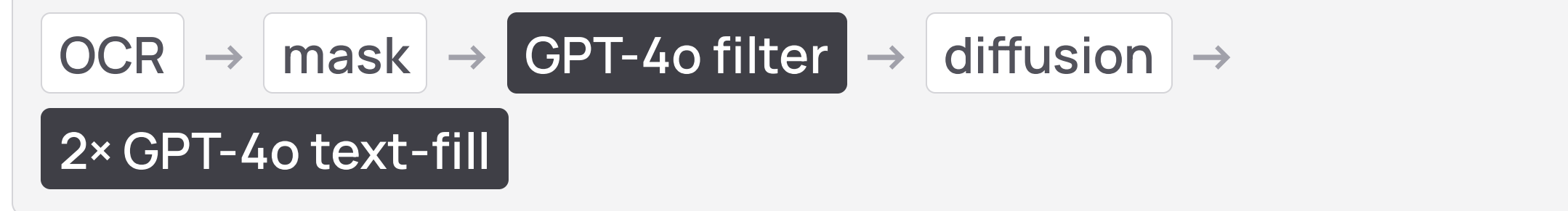




1 Motivation

Mobile-GUI **World Models** (WMs) predict future states to boost GUI agents at both **train-** and **inference-time**, but face a hard trade-off:

- > **Text WMs** abstract the state into text, discarding layout, icons, typography & color, and limiting strong VLM policies.
- > **Pixel WMs** (e.g. VIMO) need a slow, multi-model pipeline with closed weights:



Our goal: match this visual fidelity with a **single, open, fast** model. gWorld runs in **0.55 s** vs VIMO's **~160 s**.

2 Key Idea

Emulate the next state as **executable web code** that renders to pixels, not raw pixels.

One VLM keeps both strengths:

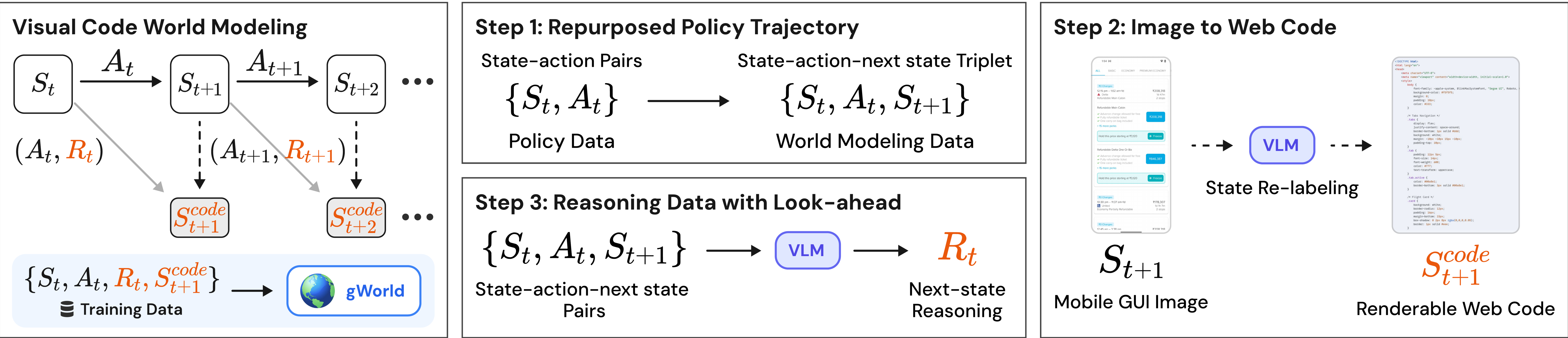
- > **Linguistic priors:** precise, legible text & realistic content.
- > **Web-code pre-training:** structurally accurate GUI layouts.
- > **Self-contained:** no OCR, masking, diffusion, or external VLM calls.

The reasoning trace R_t decomposes the task: first **describe** the state change in language, then **write the code** that renders it.

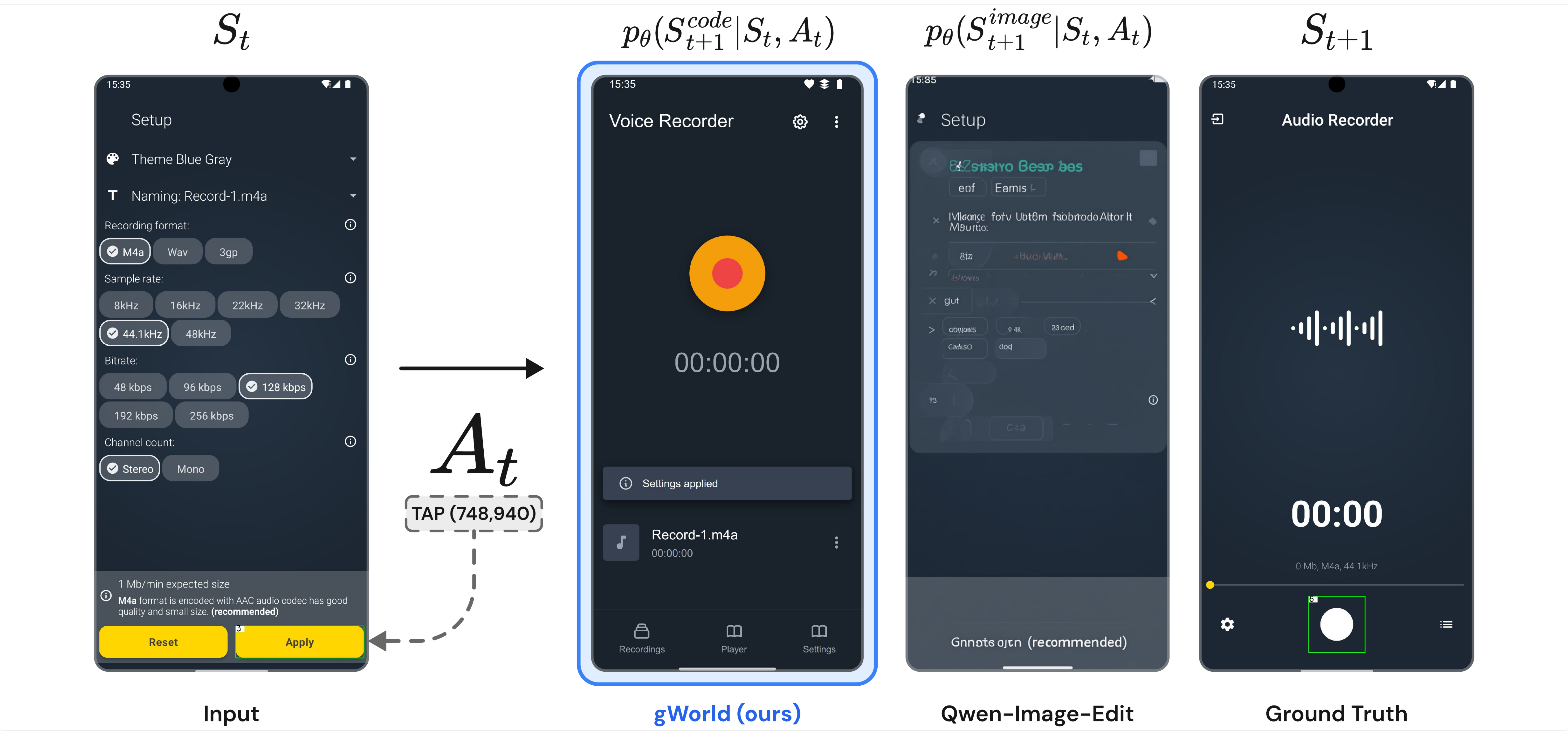
★ Contributions

- > **gWorld (8B, 32B)**: first open-weight, single-model visual mobile-GUI WMs via code generation.
- > An **automatic data-generation framework** repurposing policy trajectories into code + look-ahead reasoning.
- > **MWMBench**: a 6-benchmark suite (4 ID + 2 OOD) on the native visual, coordinate action space.
- > A **new Pareto frontier**, predictable **scaling laws**, and real **policy gains**.
- > A **fast, reproducible** alternative to multi-stage pixel pipelines, **291×** faster than VIMO.

3 Method: Automatic World Model Data Synthesis



4 Code vs. Pixel Generation

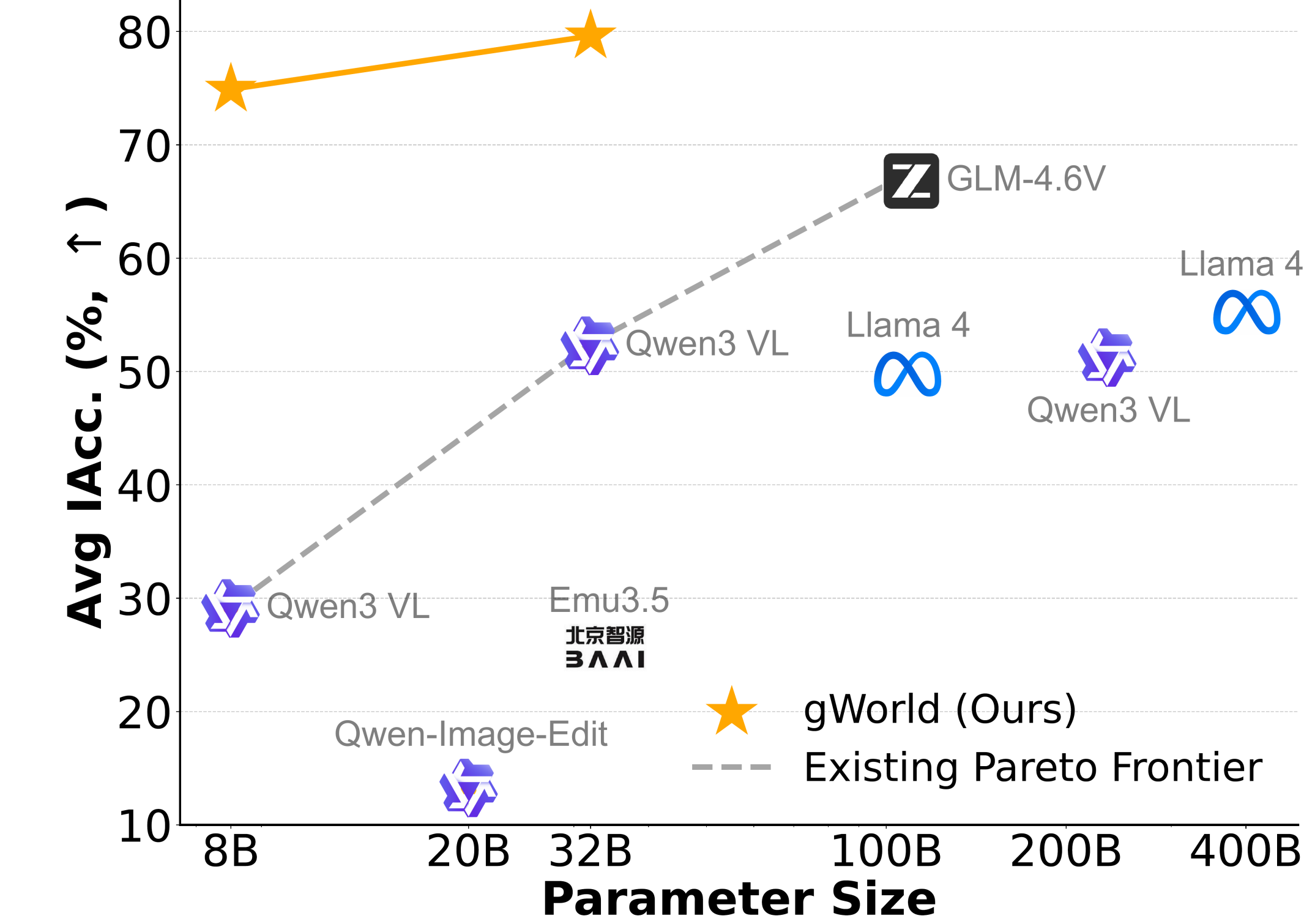


5 MWMBench: Comprehensive WM Benchmark

Benchmark	Modality	Actions	OOD
MobileWorldBench	Text	Converted to Text	No
VIMO's Benchmark	Visual	Converted to Text	No
MWMBench (ours)	Visual	Original Coord.	2 sets

4 ID: AitW · GUIO · AC · AMEX 2 OOD: AndroidWorld · KApps. **Metrics:** IAcc (3 frontier judges) & DINO similarity.

6 New Pareto Frontier



6 Main Results

Model	Size	IAcc (%) ↑	Fail (%) ↓
Qwen-I-E	20B	13.4	–
Emu3.5	34B	25.8	–
Qwen3 VL	8B	29.2	40.1
Llama 4	402B	55.7	9.2
GLM-4.6V	106B	67.4	2.5
gWorld	8B	74.9	1.4
gWorld	32B	79.6	0.6

Average over 6 benchmarks. **+45.7% / +27.1%** over base 8B / 32B.

50.25×
smaller than largest model it beats

<1%
render failure

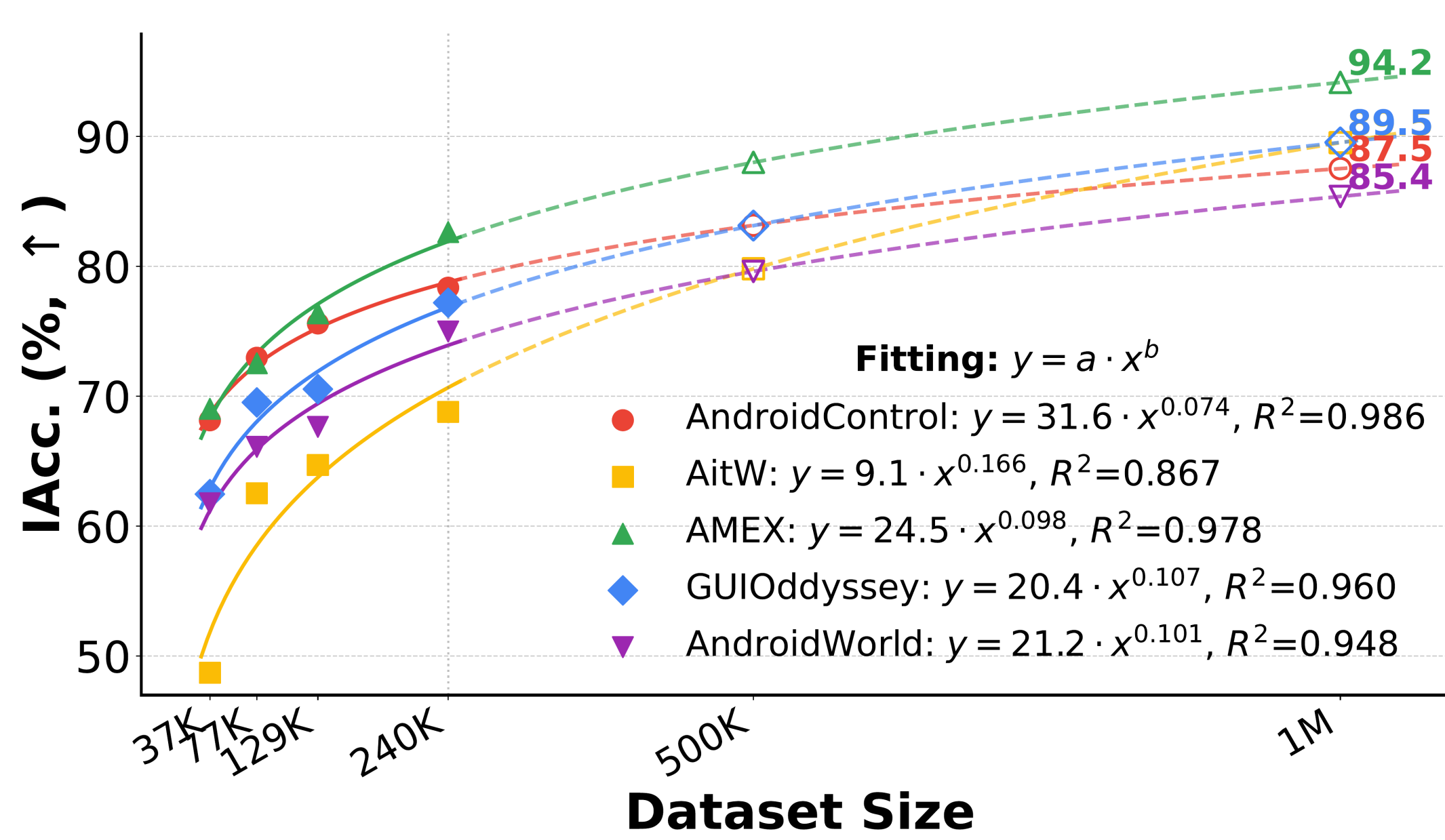
291×
faster than VIMO

#1/#2
human-pref rank

$R^2 \approx .95$
scaling-law fit

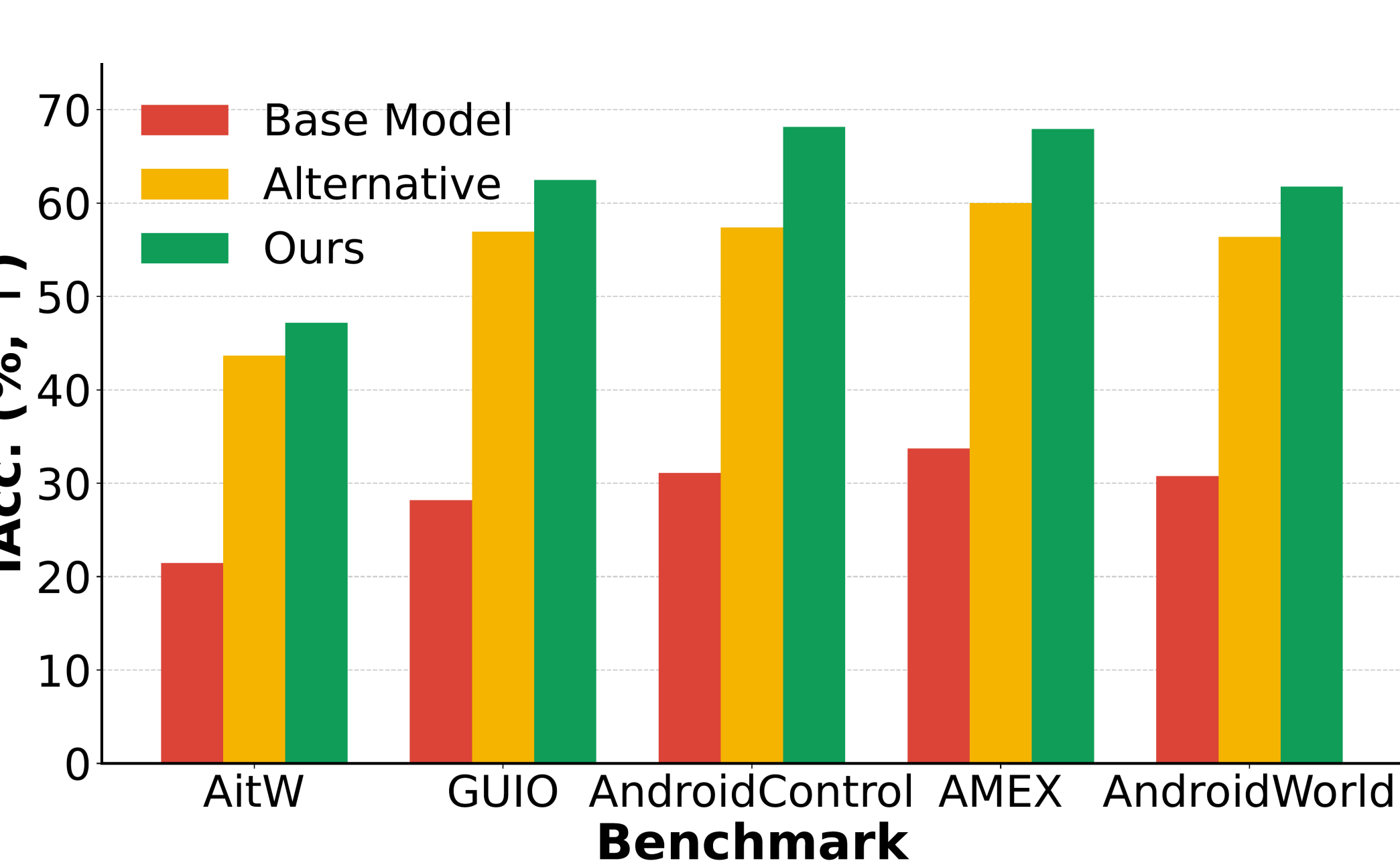
+22.4%
downstream policy

7 Data Scaling Laws



Power law ($R^2 \geq 0.94$), **non-saturating** → projects to **3.7M** transitions.

8 Ablations



Each component helps. Base → Alternative → Ours.

News

- > **gWorld** cited and adopted by **Qwen-AgentWorld**.
- > **gWorld** repo passes **60 GitHub stars**.

- > **gWorld** surpasses **9K total downloads** (including community quantized versions).

- > **gWorld** was on the front page of **HuggingFace Daily Papers** and **r/LocalLLaMA**.

- > **gWorld** featured in **over 10 news outlets**.