forest Labs	FLUX.1-Kontext-dev Flow Matching	38.71	Jun 2025
9	VectorSpaceLab	OmniGen Flow Matching	29.91	Sep 2024
10		UltraEdit Diffusion	22.89	Jul 2024
11		AnyEdit Diffusion	22.50	Nov 2024
12		MagicBrush Diffusion	19.41	Jun 2023
13		InstructPix2Pix Diffusion	12.99	Dec 2023

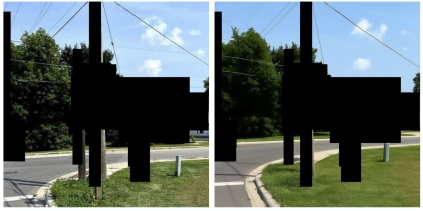
EdiVal-Agent — Qualitative Visualization

EdiVal-CC: Object Consistency

DINO similarity of unchanged objects between I_0 and I_t

(a) Base Image (b) GPT-Image-1 (95.19) (c) Nano Banana (98.05) (d) Qwen-Image-Edit (94.96)

EdiVal-CC: Background Consistency



Base image Qwen-Image-Edit

EdiVal-VQ: Beautification vs. Preservation

(a) Base image (b) GPT-Image-1 (c) FLUX.1-max

EdiVal-Agent — Marginal Task Success Rate

Image vs. Marginal Task Success

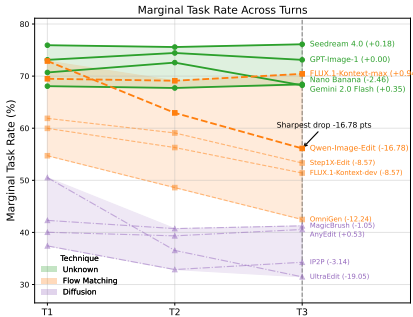
Image Success requires *all* turns up to t to succeed (cumulative), while **Marginal Task Success** measures each turn independently (marginal).

Key observation

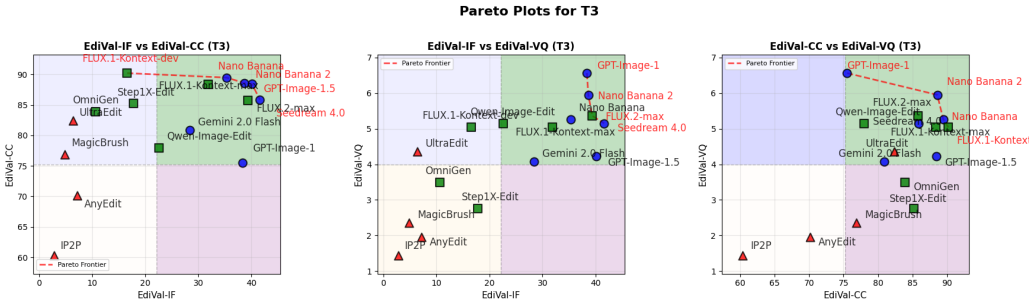
- For closed-source models, marginal task success stays relatively stable across turns.
- Open-source models suffer from exposure bias and accumulate errors across turns.

Implication

A huge improvement room for open-source models in multi-turn editing.



EdiVal Bench — Pareto Front: IF vs. CC vs. VQ Trade-off



What the plot shows

Each point is a model; axes are EdiVal-IF, EdiVal-CC and EdiVal-VQ.

Takeaway

No single model dominates all the 3 metrics—the IF–CC–VQ trade-off is a fundamental axis for future model design.

EdiVal-Agent — Summary & Follow-up

Key Takeaways

- Shifts evaluation from *holistic reference matching* to *compositional, object-level* reasoning.
- Achieves significantly stronger human agreement accuracy than all prior metrics.
- Exposes compounding failure modes that single-turn benchmarks systematically miss.
- Fine-grained, per-type analysis reveals that no single model dominates—informing targeted model improvements.

Follow-up: MT-EditFlow

MT-EditFlow: Reinforcement Learning for Multi-Turn Image Editing with Flow Matching

- Developed MT-EditFlow, a flow-matching reinforcement learning framework that optimizes reward signals specifically for multi-turn image editing applicable to both GRPO and NFT-based RL methods.
- Mitigated compounding errors and exposure bias inherent in iterative editing processes and boosted the FLUX.1- Kontext-dev model's turn-3 performance by 6.85 points, and FLUX.2-Klein's performance by 2.89 points.

PART II

Aligning

How to better align LLM-as-a-Judge
with human preference?

REAL: Regression-Aware Reinforcement Learning for LLM-as-a-Judge. ICML 2026.

EBAMA: Object-Conditioned Energy-Based Attention Map
Alignment in Text-to-Image Diffusion Models. ECCV 2024.

REAL — Motivation

My Background

Alignment has been a long-standing thread of my work—from *attention-level* alignment (EBAMA, ECCV 2024) to *reward-level* alignment (REAL, ICML 2026).

Why LLM-as-a-Judge?

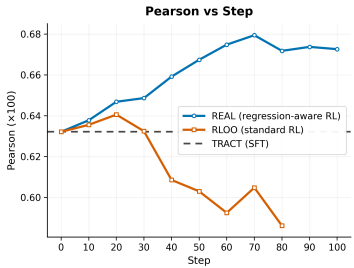
As models scale, human evaluation becomes the bottleneck. LLM/VLM judges that assign *numeric scores* are the default choice for now.

Why current RL training for LLM-as-a-Judge fails

- Intuitively, **Standard RL** uses binary 0/1 rewards: predicting 4 when truth is 5 is penalized the same as predicting 1.
- Empirically, Correlation metrics *collapse* during **standard RL** training (see right figure.).

Why it matters downstream

A miscalibrated judge corrupts preference data, enables reward hacking, and undermines the entire alignment pipeline.



Standard RL degrades Pearson correlation.

REAL — Preliminaries

Standard RL for LLMs

Given prompt x, CoT c, answer y, policy π_θ, the RL objective maximizes a reward:

$$\mathcal{L}(\theta) = \mathbb{E}_{x\sim\mathcal{D},\,(c,y)\sim\pi_\theta(\cdot|x)}\big[r(x,y)\big]$$

Standard REINFORCE Gradient

Treat the full completion as a single action (bandit setting). The policy gradient becomes:

$$\nabla_\theta\mathcal{L}(\theta) = \mathbb{E}\big[r(x,y)\,\nabla_\theta \log \pi_\theta(c,y \mid x)\big]$$

Key assumption: reward does not depend on θ, i.e. $\nabla_\theta r = 0$.

RAIL Inference (Lukasik, et al, 2024)

Instead of greedy decoding a single token, compute the **expected value** over the digit token set $\mathcal{K} = \{0, 1, \ldots, 9\}$:

$$\hat{y}_\theta(x,c) = \sum_{k\in\mathcal{K}} k \cdot \pi_\theta(k \mid x, c)$$

This predictor improves Pearson & Spearman **for free** at inference time.

REAL — Objective & Method

REAL Objective

Replace binary reward with a **regression-aware** reward:

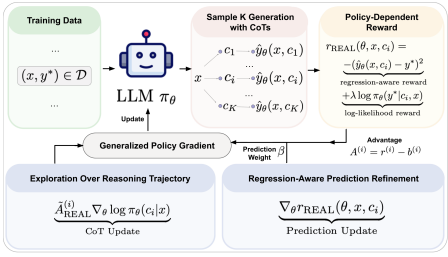
$$r_{\text{REAL}} = -(\hat{y}_\theta - y^*)^2 + \lambda \log \pi_\theta(y^*|x, c)$$

- Squared-error: predicting 4 when truth is 5 is *much better* than predicting 1.
- Log-likelihood: maintains next-token prediction ability.

The Challenge

Reward depends on θ (via $\hat{y}_\theta$ and π_θ), so $\nabla_\theta r \neq 0$—standard REINFORCE **is invalid**.

$\Rightarrow$ **Our solution: Generalized Policy Gradient**



Pipeline of REAL.

REAL — Theoretical Foundations

Lemma 3.1 Optimality of Squared Error for Pearson Correlation

Let $\hat{y}(x, c)$ be any predictor and $\mu(x, c) \triangleq \mathbb{E}[y^* \mid x, c]$.

- **Squared-error risk** $\mathcal{R}(\hat{y}) = \mathbb{E}[(\hat{y} - y^*)^2]$ is minimized by $\hat{y}^* = \mu(x, c)$.
- **Pearson correlation** $\rho(\hat{y}, y^*)$ is maximized by any positive affine transform of μ: $\hat{y} = a\mu + b,\ a > 0$.

$\Rightarrow$ **Minimizing squared error $\equiv$ maximizing Pearson correlation.**

Lemma 4.1 Generalized Policy Gradient with Policy-Dependent Rewards

Standard RL assumes $\nabla_\theta r = 0$. REAL's reward $r(\theta, x, c) = -(\hat{y}_\theta - y^*)^2 + \lambda \log \pi_\theta(y^* | x, c)$ depends on θ. The generalized gradient:

$$\nabla_\theta \mathcal{L} = \mathbb{E}_{c \sim \pi_\theta}\Big[\underbrace{r(\theta, x, c)\, \nabla_\theta \log \pi_\theta(c \mid x)}_{\text{Term 1: CoT Exploration}} + \underbrace{\nabla_\theta\, r(\theta, x, c)}_{\text{Term 2: Prediction Refinement}} \Big]$$

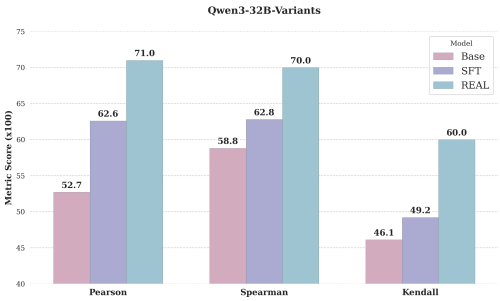
- **Term 1**: policy gradient $\times$ regression reward $\rightarrow$ explores reasoning trajectories.
- **Term 2**: backprop through $\hat{y}_\theta \rightarrow$ refines the numeric score.

$\Rightarrow$ **Jointly optimizes reasoning quality and numerical accuracy.**

REAL — Results & Takeaways

Empirical results (8B → 32B).

- Outperforms regression-aware SFT *and* standard RL.
- Strong *out-of-domain* generalization.
- Qwen3-32B: **+8.40** Pearson / **+7.20** Spearman over SFT; **+18.30** / **+11.20** over the base model.



Key Takeways

Aligning judges with the *structure of the reward* yields more *accurate* automatic evaluation.

PART III

Generating

How to learn generative models
when clean data is exhausted?

Score Distillation Beyond Acceleration:
Generative Modeling from Corrupted Data. ICLR 2026.

Flow Priors for Linear Inverse Problems via
Iterative Corrupted Trajectory Matching. NeurIPS 2024.

Restoration Score Distillation — Motivation

Clean data is running out

Frontier models already consume most high-quality data. Scaling further faces hard limits:

- Synthetic data risks *model collapse*.
- Scientific domains (MRI, astronomy) *never had* clean data.
- Real-world photos: inherently noisy, blurry, occluded.

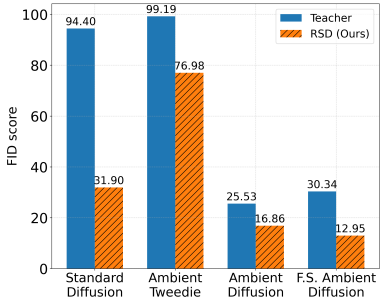
⇒ Can we learn generative models from corruption?

Conventional wisdom

Score distillation = *acceleration*: teacher → one-step student.

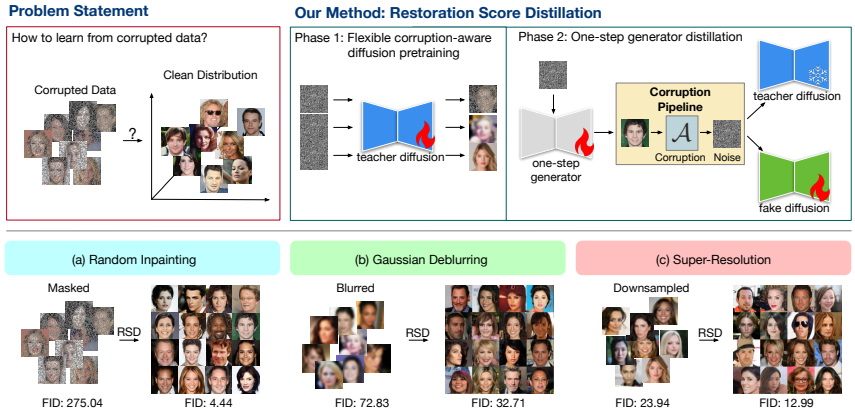
Our key insight

With corrupted data $y = \mathcal{A}(x) + \sigma\epsilon$, distillation goes *beyond* acceleration—student **restores** and **surpasses** teacher.



Student beats Teacher in FID with up to 30× speedup.

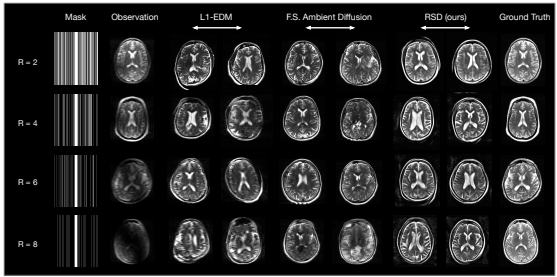
Restoration Score Distillation (RSD)



Unified framework for learning high-fidelity one-step generators from *degraded* data only.

1. Pretrain a **corruption-aware diffusion teacher** on observed measurements.
2. Distill into a **one-step generator** statistically closer to the *clean* distribution.

Restoration Score Distillation — Takeaways



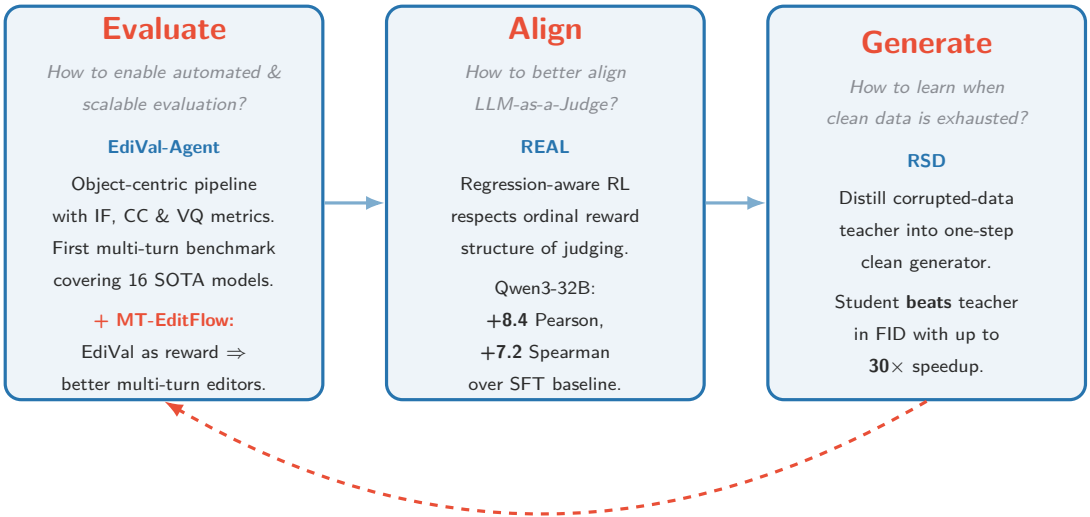
Key results

- Student surpasses teacher in FID with **one-step** sampling.
- Works across denoising, deblurring, inpainting, super-resolution, and **multi-coil MRI**.
- Up to **30**× inference speedup.

Broader impact

- Practical generative modeling where clean data is scarce (medical, scientific imaging).
- **Downstream:** learned clean prior powers *inverse problems—Flow Priors* (NeurIPS 2024).
- Distillation is not just compression—it is a *principled path to cleaner generation*.

A Virtuous Cycle — Answered



Thank You!

Questions are welcome.

References

1. Tianyu Chen*, **Yasi Zhang***, et al.
   EdiVal-Agent: An Object-Centric Framework for Automated, Fine-Grained Evaluation of Multi-Turn Editing. ICLR 2026.
   arxiv.org/abs/2509.13399

2. **Yasi Zhang***, Tianyu Chen*, Zhendong Wang, Ying Nian Wu, Mingyuan Zhou, Oscar Leong.
   Score Distillation Beyond Acceleration: Generative Modeling from Corrupted Data. ICLR 2026.
   arxiv.org/abs/2505.13377

3. **Yasi Zhang**, Peiyu Yu, Yaxuan Zhu, Yingshan Chang, Feng Gao, Ying Nian Wu, Oscar Leong.
   Flow Priors for Linear Inverse Problems via Iterative Corrupted Trajectory Matching. NeurIPS 2024.
   arxiv.org/abs/2405.18816

4. **Yasi Zhang***, Tianyu Chen*, Mingyuan Zhou, Oscar Leong, Ying Nian Wu, Michal Lukasik.
   REAL: Regression-Aware Reinforcement Learning for LLM-as-a-Judge. ICML 2026.
   arxiv.org/abs/2603.17145

5. **Yasi Zhang**, Peiyu Yu, Ying Nian Wu.
   Object-Conditioned Energy-Based Attention Map Alignment in Text-to-Image Diffusion Models. ECCV 2024.
   arxiv.org/abs/2404.07389

6. Jiahui Huang*, **Yasi Zhang***, et al.
   MT-EditFlow: Reinforcement Learning for Multi-Turn Image Editing with Flow Matching. In submission, 2026.

* Equal contribution