Video-Based Reward Modeling for Computer-Use Agents

Linxin Song[1], Jieyu Zhang[2], Huanxin Sheng[3], Taiwei Shi[1], Gupta Rahul[4], Yang Liu[4], Ranjay Krishna[2], Jian Kang[3], and Jieyu Zhao[1]

[1]University of Southern California [2]University of Washington
[3]MBZUAI [4]Amazon AGI

Abstract. Computer-use agents (CUAs) are becoming increasingly capable; however, it remains difficult to scale evaluation of whether a trajectory truly fulfills a user instruction. In this work, we study reward modeling from *execution video*: a sequence of keyframes from an agent trajectory that is independent of the agent's internal reasoning or actions. Although video-execution modeling is method-agnostic, it presents key challenges, including highly redundant layouts and subtle, localized cues that determine success. We introduce Execution Video Reward 53k (ExeVR-53k), a dataset of 53k high-quality video–task–reward triplets. We further propose adversarial instruction translation to synthesize negative samples with step-level annotations. To enable learning from long, high-resolution execution videos, we design spatiotemporal token pruning, which removes homogeneous regions and persistent tokens while preserving decisive UI changes. Building on these components, we fine-tune an Execution Video Reward Model (ExeVRM) that takes only a user instruction and a video-execution sequence to predict task success. Our ExeVRM 8B achieves 84.7% accuracy and 87.7% recall on video-execution assessment, outperforming strong proprietary models such as GPT-5.2 and Gemini-3 Pro across Ubuntu, macOS, Windows, and Android, while providing more precise temporal attribution. These results show that video-execution reward modeling can serve as a scalable, model-agnostic evaluator for CUAs.

Code: https://github.com/limenlp/ExeVRM
Model: https://huggingface.co/lime-nlp/ExeVRM-8B
Dataset: https://huggingface.co/datasets/lime-nlp/ExeVR-53K

1 Introduction

Computer-use agents (CUA) are becoming a promising paradigm for general-purpose computer-using automation [1,2,16,25,27,31,38,49,55,60,64,69,75,76]. By operating directly on real-world interfaces such as desktops, browsers, and mobile apps, these agents can execute complex, multi-step tasks in the same environments humans use. Advances in multimodal foundation models have accelerated this direction [3–5,7,17,18,22,44,52,59,62,72,79], enabling agents to

perceive visual layouts, interpret instructions, and interact across diverse platforms.

However, evaluating CUA remains challenging. Existing benchmarks often rely on handcrafted scripts or task-specific rules to verify completion [26, 41, 47], which limits scalability and transfer to new tasks and environments. A learned reward model is a more flexible alternative: given a user instruction and an execution trajectory, it judges whether the intended goal is achieved. To serve as a universal evaluator across heterogeneous CUA trajectories, the model should rely on a representation independent of internal reasoning/action formats (e.g., thoughts, tool calls, or code traces). We therefore focus on execution video, the observable sequence of interface states during interaction, which is method-agnostic and naturally comparable across different agent designs.

Building reward models from execution video introduces two key challenges. First, CUA trajectories are highly redundant: large interface regions (e.g., toolbars, backgrounds, and layout elements) remain nearly static across steps [34, 37, 70], while correctness often depends on subtle local changes such as cursor focus shifts, small text edits, or transient dialogs. The model must therefore suppress redundant content without missing decisive cues. Second, negative supervision is limited. Public computer-using datasets mainly emphasize successful or high-quality trajectories for agent training [39, 64, 70], and failure cases are rarely paired with explicit annotations of when and why trajectories diverge. This makes it difficult to build balanced, informative data for reward modeling.

To address these challenges, we construct a training dataset named Execution Video Reward 53k (ExeVR-53k), a large-scale corpus of execution video trajectories for reward modeling. We unify interaction data from multiple computer-using training datasets and agent rollouts by converting their logs into a consistent step-level video representation. Each trajectory is segmented into atomic interaction steps, and a representative key frame is extracted per step to capture the corresponding interface state. The resulting sequence forms a compact video summary that preserves the temporal progression of the task while remaining computationally manageable. To overcome the scarcity of negative supervision, we further introduce adversarial instruction translation, which synthesizes hard mismatched instruction–trajectory pairs within the same interface context. By generating plausible but semantically inconsistent instructions and identifying the step where the mismatch becomes evident, this process provides informative contrastive signals for training reward models.

To enable efficient learning from high–resolution, long–horizon execution videos, we propose a spatiotemporal token pruning strategy tailored to CUA execution video sequences. Spatial token pruning (STP) removes visually homogeneous regions that contribute little discriminative signal, such as large static backgrounds, while retaining localized UI elements that are more likely to encode task-relevant information. Temporal token pruning (TTP) further suppresses tokens that remain nearly unchanged across consecutive frames, allowing the model to focus on meaningful state transitions rather than repeated layout structure. Together, these mechanisms reduce redundancy while preserving subtle visual

cues that determine correctness, making video-based reward modeling tractable and robust for CUA environments.

With ExeVR-53k and the proposed spatiotemporal pruning strategy, we fine-tune an Execution Video Reward Model (ExeVRM) based on Qwen3-VL [72] to judge CUA task success directly from execution video demonstrations. The model takes as input a user instruction together with the corresponding execution video sequence and outputs a judgment of whether the trajectory satisfies the intended goal. By grounding its decision in the observed interface states over time, the model learns to associate subtle visual transitions with task success or failure. Experiments on ExeVR-Bench show that ExeVRM 8B reaches 84.7% accuracy and 87.7% recall, outperforming strong proprietary baselines (Seed-2.0 Pro [10]: 80.3/74.7; GPT-5.2 [43]: 75.0/66.5 in accuracy/recall) and open-weight models, while also achieving consistently better temporal attribution (higher tIoU) for localizing decisive error spans. Discussion and ablations further show that dense video context is critical compared with sparse screenshot-based evaluation, and that 720p inputs with STP+TTP improve completion judgment (notably recall) over 360p while maintaining tractable long-horizon training via favorable memory-latency trade-offs.

2 Related Work

2.1 Reward Evaluation for Visual GUI Agents

Large multimodal models have rapidly advanced GUI agents that execute natural-language instructions. Recent work spans web [20, 23, 32, 83], desktop [9, 39, 64, 70, 74], and mobile benchmarks [50], alongside specialized GUI perception/action models [25, 28, 37, 38, 40, 49, 60, 69, 76]. Yet reward evaluation remains a bottleneck.

Most existing protocols still depend on hand-crafted rules and environment-specific parsers [9,21,24,42,46,70]. Final-state checks are scalable but coarse [26, 41, 47]; full-screenshot evaluation improves coverage [13, 36] but is memory-intensive and often forces downsampling, which can hide subtle yet decisive GUI cues. Methodologically, prior work has explored Outcome Reward Models (ORMs) that score only end results [6, 48, 73]. Process Reward Models (PRMs) provide denser supervision [11, 12, 56, 57, 61, 65, 67, 68, 78, 80], but usually require $O(n)$ step-wise inference and are vulnerable to error accumulation or reward hacking [81]. Our ExeVRM instead performs holistic video-level judgment with first-failure localization, reducing step-wise tracing complexity; unlike GUI-critic-R1 and VAGEN, it relies only on external execution video, without agent-specific reasoning traces or tool calls [19, 66].

Negative supervision is another bottleneck because public computer-using data are dominated by successful trajectories [23, 70, 83]. Existing negative construction relies on passive failure collection [13, 41], expert annotation [12, 36], model annotation [48, 66], or rule-based corruption [67, 68]. In contrast, we pair successful trajectories with semantically mismatched instructions to generate

scalable hard negatives, while standardizing heterogeneous data sources into a unified representation [39, 64, 70].

2.2 Efficient Video Understanding and Token Pruning

Efficient token pruning is critical to enabling reward models to process GUI execution videos at scale. Prior methods for visual token pruning [8, 14, 15, 30, 51, 53, 54, 77] are mainly developed for natural or egocentric images or videos and often target action recognition or step localization. In GUI trajectories, however, critical evidence is frequently subtle and transient across adjacent key frames, so current methods might remove semantically important UI cues. Meanwhile, large static regions within each frame introduce substantial token redundancy that inflates memory cost with limited reward signal [29, 71]. However, methods tailored to GUI settings remain underexplored. Prior work study spatial-only dropout on single screenshot [37, 45], while we propose a more fine-grained spatiotemporal token pruning mechanism. Despite their spatiotemporal awareness, GUI-Pruner [71] and GUI-KV [29] are primarily designed for test-time memory saving during agent execution. In contrast, we adopt simpler spatial pruning and more robust temporal strategies to preserve tiny yet decision-critical visual evidence over long-horizon trajectories, which finally contributes to efficient reward modeling for outcome assessment and error attribution.

3 Execution Video Reward Modeling

3.1 ExeVR-53k

To address the data bottleneck in training reward models for computer use, we introduce Execution Video Reward 53k (ExeVR-53k), a training dataset of video trajectories curated from AgentNet [64], ScaleCUA [39], and OSWorld [70]. Existing resources are either limited in scale or lack diversity in agent behaviors; ExeVR-53k bridges this gap by combining human-supervised demonstrations (AgentNet and ScaleCUA) with solutions generated by 30 diverse computer-using agents (OSWorld), spanning end-to-end and agentic paradigms. We summarize the resulting data composition and key statistics in Table 1.

Fig. 1: Task distribution of ExeVR-53k.

We build our training corpus by leveraging existing computer-using datasets: AgentNet, ScaleCUA, and OSWorld. We convert their interaction records into

Table 1: Statistics of ExeVR-53k sources.

Dataset	Tasks	Trajectories	Trajectory Type	Accuracy
AgentNet [64]	23k	23k	Human	50% (w/ synthetic)
ScaleCUA [39]	7k	7k	Agent (human evaluation)	50% (w/ synthetic)
OSWorld [70]	361	23k	Agent (rule-based evaluation)	32%

a unified step-level video representation. Concretely, we segment each recorded trajectory into step units and retrieve a representative key frame (a screenshot after an action) for each step to capture the corresponding UI state. We then concatenate these key frames in temporal order to form a compact video summary rendered at 1 FPS, coarse-grained progression of the interaction while keeping the input length manageable. Figure 1 summarizes the task composition of ExeVR-53k across sources, highlighting the diversity of operating environments and task categories covered in our collected trajectories.

OSWorld [70] focuses on realistic computer-using evaluation over open-domain web and desktop applications, including OS file I/O, multi-application workflows, and system-level operations. It contains 369 tasks (with 361 commonly evaluated and 8 optional Google Drive tasks). Unlike demonstration datasets, OSWorld does not provide fixed ground-truth trajectories; instead, each task is specified by environment initialization plus executable evaluation scripts, and computer-using agents generate trajectories during evaluation. In our pipeline, we treat rollouts from 30 different computer-using agents [1–5, 7, 16, 18, 25, 27, 31, 38, 44, 49, 52, 59, 60, 64, 69, 75, 76] on 361 tasks as trajectories and apply the same step segmentation and key-frame summarization procedure to obtain training-ready video sequences.

AgentNet [64] provides human demonstration trajectories spanning real-world desktop and web workflows. It covers four main domains with 11 finer-grained subdomains, including e-commerce, news & entertainment, lifestyle, social media and communication, office tools, task management and collaboration, creative design and multimedia, development and engineering, knowledge discovery and research, data analysis, web tools and internet utilities, and operating systems and utilities. We use 22,625 human-labeled tasks from AgentNet. The trajectories are collected across multiple operating systems, including Windows (12K tasks), macOS (5K tasks), and Ubuntu (5K tasks), and are recorded as step-by-step human demonstrations.

ScaleCUA [39] is a large-scale, GUI-centric dataset designed to support grounding and end-to-end computer-using learning across a diverse set of platforms. The paper defines three high-level task domains centered on GUI/action grounding, navigation and interaction, and trajectory-level task execution. The dataset additionally contains multiple interaction types, although exact per-domain statistics are not fully disclosed publicly in the dataset card. ScaleCUA spans Linux,

macOS, Windows, Android, and the Web. Its data is produced via a hybrid pipeline: grounding examples are collected largely automatically and annotated with LLM assistance followed by human verification, while trajectory data relies on human interaction logs augmented with LLM-generated annotations.

3.2 Adversarial Instruction Translation

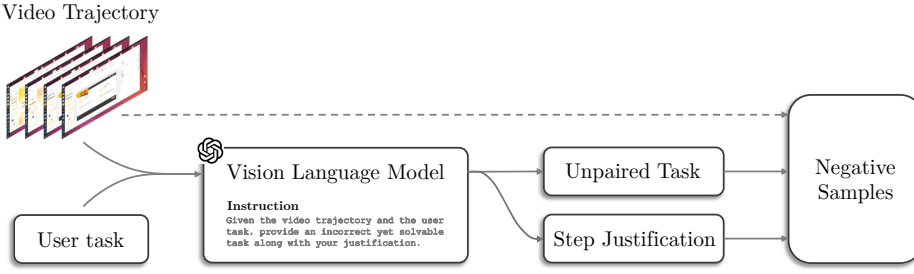

Fig. 2: Illustration of how we synthesize negative samples via adversarial instruction translation. We use GPT-5.2 as the Vision Language Model.

AgentNet and ScaleCUA predominantly provide positive supervision because they are collected as successful demonstrations for training GUI agents. To obtain informative counterexamples for reward modeling, we introduce a simple yet effective synthetic-negative construction procedure, which we term *adversarial translation*, inspired by back-translation [35]. As illustrated in Figure 2, we start from a valid trajectory segment (a short sequence of UI states and actions) and prompt a vision–language model to generate an *unpaired* task instruction that is plausible in the same interface context but does not match the demonstrated segment. During this translation process, the model is required to output both (i) a brief justification of why the segment violates the generated instruction and (ii) a reference step (i.e., the time index where the mismatch first becomes evident); we use these justifications as attribution labels for temporal grounding. This yields hard negatives that are visually similar yet semantically inconsistent with the intended goal. We then manually verify these synthetic pairs and retain only high-quality negatives; in our annotation audit, the resulting negatives achieve a 100% human pass rate on a selected subset, providing reliable contrastive supervision for training our reward model.

3.3 Spatiotemporal Token Pruning

Computer-use trajectory videos contain a large number of fine-grained elements at different scales, such as windows, icons, menus, cursors, and small text. Training reward models for computer use often benefits from high-resolution video inputs, where these subtle UI cues are clearly visible. However, the computational cost of processing high-resolution frames is substantial: naively placing

Algorithm 1 Training w/ Spatiotemporal Token Pruning

Require: Training dataset $\mathcal{D}_{\text{train}}$, vision–language model $\mathcal{M}$, STP threshold τ_s, TTP threshold τ_t, component threshold τ_{large}

Freeze vision encoder $\mathcal{V}$ and projector $\mathcal{P}$ ▷ Only train LLM parameters
for $(X, V, Y) \sim \mathcal{D}_{\text{train}}$ **do** ▷ X: text tokens, V: video frames, Y: labels
  // **Token Pruning**
  $Z_V \leftarrow \mathcal{V}(V)$ ▷ Encode video frames into patch tokens
  $\mathbf{M}_s \leftarrow \text{STP}(Z_V, \tau_s, \tau_{\text{large}})$ ▷ Algorithm 2
  $\mathbf{M}_t \leftarrow \text{TTP}(Z_V, \tau_t)$ ▷ Algorithm 3
  $\mathbf{M} \leftarrow \mathbf{M}_s \wedge \mathbf{M}_t$ ▷ Token kept only if both agree
  // **Apply Masks to Video Tokens**
  $\widetilde{Z}_V \leftarrow \text{Pack}(Z_V, \mathbf{M})$ ▷ Drop pruned tokens and re-pack sequence
  $\widetilde{H}_V \leftarrow \mathcal{P}(\widetilde{Z}_V)$ ▷ Project visual tokens to LLM input space
  // **Update Model**
  $\hat{Y} \leftarrow \mathcal{M}_{\text{LLM}}(\widetilde{H}_V, X)$
  $\mathcal{L}_{\text{rm}} \leftarrow \ell(\hat{Y}, Y)$
  Update trainable parameters with $\nabla \mathcal{L}_{\text{rm}}$
end for

all frames at full resolution into the model context quickly exhausts the token budget and leads to prohibitive memory usage, making it difficult to fit long interaction trajectories into a single forward-backward pass. To address this issue, we propose *Spatiotemporal Token Pruning*, which can significantly reduce the effective context length by pruning redundant visual tokens in both space and time. As summarized in Algorithm 1, we combine Spatial Token Pruning (STP; section 3.3) to discard spatially redundant background regions within each frame, and Temporal Token Pruning (TTP; section 3.3) to suppress temporally repeated tokens across frames, enabling efficient training with high-resolution videos without sacrificing the transient UI evidence critical for decision making.

Spatial Token Pruning Let $\mathbf{P} \in \mathbb{R}^{N \times D}$ denote the patch-feature matrix produced by the frozen vision encoder, where $N = T \times H' \times W'$ is the total number of spatiotemporal patches across T frames on an $H' \times W'$ grid, and D is the feature dimension. Inspired by ShowUI [37], we construct a per-frame UI-connected graph $\mathcal{G}^{(t)} = (V^{(t)}, E^{(t)})$, where each node in $V^{(t)}$ corresponds to a patch and edges in $E^{(t)}$ connect spatial neighbors with similar features. The resulting connected components are denoted by $\mathcal{C}^{(t)} = \{C_1^{(t)}, \ldots, C_{K_t}^{(t)}\}$. We use τ_s as the similarity threshold for establishing edges, and τ_{large} as the size threshold for identifying components to be removed. For each frame t, STP outputs a binary mask $\mathbf{M}_s^{(t)} \in \{0,1\}^{H' \times W'}$ indicating which patches are retained.

For a single frame t, we reshape patch features into a grid $\mathbf{P}^{(t)} \in \mathbb{R}^{H' \times W' \times D}$, where $H' = H/m$ and $W' = W/m$ for merge size m. We first compute local

Algorithm 2 Spatial Token Pruning (STP)

Require: Patch features $\mathbf{P}$, adjacency threshold τ_s, large-component threshold τ_{large}
Reshape $\mathbf{P} \in \mathbb{R}^{N\times D}$ into per-frame grids $\mathbf{P}^{(t)} \in \mathbb{R}^{H'\times W'\times D}$, $t \in [T]$
for $t \leftarrow 1$ **to** T. **do**
  Compute neighbor distances $d_h^{(t)}(i,j)$ and $d_v^{(t)}(i,j)$ from $\mathbf{P}^{(t)}$.
  Define adjacency indicator $e^{(t)}_{(i,j)\leftrightarrow(i',j')}$ using τ_s.
  Build graph $\mathcal{G}^{(t)} = (V^{(t)}, E^{(t)})$ from $e_{(i,j)\leftrightarrow(i',j')}$.
  Obtain connected components $\mathcal{C}^{(t)} = \{C_k^{(t)}\}$.
  Mark large components $\mathcal{R}^{(t)} \subseteq \mathcal{C}^{(t)}$ using τ_{large}.
  Construct spatial mask $\mathbf{M}_s^{(t)} \in \{0,1\}^{H'\times W'}$ from $\mathcal{R}^{(t)}$.
end for
return $\{\mathbf{M}_s^{(t)}\}_{t=1}^{T}$

feature differences between neighboring patches. The horizontal difference is

$$d_h^{(t)}(i,j) = \left\| \mathbf{P}_{i,j}^{(t)} - \mathbf{P}_{i,j+1}^{(t)} \right\|_2, \quad 1 \le i \le H',\ 1 \le j < W', \tag{1}$$

while the vertical difference is

$$d_v^{(t)}(i,j) = \left\| \mathbf{P}_{i,j}^{(t)} - \mathbf{P}_{i+1,j}^{(t)} \right\|_2, \quad 1 \le i < H',\ 1 \le j \le W'. \tag{2}$$

Let $\mathcal{N}(i,j)$ be the 4-neighborhood of location (i,j). Two patches are connected if they are neighbors and their feature distance is below τ_s:

$$e^{(t)}_{(i,j)\leftrightarrow(i',j')} = \mathbb{1}\left[(i',j') \in \mathcal{N}(i,j) \ \wedge\ \left\| \mathbf{P}_{i,j}^{(t)} - \mathbf{P}_{i',j'}^{(t)} \right\|_2 < \tau_s \right]. \tag{3}$$

We then apply Union-Find on $\mathcal{G}^{(t)}$ to obtain connected components and select large components for pruning:

$$\mathcal{C}^{(t)} = \text{Union-Find}\left(V^{(t)}, E^{(t)}\right), \quad \mathcal{R}^{(t)} = \left\{ C \in \mathcal{C}^{(t)} : |C| > \tau_{\text{large}} \right\}. \tag{4}$$

Let $C^{(t)}(i,j)$ denote the component containing patch (i,j). The spatial mask is

$$\mathbf{M}_s^{(t)}(i,j) = \begin{cases} 0 & \text{if } C^{(t)}(i,j) \in \mathcal{R}^{(t)}, \\ 1 & \text{otherwise.} \end{cases} \tag{5}$$

Applying $\mathbf{M}_s^{(t)}$ filters the original token grid and packs the remaining tokens into a shorter sequence, which is then fed to the vision-language model. This parameter-free procedure preserves local UI structure and effectively reduces redundant spatial tokens while retaining fine-grained cues.

Temporal Token Pruning In UI-dense videos, spatial redundancy alone is often insufficient: many frames share the same layout (menus, sidebars, toolbars, and background windows), while task-relevant evidence is concentrated in

Algorithm 3 Temporal Token Pruning (TTP)

Require: Video tokens $\mathbf{V} \in \mathbb{R}^{T\times N\times D}$, threshold τ_t, similarity $\mathrm{sim}_{\cos}(\cdot,\cdot)$
for $i = 0$ **to** $N-1$ **do**
  $\mathbf{v}_i^{(\mathrm{ref})} \leftarrow \mathbf{v}_i^{(0)}$ ▷ Frame 0 is the initial reference frame
  $\mathbf{M}_t(0,i) \leftarrow 1$ ▷ Always keep reference frame
  for $t = 1$ **to** $T-1$ **do**
    if $\mathrm{sim}_{\cos}(\mathbf{v}_i^{(\mathrm{ref})}, \mathbf{v}_i^{(t)}) > \tau_t$ **then**
      $\mathbf{M}_t(t,i) \leftarrow 0$ ▷ Similar to reference, remove
    else
      $\mathbf{M}_t(t,i) \leftarrow 1$ ▷ Different from reference, keep
      $\mathbf{v}_i^{(\mathrm{ref})} \leftarrow \mathbf{v}_i^{(t)}$ ▷ Update reference
    end if
  end for
end for
return $\mathbf{M}_t$

small, transient changes (e.g., a highlighted button, a newly opened dropdown, a scrolling offset, or a cursor-driven text update). In such cases, STP may still retain many tokens per frame because the UI contains high-contrast elements everywhere, even though most of them remain unchanged across adjacent frames. Moreover, STP can be unreliable in the presence of visually significant yet semantically irrelevant content (e.g., image-dense webpages). Therefore, we further propose temporal token pruning (TTP), which complements STP by removing temporally repeated tokens and reallocating the token budget to regions that actually evolve over time.

We denote the video token tensor by $\mathbf{V} \in \mathbb{R}^{T\times N\times D}$, where T is the number of frames, N is the number of (merged) spatial tokens per frame, and D is the token dimension. Let $\mathbf{v}_i^{(t)} \in \mathbb{R}^D$ be the token at spatial location i in frame t. Analogous to STP, our goal is to identify and remove tokens that contribute little new information under the UI-centric computer-using setting. While STP suppresses spatially homogeneous regions within each frame, TTP targets redundancy across time by filtering tokens whose appearance remains nearly unchanged over consecutive frames.

Specifically, for each spatial location i, we maintain a reference token $\mathbf{v}_i^{(\mathrm{ref})}$ initialized from the first frame and compare subsequent tokens against this reference using cosine similarity $\mathrm{sim}_{\cos}(\cdot,\cdot)$. We set the temporal mask as

$$\mathbf{M}_t(t,i) = \mathbb{1}\left[\mathrm{sim}_{\cos}\left(\mathbf{v}_i^{(\mathrm{ref})}, \mathbf{v}_i^{(t)}\right) \le \tau_t\right], \quad t \in \{1,\dots,T-1\}. \tag{6}$$

We always keep the first-frame token by setting $\mathbf{M}_t(0,i) = 1$. Equivalently, a token is pruned when $\mathrm{sim}_{\cos}(\mathbf{v}_i^{(\mathrm{ref})}, \mathbf{v}_i^{(t)}) > \tau_t$. The reference update follows

$$\mathbf{v}_i^{(\mathrm{ref})} \leftarrow \begin{cases} \mathbf{v}_i^{(t)} & \text{if } \mathbf{M}_t(t,i) = 1, \\ \mathbf{v}_i^{(\mathrm{ref})} & \text{otherwise.} \end{cases} \tag{7}$$

where τ_t is the similarity threshold. This update rule ensures that each location is compared against its most recent "distinct" state, effectively capturing step-wise UI transitions (cursor moves, menu expansion, scrolling, and window switching) while ignoring static backgrounds. Finally, Algorithm 3 constructs a binary temporal mask $\mathbf{M}_t \in \{0,1\}^{T\times N}$ which can be applied to the token sequence to drop temporally redundant tokens.

4 Experiments

4.1 Implementation Details

Our approach introduces three key hyperparameters controlling token filtering: the spatial threshold τ_s for STP, the temporal similarity threshold τ_t for TTP, and a "large-component" threshold τ_{large} used to guard against over-pruning when abrupt UI transitions occur. Intuitively, decreasing τ_s makes STP more aggressive by removing a larger portion of spatially low-saliency tokens, while increasing τ_s retains more within-frame context at the cost of a higher token budget. Similarly, increasing τ_t makes TTP stricter (i.e., only near-identical tokens are removed), whereas lowering τ_t prunes more temporal repeats but risks discarding subtle yet task-relevant UI updates. We set $\tau_s = 0.3$, $\tau_t = 0.9999$, and $\tau_{\text{large}} = 40$ for all experiments. We instantiate our video reward model from `Qwen3-VL-4B-Instruct` and `Qwen3-VL-8B-Instruct`, and fine-tune with a learning rate of 5×10^{-6} using a cosine decay schedule. All training runs are performed on 8×NVIDIA A100 GPUs with 80 GB memory each. Our training is based on a modified LLaMA-Factory [82] framework.

4.2 Evaluation Protocol

ExeVR-Bench We build our evaluation benchmark, ExeVR-Bench, from a held-out split of ExeVR-53k. Each instance pairs a *user instruction* with a *video trajectory*. We evaluate two settings: (i) binary judgment (`correct` vs. `incorrect`), and (ii) *attribution judgment*, which additionally requires a short time range indicating where the first error occurs. The initial pool contains 800 tasks, evenly split across Ubuntu (Agent), Ubuntu (Human), Mac/Win, and Android. For Ubuntu (Agent), we sample 10 tasks from OSWorld and collect 200 distinct solutions to capture trajectory diversity under shared goals; for Ubuntu (Human), Mac/Win, and Android, we separate 200 unique task–solution pairs per category from ExeVR-53k to broaden task coverage. We also annotate 200 instances with first-deviation time ranges for fine-grained temporal localization evaluation. For consistent visual input, all videos are rendered at 720p, and models receive up to 100 frames at 1 FPS (uniformly sampled when longer). After removing unsolvable tasks, ExeVR-Bench contains 789 instances, with an approximately balanced class ratio (49.94% positive vs. 50.06% negative).

Table 2: Detailed performance on Agent/Human trajectory evaluation on ExeVR-Bench. We report Accuracy (Acc), Precision (Prec), and Recall (Rec). Underline indicates the second-best result, and **bold** indicates the best result. ExeVRM 8B achieves the best overall balance, outperforming open-source baselines and matching or exceeding strong proprietary models on most settings.

Model	Ubuntu (Agent)			Ubuntu (Human)			Mac/Win			Android			Overall		
	Acc.	Prec.	Rec.	Acc.	Prec.	Rec.	Acc.	Prec.	Rec.	Acc.	Prec.	Rec.	Acc.	Prec.	Rec.
Proprietary Models															
Gemini 2.5 Pro [18]	84.6	85.6	82.8	64.5	82.2	37.0	69.0	83.9	47.0	74.0	81.6	62.0	72.8	83.5	56.7
Gemini 3 Pro [58]	80.4	75.0	**90.0**	71.2	71.7	71.0	75.6	75.3	76.8	73.5	74.7	71.0	75.1	74.2	76.7
GPT-5.2 [43]	82.5	85.1	78.7	74.0	**84.3**	59.0	74.5	**88.9**	68.7	74.5	75.5	75.8	75.0	82.7	66.5
Seed-2.0 Pro [10]	**85.1**	85.1	85.1	77.2	81.7	69.1	81.0	86.1	74.0	78.0	**82.6**	71.0	80.3	**83.9**	74.7
Open-sourced Models															
LLaVA-Next-Video 7B [33]	50.8	51.6	17.0	48.5	46.7	21.0	49.5	48.8	21.0	49.0	47.5	19.0	13.6	48.4	19.4
InternVL-3.5 8B [63]	64.6	64.8	55.9	52.0	52.2	48.0	59.5	58.6	65.0	57.5	59.3	48.0	56.6	58.9	55.9
Qwen2.5-VL 7B [7]	64.6	65.5	60.6	63.0	65.1	56.0	68.0	74.3	55.0	63.5	62.7	66.0	64.5	66.6	59.2
Qwen3-VL 4B [72]	76.2	72.9	82.9	72.5	76.5	65.0	74.5	74.3	75.0	68.5	70.8	63.0	72.9	73.7	71.3
Qwen3-VL 8B [72]	71.4	74.4	64.9	69.0	86.5	45.0	66.5	81.1	43.0	64.5	71.6	48.0	67.7	77.6	49.9
ExeVRM 4B	77.8	83.3	69.1	81.0	82.9	78.0	**90.0**	87.7	93.0	72.5	66.4	91.0	80.1	79.2	82.5
ExeVRM 8B	82.5	**85.9**	77.7	**84.0**	84.0	**84.0**	89.0	85.5	**94.0**	**83.5**	77.2	**95.0**	**84.7**	82.9	**87.7**

Evaluation metrics We report standard classification metrics, including accuracy, precision, and recall, to measure how well the reward model distinguishes positive from negative trajectories. In addition, to evaluate temporal grounding quality, *i.e.*, whether the model localizes the critical time span responsible for failure, we compute a temporal intersection-over-union (tIoU) between the model-predicted interval $\hat{\mathcal{I}} = [\hat{t}_s, \hat{t}_e]$ and the ground-truth interval $\mathcal{I} = [t_s, t_e]$:

$$\text{tIoU}(\hat{\mathcal{I}}, \mathcal{I}) = \frac{|\hat{\mathcal{I}} \cap \mathcal{I}|}{|\hat{\mathcal{I}} \cup \mathcal{I}|} = \frac{\max\big(0,\, \min(\hat{t}_e, t_e) - \max(\hat{t}_s, t_s)\big)}{\max(\hat{t}_e, t_e) - \min(\hat{t}_s, t_s)}. \tag{8}$$

This metric rewards accurate localization and penalizes overly broad predictions.

4.3 Main Result

Performance comparison Consistent with Table 2, ExeVRM 8B ranks first overall with **84.7** accuracy, **82.9** precision, and **87.7** recall, surpassing the strongest proprietary baselines Seed-2.0 Pro (80.3/83.9/74.7) and GPT-5.2 (75.0/82.7/66.5). It also outperforms the best open-weight models by +17.1 accuracy points over Qwen3-VL 8B and +28.2 over InternVL-3.5 8B, while substantially improving recall (87.7 vs. 49.9 and 55.9). Gains are consistent across settings: Mac/Win reaches 89.0 accuracy with **94.0** recall, Android reaches **83.5** accuracy with **95.0** recall, and Ubuntu reaches 82.5/77.7 (Agent) and **84.0**/**84.0** (Human) for accuracy/recall, all leading their baselines. Scaling from 4B to 8B further brings +4.6 overall accuracy (84.7 vs. 80.1) and +5.2 recall (87.7 vs. 82.5), with the largest gains on Android (+11.0) and Ubuntu (Agent, +4.8).

Attribution comparison Beyond overall accuracy, we analyze attribution quality: *how well a model can localize the decisive temporal window that determines whether a trajectory succeeds or fails.* We compute temporal IoU (tIoU) between each model's predicted salient segment and human-annotated evidence spans on ExeVR-Bench. As shown in Figure 3, ExeVRM attains consistently higher tIoU than all baselines, indicating more precise and less noisy credit assignment over time. This suggests that explicitly training a reward model on computer-using trajectories not only improves preference prediction, but also yields explanations that better align with human judgments about when the outcome is decided. In practice, stronger attribution makes downstream agent development easier: it highlights the exact interaction steps that cause failures (e.g., an incorrect click or a missed UI state change), enabling faster debugging and more targeted data collection.

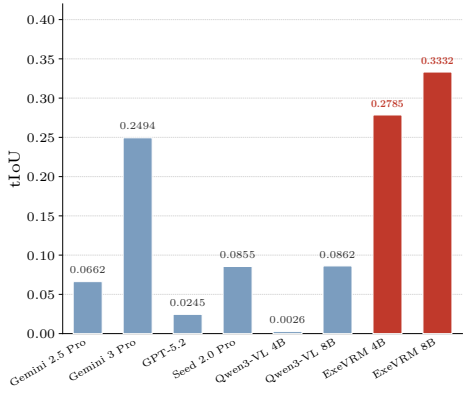

Fig. 3: Comparison of temporal IoU (tIoU) scores across models on ExeVR-Bench.

Table 3: Performance comparison across different evaluation strategies. Here Tail (T) means the last screenshot and Head (H) means the first screenshot. Compact denotes a compact list of action history, while Details denotes a structured action history. Oneshot means there is an example in the prompt as a reference.

Model	Method	Res.	Visual	Info.	Ubuntu(Human)			Mac/Win		
					Acc.	Prec.	Rec.	Acc.	Prec.	Rec.
Qwen3-VL 4B	AER [47]	720p	Tail	Compact	58.5	53.3	70.3	69.0	84.0	72.9
	Simplified Judge [41]	720p	H+T	Details	57.5	52.1	**81.3**	70.8	87.1	71.5
	SE-WSM [57]	360p	Full	None	50.5	44.6	36.3	46.0	84.1	35.1
	ZeroGUI [73]	360p	Full	Oneshot	48.0	42.5	40.7	51.5	81.4	46.4
	ExeVRM 4B (Ours)	720p	Video	None	**81.0**	**82.9**	78.0	**90.0**	**87.7**	**93.0**
Qwen3-VL 8B	AER [47]	720p	Tail	Compact	55.0	50.5	61.5	68.0	80.4	76.2
	Simplified Judge [41]	720p	H+T	Details	58.5	53.5	68.1	78.2	**89.6**	80.1
	SE-WSM [57]	360p	Full	None	48.5	43.2	41.8	51.0	81.5	46.4
	ZeroGUI [73]	360p	Full	Oneshot	50.0	46.7	61.5	63.5	81.5	66.9
	ExeVRM 8B (Ours)	720p	Video	None	**84.0**	**84.0**	**84.0**	**89.0**	85.5	**94.0**

5 Discussion & Ablation Studies

Findings 1: Dense video context outperforms sparse snapshots Previous work on task-completion assessment in CUA largely relies on screenshots rather than execution video. In particular, AER [47] evaluates only the final-state

Table 4: Effect of input resolution on reward-model performance. Increasing the input from 360p to 720p with spatiotemporal token pruning (STP&TTP) consistently improves accuracy and recall for both Qwen3-VL 4B and 8B.

Model	Resolution	Accuracy	Precision	Recall
Qwen3-VL 4B	360p	79.3	**80.6**	77.8
	720p (w/STP & TTP)	**80.1**	79.2	**82.5**
Qwen3-VL 8B	360p	81.5	82.5	80.5
	720p (w/STP & TTP)	**84.7**	**82.9**	**87.7**

Table 5: Ablation of spatiotemporal token pruning on Qwen3-VL 4B. STP denotes spatial token pruning and TTP denotes temporal token pruning. TTP contributes stronger gains than STP when used alone, while the full STP+TTP configuration remains competitive and achieves the highest recall.

Model	w/STP	w/TTP	Accuracy	Precision	Recall
Qwen3-VL 4B	✓		77.6	81.7	72.6
		✓	80.3	81.3	79.3
	✓	✓	80.1	79.2	82.5

screenshot, while Simplified Judge [41] add initial screenshot into considerations. However, these sparse-observation methods underperform (Table 3), indicating that ignoring causal transitions throughout the interaction makes accurate completion judgment difficult. SE-WSM [57] and ZeroGUI [73] instead feed every key frame to the model, which leads to OOM on a single A100 (80GB) without token pruning. If we reduce the resolution to 360p to meet the constraint, performance can even fall below AER [47] that only judges final state. This finding highlights the necessity of both video-based assessment and token pruning.

Findings 2: Higher resolution brings more benefit for reward modeling As shown in Table 4, moving from 360p to 720p yields clear gains in completion judgment quality, especially in recall. For Qwen3-VL 4B, accuracy improves from 79.3 to 80.1 and recall rises from 77.8 to 82.5 (+4.7), with a modest precision drop (80.6→79.2). For Qwen3-VL 8B, all three metrics improve: accuracy increases from 81.5 to 84.7 (+3.2), precision from 82.5 to 82.9, and recall from 80.5 to 87.7 (+7.2). These results indicate that higher-resolution inputs preserve fine-grained GUI cues (e.g., small text edits and localized state changes) that are frequently decisive for reward prediction, while STP/TTP keeps 720p training and inference tractable.

Findings 3: Asymmetric Effects of Spatial and Temporal Pruning Table 5 reveals an asymmetric contribution of the two pruning modules. Applying STP alone yields lower accuracy and recall (77.9/72.6), whereas TTP alone provides the strongest overall balance (80.3/79.3). We conjecture this trend follows

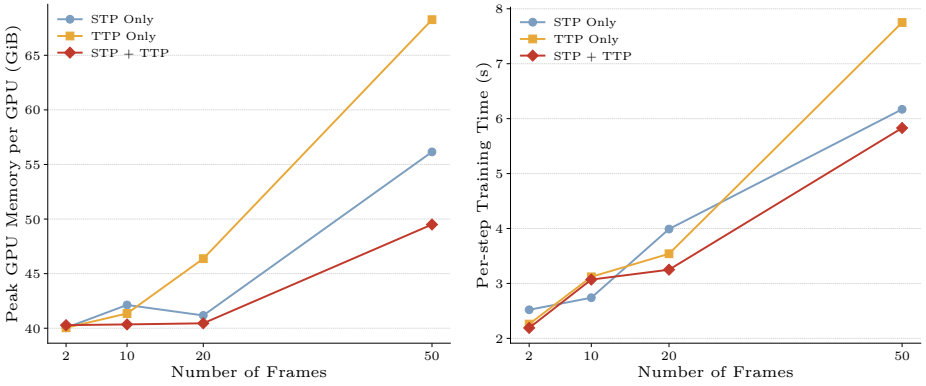

Fig. 4: Efficiency analysis. Left: memory usage. Right: runtime.

from the nature of GUI trajectories: within-frame saliency is often dominated by visually rich yet task-irrelevant regions (e.g., dense menus, icons, and textured webpage content), so spatial pruning may occasionally discard fine-grained cues that are weak visually but decisive for completion judgment. By contrast, reward prediction is primarily driven by inter-frame state transitions, which are directly targeted by TTP through suppressing temporally repeated tokens and preserving changing evidence. Notably, enabling STP and TTP jointly causes only a marginal accuracy change relative to TTP-only (80.1 vs. 80.3) while improving recall to 82.5, suggesting that the full configuration remains robust and mainly induces a precision–recall trade-off rather than a substantial performance drop.

Findings 4: Spatiotemporal pruning improves training efficiency As shown in Figure 4, both peak GPU memory and per-step training time increase with the number of input frames for all variants, but the growth rates differ substantially. With 50 frames, STP-only reaches about 56 GiB memory and 6.2 s/step, while TTP-only rises further to about 68 GiB and 7.8 s/step. In contrast, combining STP and TTP keeps the footprint much lower (∼49.5 GiB and 5.8 s/step), and remains close to the 40 GiB regime up to 20 frames. This trend is consistent with our method design: spatial pruning removes large homogeneous regions in each frame, whereas temporal pruning suppresses repeated tokens across adjacent frames. Jointly applying both reduces redundancy along both spatial and temporal dimensions, so the benefit becomes more pronounced as trajectories get longer. These efficiency gains are what make 720p long-horizon reward-model training practical under our hardware setup (8×A100-80GB), without resorting to aggressive resolution downsampling. Additional analyses and visualizations of STP and TTP are provided in the supplementary materials.

Fig. 5: Comparison of STP and TTP

5.1 Visualization of STP and TTP

Figure 5 visualizes how STP and TTP reduce redundancy in video tokens. Under Qwen3-VL tokenization, each frame is represented by 880 tokens after spatial patch merging, and adjacent frames are further merged temporally. To stay consistent with this tokenization pipeline, we first compute STP masks on the original spatial patch grid, and then merge masks for temporally adjacent frames with a logical OR before applying them to merged tokens. This implementation detail explains the blank area on the right side of the first panel in frame 1: if a region is pruned in either of the two frames that are temporally merged, the corresponding merged token remains masked. TTP is then applied directly in token space. Because each token already carries information from neighboring frames, some content that appears early in the trajectory can remain visible in later visualizations. Nevertheless, TTP still removes a large amount of persistent redundancy, including repeated sidebars, window chrome, and static wallpaper/background regions.

6 Conclusion

In this work, we present a video-execution paradigm for reward modeling of computer-using agents (CUA). We introduce ExeVR-53k, a 53k–scale corpus of instruction–video–reward triplets, and propose adversarial instruction translation to synthesize hard negative pairs with step-level mismatch signals while producing justification-based attribution labels for temporal grounding. To make long-horizon, high-resolution execution videos trainable, we further develop spatiotemporal token pruning (STP+TTP), which reduces redundant spatial and temporal tokens while preserving decisive UI evidence.

References

1. Agashe, S., Wong, K., Tu, V., Yang, J., Li, A., Wang, X.E.: Agent s2: A compositional generalist-specialist framework for computer use agents (2025), https://arxiv.org/abs/2504.00906
2. AGI, Inc.: Agent - make your devices ai-native. https://www.theagi.company/agent (2025), accessed: February 14, 2026
3. Anthropic: Claude 3.7 Sonnet and Claude Code (February 2025), https://www.anthropic.com/news/claude-3-7-sonnet, accessed: 2026-02-14
4. Anthropic: Introducing Claude 4. https://www.anthropic.com/news/claude-4 (May 22 2025), accessed: 2026-02-14
5. Anthropic: Introducing claude sonnet 4.5 (September 29 2025), https://www.anthropic.com/news/claude-sonnet-4-5, accessed: 2026-02-14
6. Bai, H., Zhou, Y., Cemri, M., Pan, J., Suhr, A., Levine, S., Kumar, A.: Digirl: Training in-the-wild device-control agents with autonomous reinforcement learning (2024), https://arxiv.org/abs/2406.11896
7. Bai, S., Chen, K., Liu, X., Wang, J., Ge, W., Song, S., Dang, K., Wang, P., Wang, S., Tang, J., Zhong, H., Zhu, Y., Yang, M., Li, Z., Wan, J., Wang, P., Ding, W., Fu, Z., Xu, Y., Ye, J., Zhang, X., Xie, T., Cheng, Z., Zhang, H., Yang, Z., Xu, H., Lin, J.: Qwen2.5-vl technical report (2025), https://arxiv.org/abs/2502.13923
8. Bolya, D., Fu, C.Y., Dai, X., Zhang, P., Feichtenhofer, C., Hoffman, J.: Token merging: Your vit but faster (2023), https://arxiv.org/abs/2210.09461
9. Bonatti, R., Zhao, D., Bonacci, F., Dupont, D., Abdali, S., Li, Y., Lu, Y., Wagle, J., Koishida, K., Bucker, A., Jang, L.K., Hui, Z.: Windows agent arena: Evaluating multi-modal OS agents at scale. In: Forty-second International Conference on Machine Learning (2025), https://openreview.net/forum?id=W9s817KqYf
10. Bytedance Seed Team: Seed2.0 model card: Towards intelligence frontier for real-world complexity (2026)
11. Chae, H., Kim, S., Cho, J., Kim, S., Moon, S., et al.: Web-shepherd: Advancing prms for reinforcing web agents. arXiv preprint arXiv:2505.15277 (2025), https://arxiv.org/abs/2505.15277
12. Chen, C., Ji, K., Zhong, H., Zhu, M., Li, A., Gan, G., Huang, Z., Zou, C., Liu, J., Chen, J., Chen, H., Shen, C.: Gui-shepherd: Reliable process reward and verification for long-sequence gui tasks (2025), https://arxiv.org/abs/2509.23738
13. Chen, J., Yuen, D., Xie, B., Yang, Y., Chen, G., Wu, Z., Yixing, L., Zhou, X., Liu, W., Wang, S., Zhou, K., Shao, R., Nie, L., Wang, Y., Hao, J., Wang, J., Shao, K.: Spa-bench: A comprehensive benchmark for smartphone agent evaluation (2025), https://arxiv.org/abs/2410.15164
14. Chen, L., Zhao, H., Liu, T., Bai, S., Lin, J., Zhou, C., Chang, B.: An image is worth 1/2 tokens after layer 2: Plug-and-play inference acceleration for large vision-language models (2024), https://arxiv.org/abs/2403.06764
15. Chen, R., Chen, X., Wang, S., Kong, S., Yu, J.: Let-us: Long event-text understanding of scenes (2025), https://arxiv.org/abs/2508.07401
16. Cheng, Y., Tang, L., Li, S., Huo, Y., Duan, T., Huang, K., Jing, Y., Yan, Y.: Evolving in tasks: Empowering the multi-modality large language model as the computer use agent (2026), https://arxiv.org/abs/2508.04037
17. Clark, C., Zhang, J., Ma, Z., Park, J.S., Salehi, M., Tripathi, R., Lee, S., Ren, Z., Kim, C.D., Yang, Y., Shao, V., Yang, Y., Huang, W., Gao, Z., Anderson, T., Zhang, J., Jain, J., Stoica, G., Han, W., Farhadi, A., Krishna, R.: Molmo2: Open weights and data for vision-language models with video understanding and grounding (2026), https://arxiv.org/abs/2601.10611

18. Comanici, G., Bieber, E., Schaekermann, M., Pasupat, I., Sachdeva, N., Dhillon, I., Blistein, M., Ram, O., Zhang, D., Rosen, E., Marris, L., Petulla, S., Gaffney, C., Aharoni, A., Lintz, N., Pais, T.C., Jacobsson, H., Szpektor, I., Jiang, N.J., Haridasan, K., Omran, A., Saunshi, N., Bahri, D., Mishra, G., Chu, E., Boyd, T., Hekman, B., Parisi, A., Zhang, C., Kawintiranon, K., Bedrax-Weiss, T., Wang, O., Xu, Y., Purkiss, O., Mendlovic, U., Deutel, I., Nguyen, N., Langley, A., Korn, F., Rossazza, L., Ramé, A., Waghmare, S., Miller, H., Byrd, N., Sheshan, A., Hadsell, R., Bhardwaj, S., Janus, P., Rissa, T., Horgan, D., Abdagic, A., Belenki, L., Allingham, J., Singh, A., Guidroz, T., Srinivasan, S., Schmit, H., Chiafullo, K., Elisseeff, A., Jha, N., Kolhar, P., Berrada, L., Ding, F., Si, X., Mallick, S.B., Och, F., Erell, S., Ni, E., Latkar, T., Yang, S., Sirkovic, P., Feng, Z., Leland, R., Hornung, R., Wu, G., Blundell, C., Alvari, H., Huang, P.S., Yip, C., Deur, S., Liu, L., Surita, G., Duque, P., Damen, D., Jia, J., Guez, A., Mircea, M., Sinha, A., Magni, A., Stradomski, P., Marian, T., Galić, V., Chen, W., Husain, H., Singhal, A., Grewe, D., Aubet, F.X., Song, S., Blanco, L., Rechis, L., Ho, L., Munoz, R., Zheng, K., Hamrick, J., Mather, K., Taitelbaum, H., Rutherford, E., Lei, Y., Chen, K., Shukla, A., Moreira, E., Doi, E., Isik, B., Shabat, N., Rogozińska, D., Kolipaka, K., Chang, J., Vušak, E., Venkatachary, S., Noghabi, S., Bharti, T., Jun, Y., Zaks, A., Green, S., Challagundla, J., Wong, W., Mohammad, M., Hirsch, D., Cheng, Y., Naim, I., Proleev, L., Vincent, D., Singh, A., Krikun, M., Krishnan, D., Ghahramani, Z., Atias, A., Aggarwal, R., Kirov, C., Vytiniotis, D., Koh, C., Chronopoulou, A., Dogra, P., Ion, V.D., Tyen, G., Lee, J., Weissenberger, F., Strohman, T., Balakrishna, A., Rae, J., Velic, M., de Liedekerke, R., Elyada, O., Yuan, W., Liu, C., Shani, L., Kishchenko, S., Alessio, B., Li, Y., Song, R., Kwei, S., Jankowski, O., Pappu, A., Namiki, Y., Ma, Y., Tripuraneni, N., Cherry, C., Ikonomidis, M., Ling, Y.C., Ji, C., Westberg, B., Wright, A., Yu, D., Parkinson, D., Ramaswamy, S., Connor, J., Yeganeh, S.H., Grover, S., Kenwright, G., Litchev, L., Apps, C., Tomala, A., Halim, F., Castro-Ros, A., Li, Z., Boral, A., Sho, P., Yarom, M., Malmi, E., Klinghoffer, D., Lin, R., Ansell, A., S, P.K., Zhao, S., Zuo, S., Santoro, A., Cheng, H.T., Demmessie, S., Liu, Y., Brichtova, N., Culp, A., Braun, N., Graur, D., Ng, W., Mehta, N., Phillips, A., Sundberg, P., Godbole, V., Liu, F., Katariya, Y., Rim, D., Seyedhosseini, M., Ammirati, S., Valfridsson, J., Malihi, M., Knight, T., Toor, A., Lampe, T., Ittycheriah, A., Chiang, L., Yeung, C., Fréchette, A., Rao, J., Wang, H., Srivastava, H., Zhang, R., Rhodes, R., Brand, A., Weesner, D., Figotin, I., Gimeno, F., Fellinger, R., Marcenac, P., Leal, J., Marcus, E., Cotruta, V., Cabrera, R., Luo, S., Garrette, D., Axelrod, V., Baltateanu, S., Barker, D., Chen, D., Toma, H., Ingram, B., Riesa, J., Kulkarni, C., Zhang, Y., Liu, H., Wang, C., Polacek, M., Wu, W., Hui, K., Reyes, A.N., Su, Y., Barnes, M., Malhi, I., Siddiqui, A., Feng, Q., Damaschin, M., Pighin, D., Steiner, A., Yang, S., Boppana, R.S., Ivanov, S., Kandoor, A., Shah, A., Mujika, A., Huang, D., Choquette-Choo, C.A., Patel, M., Yu, T., Creswell, T., Jerry, Liu, Barros, C., Razeghi, Y., Roy, A., Culliton, P., Xiong, B., Pan, J., Strohmann, T., Powell, T., Seal, B., DeCarlo, D., Shyam, P., Katircioglu, K., Wang, X., Hardin, C., Odisho, I., Broder, J., Chang, O., Nair, A., Shtefan, A., O'Brien, M., Agarwal, M., Potluri, S., Goyal, S., Jhindal, A., Thakur, S., Stuken, Y., Lyon, J., Toutanova, K., Feng, F., Wu, A., Horn, B., Wang, A., Cullum, A., Taubman, G., Shrivastava, D., Shi, C., Tomlinson, H., Patel, R., Tu, T., Oflazer, A.M., Pongetti, F., Yang, M., Taïga, A.A., Perot, V., Pierse, N.W., Han, F., Drori, Y., Iturrate, I., Chakrabarti, A., Yeung, L., Dopson, D., ting Chen, Y., Kulshreshtha, A., Guo, T., Pham, P., Schuster, T., Chen, J., Polozov, A., Xing, J., Zhou, H., Kacham, P., Kukliansky, D., Miech, A., Yaroshenko, S., Chi, E., Dou-

glas, S., Fei, H., Blondel, M., Myla, P., Madmoni, L., Wu, X., Keysers, D., Kjems, K., Albuquerque, I., Yu, L., D'sa, J., Plantan, M., Ionescu, V., Elias, J.S., Gupta, A., Vuyyuru, M.R., Alcober, F., Zhou, T., Ji, K., Hartmann, F., Puttagunta, S., Song, H., Amid, E., Stefanoiu, A., Lee, A., Pucciarelli, P., Wang, E., Raul, A., Petrov, S., Tian, I., Anklin, V., Nti, N., Gomes, V., Schumacher, M., Vesom, G., Panagopoulos, A., Bousmalis, K., Andor, D., Jacob, J., Zhang, Y., Rosgen, B., Kecman, M., Tung, M., Belias, A., Goodman, N., Covington, P., Wieder, B., Saxena, N., Davoodi, E., Huang, M., Maddineni, S., Roulet, V., Campbell-Ajala, F., Sessa, P.G., Xintian, Wu, Lai, G., Collins, P., Haig, A., Sakenas, V., Xu, X., Giustina, M., Shafey, L.E., Charoenpanit, P., Garg, S., Ainslie, J., Severson, B., Arenas, M.G., Pathak, S., Rajayogam, S., Feng, J., Bakker, M., Li, S., Wichers, N., Rogers, J., Geng, X., Li, Y., Jagerman, R., Jia, C., Olmert, N., Sharon, D., Mauger, M., Mariserla, S., Ma, H., Mohabey, M., Kim, K., Andreev, A., Pollom, S., Love, J., Jain, V., Agrawal, P., Schroecker, Y., Fortin, A., Warmuth, M., Liu, J., Leach, A., Blok, I., Girirajan, G.P., Aharoni, R., Uria, B., Sozanschi, A., Goldberg, D., Ionita, L., Ribeiro, M.T., Zlocha, M., Birodkar, V., Lachgar, S., Yuan, L., Choudhury, H., Ginsberg, M., Zheng, F., Dibb, G., Graves, E., Lokhande, S., Rasskin, G., Muraru, G.C., Quick, C., Tata, S., Sermanet, P., Chawla, A., Karo, I., Wang, Y., Zhang, S., Keller, O., Dragan, A., Su, G., Chou, I., Liu, X., Tao, Y., Prabhakara, S., Wilson, M., Liu, R., Wang, S., Evans, G., Du, D., Castaño, A., Prasad, G., Mahdy, M.E., Gerlach, S., Reid, M., Kahn, J., Zait, A., Pillai, T.S., Ulrich, T., Wang, G., Wassenberg, J., Farkash, E., Yalasangi, K., Wang, C., Bauza, M., Bucher, S., Liu, T., Yan, J., Leung, G., Sindhwani, V., Barnes, P., Singh, A., Jurin, I., Chang, J., Bhumihar, N.K., Eiger, S., Citovsky, G., Withbroe, B., Li, Z., Xue, S., Santo, N.D., Stoyanov, G., Raimond, Y., Zheng, S., Gao, Y., Listík, V., Kwasiborski, S., Saputro, R., Ozturel, A., Mallya, G., Majmundar, K., West, R., Caron, P., Wei, J., Castrejon, L., Vikram, S., Ramachandran, D., Dhawan, N., Park, J., Smoot, S., van den Driessche, G., Blau, Y., Malik, C., Liang, W., Hirsch, R., dos Santos, C.N., Weinstein, E., van den Oord, A., Lall, S., FitzGerald, N., Jiang, Z., Yang, X., Webster, D., Elqursh, A., Pope, A., Rotival, G., Raposo, D., Zhu, W., Dean, J., Alabed, S., Tran, D., Gupta, A., Gleicher, Z., Austin, J., Rosseel, E., Umekar, M., Das, D., Sun, Y., Chen, K., Misiunas, K., Zhou, X., Di, Y., Loo, A., Newlan, J., Li, B., Ramasesh, V., Xu, Y., Chen, A., Gandhe, S., Soricut, R., Gupta, N., Hu, S., El-Sayed, S., Garcia, X., Brusilovsky, I., Chen, P.C., Bolt, A., Huang, L., Gurney, A., Zhang, Z., Pritzel, A., Wilkiewicz, J., Seybold, B., Shamanna, B.K., Fischer, F., Dean, J., Gill, K., Mcilroy, R., Bhowmick, A., Selier, J., Yang, A., Cheng, D., Magay, V., Tan, J., Varma, D., Walder, C., Kocisky, T., Nakashima, R., Natsev, P., Kwong, M., Gog, I., Zhang, C., Dieleman, S., Jimma, T., Ryabtsev, A., Brahma, S., Steiner, D., Du, D., Žužul, A., Žanić, M., Raghavachari, M., Gierke, W., Zheng, Z., Petrova, D., Dauphin, Y., Liu, Y., Kessler, I., Hand, S., Duvarney, C., Kim, S., Lee, H., Hussenot, L., Hui, J., Smith, J., Jain, D., Xia, J., Tomar, G.S., Amiri, K., Phan, D., Fuchs, F., Weyand, T., Tomasev, N., Cordell, A., Liu, X., Mallinson, J., Joshi, P., Crawford, A., Suggala, A., Chien, S., Fernando, N., Sanchez-Vargas, M., Williams, D., Crone, P., Luo, X., Karpov, I., Shan, J., Thurk, T., Strudel, R., Voigtlaender, P., Patil, P., Dozat, T., Khodaei, A., Singla, S., Ambroszczyk, P., Wu, Q., Chang, Y., Roark, B., Hegde, C., Ding, T., Filos, A., Wu, Z., Pinto, A.S., Liu, S., Khanna, S., Pandey, A., Mcloughlin, S., Li, Q., Haves, S., Zhou, A., Buchatskaya, E., Leal, I., de Boursac, P., Akazawa, N., Anderson, N., Chen, T., Somandepalli, K., Liang, C., Goenka, S., Winkler, S., Grushetsky, A., Ding, Y., Smith, J., Ye, F., Pont-Tuset, J., Li, E., Li, R., Golany, T., Wegner, D., Jiang, T., Barak, O.,

Shangguan, Y., Vértes, E., Wong, R., Bornschein, J., Tudor, A., Bevilacqua, M., Schaul, T., Rawat, A.S., Zhao, Y., Axiotis, K., Meng, L., McLean, C., Lai, J., Beattie, J., Kushman, N., Liu, Y., Kutzman, B., Lang, F., Ye, J., Netrapalli, P., Mishra, P., Khan, M., Goel, M., Willoughby, R., Tian, D., Zhuang, H., Chen, J., Tsai, Z., Kementsietsidis, T., Khare, A., Keeling, J., Xu, K., Waters, N., Altché, F., Popat, A., Mittal, B., Saxton, D., Badawy, D.E., Mathieu, M., Zheng, Z., Zhou, H., Ranka, N., Shin, R., Duan, Q., Salimans, T., Mihailescu, I., Shaham, U., Chang, M.W., Assael, Y., Dikkala, N., Izzard, M., Cohen-Addad, V., Graves, C., Feinberg, V., Chung, G., Strouse, D., Karmon, D., Sharifzadeh, S., Ashwood, Z., Pham, K., Blanton, J., Vasiloff, A., Barber, J., Geller, M., Zhou, A., Zubach, F., Huang, T.K., Zhang, L., Gupta, H., Young, M., Proskurnia, J., Votel, R., Gabeur, V., Barcik, G., Tripathi, A., Yu, H., Yan, G., Changpinyo, B., Pavetić, F., Coyle, A., Fujii, Y., Mendez, J.G., Zhou, T., Rajamani, H., Hechtman, B., Cao, E., Juan, D.C., Tan, Y.X., Dalibard, V., Du, Y., Clay, N., Yao, K., Jia, W., Vijaykumar, D., Zhou, Y., Bai, X., Hung, W.C., Pecht, S., Todorov, G., Khadke, N., Gupta, P., Lahoti, P., Autef, A., Duddu, K., Lee-Thorp, J., Bykovsky, A., Misiunas, T., Flennerhag, S., Thangaraj, S., McGiffin, J., Nado, Z., Kunesch, M., Noever, A., Hertz, A., Liang, M., Stone, V., Palmer, E., Daruki, S., Pramanik, A., Põder, S., Kyker, A., Khan, M., Sluzhaev, E., Ritter, M., Ruderman, A., Zhou, W., Nagpal, C., Vodrahalli, K., Necula, G., Barham, P., Pavlick, E., Hartford, J., Shafran, I., Zhao, L., Mikuła, M., Eccles, T., Shimokawa, H., Garg, K., Vilnis, L., Chen, H., Shumailov, I., Lee, K.H., Abdelhamed, A., Xie, M., Cohen, V., Hlavnova, E., Malkin, D., Sitawarin, C., Lottes, J., Coquinot, P., Yu, T., Kumar, S., Zhang, J., Mahendru, A., Ahmed, Z., Martens, J., Chen, T., Boag, A., Peng, D., Devin, C., Klimovskiy, A., Phuong, M., Vainstein, D., Xie, J., Ramabhadran, B., Howard, N., Yu, X., Goswami, G., Cui, J., Shleifer, S., Pinto, M., Yeh, C.K., Yang, M.H., Javanmardi, S., Ethier, D., Lee, C., Orbay, J., Kotecha, S., Bromberg, C., Shaw, P., Thornton, J., Rosenthal, A.G., Gu, S., Thomas, M., Gemp, I., Ayyar, A., Ushio, A., Selvan, A., Wee, J., Liu, C., Majzoubi, M., Yu, W., Abernethy, J., Liechty, T., Pan, R., Nguyen, H., Qiong, Hu, Perrin, S., Arora, A., Pitler, E., Wang, W., Shivakumar, K., Prost, F., Limonchik, B., Wang, J., Gao, Y., Cour, T., Buch, S., Gui, H., Ivanova, M., Neubeck, P., Chan, K., Kim, L., Chen, H., Goyal, N., Chung, D.W., Liu, L., Su, Y., Petrushkina, A., Shen, J., Joulin, A., Xu, Y., Lin, S.X., Kulizhskaya, Y., Chelba, C., Vasudevan, S., Collins, E., Bashlovkina, V., Lu, T., Fritz, D., Park, J., Zhou, Y., Su, C., Tanburn, R., Sushkov, M., Rasquinha, M., Li, J., Prendki, J., Li, Y., LV, P., Sharma, S., Fitoussi, H., Huang, H., Dai, A., Dao, P., Burrows, M., Prior, H., Qin, D., Pundak, G., Sjoesund, L.L., Khurshudov, A., Zhu, Z., Webson, A., Kemp, E., Tan, T., Agrawal, S., Sargsyan, S., Cheng, L., Stephan, J., Kwiatkowski, T., Reid, D., Byravan, A., Michaely, A.H., Heess, N., Zhou, L., Goenka, S., Carpenter, V., Levskaya, A., Wang, B., Roberts, R., Leblond, R., Chikkerur, S., Ginzburg, S., Chang, M., Riachi, R., Chuqiao, Xu, Borsos, Z., Pliskin, M., Pawar, J., Lustman, M., Kirkwood, H., Anand, A., Chaudhary, A., Kalb, N., Milan, K., Augenstein, S., Goldie, A., Prince, L., Raman, K., Sun, Y., Xia, V., Cohen, A., Huo, Z., Camp, J., Ellis, S., Zilka, L., Torres, D.V., Patel, L., Arora, S., Chan, B., Adler, J., Ayoub, K., Liang, J., Jamil, F., Jiang, J., Baumgartner, S., Sun, H., Karov, Y., Akulov, Y., Zheng, H., Cai, I., Fantacci, C., Rubin, J., Acha, A.R., Wang, M., D'Souza, N., Sathyanarayana, R., Dai, S., Rowe, S., Simanovsky, A., Goldman, O., Kuang, Y., Pan, X., Rosenberg, A., Rojas-Esponda, T., Dutta, P., Zeng, A., Jurenka, I., Farquhar, G., Bansal, Y., Iqbal, S., Roelofs, B., Joung, G.Y., Beak, P., Ryu, C., Poplin, R., Wu, Y., Alayrac, J.B., Buthpitiya, S., Ronneberger, O., Habtegebriel,

C., Li, W., Cavallaro, P., Wei, A., Bensky, G., Denk, T., Ganapathy, H., Stanway, J., Joshi, P., Bertolini, F., Lo, J., Ma, O., Charles, Z., Sampemane, G., Sahni, H., Chen, X., Askham, H., Gaddy, D., Young, P., Tan, J., Eyal, M., Bražinskas, A., Zhong, L., Wu, Z., Epstein, M., Bailey, K., Hard, A., Lee, K., Goldshtein, S., Ruiz, A., Badawi, M., Lochbrunner, M., Kearns, J., Brown, A., Pardo, F., Weber, T., Yang, H., Jiang, P.P., Akin, B., Fu, Z., Wainwright, M., Zou, C., Gaba, M., Manzagol, P.A., Kan, W., Song, Y., Zainullina, K., Lin, R., Ko, J., Deshmukh, S., Jindal, A., Svensson, J., Tyam, D., Zhao, H., Kaeser-Chen, C., Baird, S., Moradi, P., Hall, J., Guo, Q., Tsang, V., Liang, B., Pereira, F., Ganesh, S., Korotkov, I., Adamek, J., Thiagarajan, S., Tran, V., Chen, C., Tar, C., Jain, S., Dasgupta, I., Bilal, T., Reitter, D., Zhao, K., Vezzani, G., Gehman, Y., Mehta, P., Beltrone, L., Dotiwalla, X., Guadarrama, S., Abbas, Z., Karp, S., Georgiev, P., Ferng, C.S., Brockschmidt, M., Peng, L., Hirnschall, C., Verma, V., Bi, Y., Xiao, Y., Dabush, A., Xu, K., Wallis, P., Parker, R., Wang, Q., Xu, Y., Safarli, I., Tewari, D., Zhang, Y., Kim, S., Gesmundo, A., Thomas, M., Levi, S., Chowdhury, A., Rao, K., Garst, P., Conway-Rahman, S., Ran, H., McKinney, K., Xiao, Z., Yu, W., Agrawal, R., Stjerngren, A., Ionescu, C., Chen, J., Sharma, V., Chiu, J., Liu, F., Franko, K., Sanford, C., Cai, X., Michel, P., Ganapathy, S., Labanowski, J., Garrett, Z., Vargas, B., Sun, S., Gale, B., Buschmann, T., Desjardins, G., Ghelani, N., Jain, P., Verma, M., Asawaroengchai, C., Eisenschlos, J., Harlalka, J., Kazawa, H., Metzler, D., Howland, J., Jian, Y., Ades, J., Shah, V., Gangwani, T., Lee, S., Ring, R., Hernandez, S.M., Reich, D., Sinha, A., Sathe, A., Kovac, J., Gill, A., Kannan, A., D'olimpio, A., Sevenich, M., Whang, J., Kim, B., Sim, K.C., Chen, J., Zhang, J., Lall, S., Matias, Y., Jia, B., Friesen, A., Nasso, S., Thapliyal, A., Perozzi, B., Yu, T., Shekhawat, A., Huda, S., Grabowski, P., Wang, E., Sreevatsa, A., Dib, H., Hassen, M., Schuh, P., Milutinovic, V., Welty, C., Quinn, M., Shah, A., Wang, B., Barth-Maron, G., Frye, J., Axelsson, N., Zhu, T., Ma, Y., Giannoumis, I., Sedghi, H., Ye, C., Luan, Y., Aydin, K., Chandra, B., Sampathkumar, V., Huang, R., Lavrenko, V., Eleryan, A., Hong, Z., Hansen, S., Carthy, S.M., Samanta, B., Ćevid, D., Wang, X., Li, F., Voznesensky, M., Hoffman, M., Terzis, A., Sehwag, V., Fidel, G., He, L., Cai, M., He, Y., Feng, A., Nikoltchev, M., Phatale, S., Chase, J., Lawton, R., Zhang, M., Ouyang, T., Tragut, M., Manshadi, M.H., Narayanan, A., Shen, J., Gao, X., Bolukbasi, T., Roy, N., Li, X., Golovin, D., Panait, L., Qin, Z., Han, G., Anthony, T., Kudugunta, S., Patraucean, V., Ray, A., Chen, X., Yang, X., Bhatia, T., Talluri, P., Morris, A., Ražnatović, A., Brownfield, B., An, J., Peng, S., Kane, P., Zheng, C., Duduta, N., Kessinger, J., Noraky, J., Liu, S., Rong, K., Veličković, P., Rush, K., Goldin, A., Wei, F., Garlapati, S.M.R., Pantofaru, C., Kwon, O., Ni, J., Noland, E., Trapani, J.D., Beaufays, F., Roy, A.G., Chow, Y., Turker, A., Cideron, G., Mei, L., Clark, J., Dou, Q., Bošnjak, M., Leith, R., Du, Y., Yazdanbakhsh, A., Nasr, M., Kwak, C., Sheth, S.S., Kaskasoli, A., Anand, A., Lakshminarayanan, B., Jerome, S., Bieber, D., Chu, C.T., Senges, A., Shen, T., Sridhar, M., Ndebele, N., Beyret, B., Mohamed, S., Chen, M., Freitag, M., Guo, J., Liu, L., Roit, P., Chen, H., Yan, S., Stone, T., Co-Reyes, J., Cole, J., Scellato, S., Azizi, S., Hashemi, H., Jin, A., Iyer, A., Valentine, M., György, A., Ahuja, A., Diaz, D.H., Lee, C.Y., Clement, N., Kong, W., Garmon, D., Watts, I., Bhatia, K., Gupta, K., Miecnikowski, M., Vallet, H., Taly, A., Loper, E., Joshi, S., Atwood, J., Chick, J., Collier, M., Iliopoulos, F., Trostle, R., Gunel, B., Leal-Cavazos, R., Hrafnkelsson, A.M., Guzman, M., Ju, X., Forbes, A., Emond, J., Chauhan, K., Caine, B., Xiao, L., Zeng, W., Moufarek, A., Murphy, D., Meng, M., Gupta, N., Riedel, F., Das, A., Lawal, E., Narayan, S., Sosea, T., Swirhun, J., Friso, L., Neyshabur, B.,

Lu, J., Girgin, S., Wunder, M., Yvinec, E., Pyne, A., Carbune, V., Rijhwani, S., Guo, Y., Doshi, T., Briukhov, A., Bain, M., Hitron, A., Wang, X., Gupta, A., Chen, K., Du, C., Zhang, W., Shah, D., Akula, A., Dylla, M., Kachra, A., Kuo, W., Zou, T., Wang, L., Xu, L., Zhu, J., Snyder, J., Menon, S., Firat, O., Mordatch, I., Yuan, Y., Ponomareva, N., Blevins, R., Moore, L., Wang, W., Chen, P., Scholz, M., Dwornik, A., Lin, J., Li, S., Antognini, D., I, T., Song, X., Miller, M., Kalra, U., Raveret, A., Akerlund, O., Wu, F., Nystrom, A., Godbole, N., Liu, T., DeBalsi, H., Zhao, J., Liu, B., Caciularu, A., Lax, L., Khandelwal, U., Langston, V., Bailey, E., Lattanzi, S., Wang, Y., Kovelamudi, N., Mondal, S., Guruganesh, G., Hua, N., Roval, O., Wesołowski, P., Ingale, R., Halcrow, J., Sohn, T., Angermueller, C., Raad, B., Stickgold, E., Lu, E., Kosik, A., Xie, J., Lillicrap, T., Huang, A., Zhang, L.L., Paulus, D., Farabet, C., Wertheim, A., Wang, B., Joshi, R., ling Ko, C., Wu, Y., Agrawal, S., Lin, L., Sheng, X., Sung, P., Breland-King, T., Butterfield, C., Gawde, S., Singh, S., Zhang, Q., Apte, R., Shetty, S., Hutter, A., Li, T., Salesky, E., Lebron, F., Kanerva, J., Paganini, M., Nguyen, A., Vallu, R., Peter, J.T., Velury, S., Kao, D., Hoover, J., Bortsova, A., Bishop, C., Jakobovits, S., Agostini, A., Agarwal, A., Liu, C., Kwong, C., Tavakkol, S., Bica, I., Greve, A., GP, A., Marcus, J., Hou, L., Duerig, T., Moroshko, R., Lacey, D., Davis, A., Amelot, J., Wang, G., Kim, F., Strinopoulos, T., Wan, H., Lan, C.L., Krishnan, S., Tang, H., Humphreys, P., Bai, J., Shtacher, I.H., Machado, D., Pang, C., Burke, K., Liu, D., Aravamudhan, R., Song, Y., Hirst, E., Singh, A., Jou, B., Bai, L., Piccinno, F., Fu, C.K., Alazard, R., Meiri, B., Winter, D., Chen, C., Zhang, M., Heitkaemper, J., Lambert, J., Lee, J., Frömmgen, A., Rogulenko, S., Nair, P., Niemczyk, P., Bulyenov, A., Xu, B., Shemtov, H., Zadimoghaddam, M., Toropov, S., Wirth, M., Dai, H., Gollapudi, S., Zheng, D., Kurakin, A., Lee, C., Bullard, K., Serrano, N., Balazevic, I., Li, Y., Schalkwyk, J., Murphy, M., Zhang, M., Sequeira, K., Datta, R., Agrawal, N., Sutton, C., Attaluri, N., Chiang, M., Farhan, W., Thornton, G., Lin, K., Choma, T., Nguyen, H., Dasgupta, K., Robinson, D., Comşa, I., Riley, M., Pillai, A., Mustafa, B., Golan, B., Zandieh, A., Lespiau, J.B., Porter, B., Ross, D., Rajayogam, S., Agarwal, M., Venugopalan, S., Shahriari, B., Yan, Q., Xu, H., Tobin, T., Dubov, P., Shi, H., Recasens, A., Kovsharov, A., Borgeaud, S., Dery, L., Vasanth, S., Gribovskaya, E., Qiu, L., Mahdieh, M., Skut, W., Nielsen, E., Zheng, C., Yu, A., Bostock, C.G., Gupta, S., Archer, A., Rawles, C., Davies, E., Svyatkovskiy, A., Tsai, T., Halpern, Y., Reisswig, C., Wydrowski, B., Chang, B., Puigcerver, J., Taege, M.H., Li, J., Schnider, E., Li, X., Dena, D., Xu, Y., Telang, U., Shi, T., Zen, H., Kastner, K., Ko, Y., Subramaniam, N., Kumar, A., Blois, P., Dai, Z., Wieting, J., Lu, Y., Zeldes, Y., Xie, T., Hauth, A., Ţifrea, A., Li, Y., El-Husseini, S., Abolafia, D., Zhou, H., Ding, W., Ghalebikesabi, S., Guía, C., Maksai, A., Ágoston Weisz, Arik, S., Sukhanov, N., Świetlik, A., Jia, X., Yu, L., Wang, W., Brand, M., Bloxwich, D., Kirmani, S., Chen, Z., Go, A., Sprechmann, P., Kannen, N., Carin, A., Sandhu, P., Edkins, I., Nooteboom, L., Gupta, J., Maggiore, L., Azizi, J., Pritch, Y., Yin, P., Gupta, M., Tarlow, D., Smith, D., Ivanov, D., Babaeizadeh, M., Goel, A., Kambala, S., Chu, G., Kastelic, M., Liu, M., Soltau, H., Stone, A., Agrawal, S., Kim, M., Soparkar, K., Tadepalli, S., Bunyan, O., Soh, R., Kannan, A., Kim, D., Chen, B.J., Halumi, A., Roy, S., Wang, Y., Sercinoglu, O., Gibson, G., Bhatnagar, S., Sano, M., von Dincklage, D., Ren, Q., Mitrevski, B., Olšák, M., She, J., Doersch, C., Jilei, Wang, Liu, B., Tan, Q., Yakar, T., Warkentin, T., Ramirez, A., Lebsack, C., Dillon, J., Mathews, R., Cobley, T., Wu, Z., Chen, Z., Simon, J., Nath, S., Sainath, T., Bendebury, A., Julian, R., Mankalale, B., Ćurko, D., Zacchello, P., Brown, A.R., Sodhia, K., Howard, H., Caelles, S., Gupta,

A., Evans, G., Bulanova, A., Katzen, L., Goldenberg, R., Tsitsulin, A., Stanton, J., Schillings, B., Kovalev, V., Fry, C., Shah, R., Lin, K., Upadhyay, S., Li, C., Radpour, S., Maggioni, M., Xiong, J., Haas, L., Brennan, J., Kamath, A., Savinov, N., Nagrani, A., Yacovone, T., Kappedal, R., Andriopoulos, K., Lao, L., Li, Y., Rozhdestvenskiy, G., Hashimoto, K., Audibert, A., Austin, S., Rodriguez, D., Ruoss, A., Honke, G., Karkhanis, D., Xiong, X., Wei, Q., Huang, J., Leng, Z., Premachandran, V., Bileschi, S., Evangelopoulos, G., Mensink, T., Pavagadhi, J., Teplyashin, D., Chang, P., Xue, L., Tanzer, G., Goldman, S., Patel, K., Li, S., Wiesner, J., Zheng, I., Stewart-Binks, I., Han, J., Li, Z., Luo, L., Lenc, K., Lučić, M., Xue, F., Mullins, R., Guseynov, A., Chang, C.C., Galatzer-Levy, I., Zhang, A., Bingham, G., Hu, G., Hartman, A., Ma, Y., Griffith, J., Irpan, A., Radebaugh, C., Yue, S., Fan, L., Ungureanu, V., Sorokin, C., Teufel, H., Li, P., Anil, R., Paparas, D., Wang, T., Lin, C.C., Peng, H., Shum, M., Petrovic, G., Brady, D., Nguyen, R., Macherey, K., Li, Z., Singh, H., Yenugula, M., Iinuma, M., Chen, X., Kopparapu, K., Stern, A., Dave, S., Thekkath, C., Perot, F., Kumar, A., Li, F., Xiao, Y., Bilotti, M., Bateni, M.H., Noble, I., Lee, L., Vázquez-Reina, A., Salazar, J., Yang, X., Wang, B., Gruzewska, E., Rao, A., Raghuram, S., Xu, Z., Ben-David, E., Mei, J., Dalmia, S., Zhang, Z., Liu, Y., Bansal, G., Pankov, H., Schwarcz, S., Burns, A., Chan, C., Sanghai, S., Liang, R., Liang, E., He, A., Stuart, A., Narayanan, A., Zhu, Y., Frank, C., Fatemi, B., Sabne, A., Lang, O., Bhattacharya, I., Settle, S., Wang, M., McMahan, B., Tacchetti, A., Soares, L.B., Hadian, M., Cabi, S., Chung, T., Putikhin, N., Li, G., Chen, J., Tarango, A., Michalewski, H., Kazemi, M., Masoom, H., Sheftel, H., Shivanna, R., Vadali, A., Comanescu, R., Reid, D., Moore, J., Neelakantan, A., Sander, M., Herzig, J., Rosenberg, A., Dehghani, M., Choi, J., Fink, M., Hayes, R., Ge, E., Weng, S., Ho, C.H., Karro, J., Krishna, K., Thiet, L.N., Skerry-Ryan, A., Eppens, D., Andreetto, M., Sarma, N., Bonacina, S., Ayan, B.K., Nawhal, M., Shan, Z., Dusenberry, M., Thakoor, S., Gubbi, S., Nguyen, D.D., Tsarfaty, R., Albanie, S., Mitrović, J., Gandhi, M., Chen, B.J., Epasto, A., Stephanov, G., Jin, Y., Gehman, S., Amini, A., Weber, J., Behbahani, F., Xu, S., Allamanis, M., Chen, X., Ott, M., Sha, C., Jastrzebski, M., Qi, H., Greene, D., Wu, X., Toki, A., Vlasic, D., Shapiro, J., Kotikalapudi, R., Shen, Z., Saeki, T., Xie, S., Cassirer, A., Bharadwaj, S., Kiyono, T., Bhojanapalli, S., Rosenfeld, E., Ritter, S., Mao, J., Oliveira, J.G., Egyed, Z., Bandemer, B., Parisotto, E., Kinoshita, K., Pluto, J., Maniatis, P., Li, S., Guo, Y., Ghiasi, G., Tarbouriech, J., Chatterjee, S., Jin, J., Katrina, Xu, Palomaki, J., Arnold, S., Sewak, M., Piccinini, F., Sharma, M., Albrecht, B., Purser-haskell, S., Vaswani, A., Chen, C., Wisniewski, M., Cao, Q., Aslanides, J., Phu, N.M., Sieb, M., Agubuzu, L., Zheng, A., Sohn, D., Selvi, M., Andreassen, A., Subudhi, K., Eruvbetine, P., Woodman, O., Mery, T., Krause, S., Ren, X., Ma, X., Luo, J., Chen, D., Fan, W., Griffiths, H., Schuler, C., Li, A., Zhang, S., Sarr, J.M., Luo, S., Patana, R., Watson, M., Naboulsi, D., Collins, M., Sidhwani, S., Hoogeboom, E., Silver, S., Caveness, E., Zhao, X., Rodriguez, M., Deines, M., Bai, L., Griffin, P., Tagliasacchi, M., Xue, E., Babbula, S.R., Pang, B., Ding, N., Shen, G., Peake, E., Crocker, R., Raghvendra, S.S., Swisher, D., Han, W., Singh, R., Wu, L., Pchelin, V., Munkhdalai, T., Alon, D., Bacon, G., Robles, E., Bulian, J., Johnson, M., Powell, G., Ferreira, F.T., Li, Y., Benzing, F., Velimirović, M., Soyer, H., Kong, W., Tony, Nguyên, Yang, Z., Liu, J., van Amersfoort, J., Gillick, D., Sun, B., Rauschmayr, N., Zhang, K., Zhan, S., Zhou, T., Frolov, A., Yang, C., Vnukov, D., Rouillard, L., Li, H., Mandhane, A., Fallen, N., Venkataraman, R., Hu, C.H., Brennan, J., Lee, J., Chang, J., Sundermeyer, M., Pan, Z., Ke, R., Tong, S., Fabrikant, A., Bono, W., Gu, J., Foley, R., Mao, Y.,

Delakis, M., Bhaswar, D., Frostig, R., Li, N., Zipori, A., Hope, C., Kozlova, O., Mishra, S., Djolonga, J., Schiff, C., Merey, M.A., Briakou, E., Morgan, P., Wan, A., Hassidim, A., Skerry-Ryan, R., Sengupta, K., Jasarevic, M., Kallakuri, P., Kunkle, P., Brennan, H., Lieber, T., Mansoor, H., Walker, J., Zhang, B., Xie, A., Žužić, G., Chukwuka, A., Druinsky, A., Cho, D., Yao, R., Naeem, F., Butt, S., Kim, E., Jia, Z., Jordan, M., Lelkes, A., Kurzeja, M., Wang, S., Zhao, J., Over, A., Chakladar, A., Prasetya, M., Jha, N., Ganapathy, S., Cong, Y., Shroff, P., Saroufim, C., Miryoosefi, S., Hammad, M., Nasir, T., Xi, W., Gao, Y., Maeng, Y., Hora, B., Cheng, C.Y., Haghani, P., Lewenberg, Y., Lu, C., Matysiak, M., Raisinghani, N., Wang, H., Baugher, L., Sukthankar, R., Giang, M., Schultz, J., Fiedel, N., Chen, M., Lee, C.C., Dey, T., Zheng, H., Paul, S., Smith, C., Ly, A., Wang, Y., Bansal, R., Perz, B., Ricco, S., Blank, S., Keshava, V., Sharma, D., Chow, M., Lad, K., Jalan, K., Osindero, S., Swanson, C., Scott, J., Ilić, A., Li, X., Jonnalagadda, S.R., Soudagar, A.S., Xiong, Y., Batsaikhan, B.O., Jarrett, D., Kumar, N., Shah, M., Lawlor, M., Waters, A., Graham, M., May, R., Ramos, S., Lefdal, S., Cankara, Z., Cano, N., O'Donoghue, B., Borovik, J., Liu, F., Grimstad, J., Alnahlawi, M., Tsihlas, K., Hudson, T., Grigorev, N., Jia, Y., Huang, T., Igwe, T.P., Lebedev, S., Tang, X., Krivokon, I., Garcia, F., Tan, M., Jia, E., Stys, P., Vashishth, S., Liang, Y., Venkatraman, B., Gu, C., Kementsietsidis, A., Zhu, C., Jung, J., Bai, Y., Hosseini, M.J., Ahmed, F., Gupta, A., Yuan, X., Ashraf, S., Nigam, S., Vasudevan, G., Awasthi, P., Gilady, A.M., Mariet, Z., Eskander, R., Li, H., Hu, H., Garrido, G., Schlattner, P., Zhang, G., Saxena, R., Dević, P., Muralidharan, K., Murthy, A., Zhou, Y., Choi, M., Wongpanich, A., Wang, Z., Shah, P., Xu, Y., Huang, Y., Spencer, S., Chen, A., Cohan, J., Wang, J., Tompson, J., Wu, J., Haroun, R., Li, H., Huergo, B., Yang, F., Yin, T., Wendt, J., Bendersky, M., Chaabouni, R., Snaider, J., Ferret, J., Jindal, A., Thompson, T., Xue, A., Bishop, W., Phal, S.M., Sharma, A., Sung, Y., Radhakrishnan, P., Shomrat, M., Ingle, R., Vij, R., Gilmer, J., Istin, M.D., Sobell, S., Lu, Y., Nottage, E., Sadigh, D., Willcock, J., Zhang, T., Xu, S., Brown, S., Lee, K., Wang, G., Zhu, Y., Tay, Y., Kim, C., Gutierrez, A., Sharma, A., Xian, Y., Seo, S., Cui, C., Pochernina, E., Baetu, C., Jastrzębski, K., Ly, M., Elhawaty, M., Suh, D., Sezener, E., Wang, P., Yuen, N., Tucker, G., Cai, J., Yang, Z., Wang, C., Muzio, A., Qian, H., Yoo, J., Lockhart, D., McKee, K.R., Guo, M., Mehrotra, M., Mendonça, A., Mehta, S.V., Ben, S., Tekur, C., Mu, J., Zhu, M., Krakovna, V., Lee, H., Maschinot, A., Cevey, S., Choe, H., Bai, A., Srinivasan, H., Gasaway, D., Young, N., Siegler, P., Holtmann-Rice, D., Piratla, V., Baumli, K., Yogev, R., Hofer, A., van Hasselt, H., Grant, S., Chervonyi, Y., Silver, D., Hogue, A., Agarwal, A., Wang, K., Singh, P., Flynn, F., Lipschultz, J., David, R., Bellot, L., Yang, Y.Y., Le, L., Graziano, F., Olszewska, K., Hui, K., Maurya, A., Parotsidis, N., Chen, W., Oguntebi, T., Kelley, J., Baddepudi, A., Mauerer, J., Shaw, G., Siegman, A., Yang, L., Shetty, S., Roy, S., Song, Y., Stokowiec, W., Burnell, R., Savant, O., Busa-Fekete, R., Miao, J., Ghosh, S., MacDermed, L., Lippe, P., Dektiarev, M., Behrman, Z., Mentzer, F., Nguyen, K., Wei, M., Verma, S., Knutsen, C., Dasari, S., Yan, Z., Mitrichev, P., Wang, X., Shejwalkar, V., Austin, J., Sunkara, S., Potti, N., Virin, Y., Wright, C., Liu, G., Riva, O., Pot, E., Kochanski, G., Le, Q., Balasubramaniam, G., Dhar, A., Liao, Y., Bloniarz, A., Shukla, D., Cole, E., Lee, J., Zhang, S., Kafle, S., Vashishtha, S., Mahmoudieh, P., Chen, G., Hoffmann, R., Srinivasan, P., Lago, A.D., Shalom, Y.B., Wang, Z., Elabd, M., Sharma, A., Oh, J., Kothawade, S., Le, M., Monteiro, M., Yang, S., Alarakyia, K., Geirhos, R., Mincu, D., Garnes, H., Kobayashi, H., Mariooryad, S., Krasowiak, K., Zhixin, Lai, Mourad, S., Wang, M., Bu, F., Aharoni, O., Chen, G., Goyal, A.,

Zubov, V., Bapna, A., Dabir, E., Kothari, N., Lamerigts, K., Cao, N.D., Shar, J., Yew, C., Kulkarni, N., Mahaarachchi, D., Joshi, M., Zhu, Z., Lichtarge, J., Zhou, Y., Muckenhirn, H., Selo, V., Vinyals, O., Chen, P., Brohan, A., Mehta, V., Cogan, S., Wang, R., Geri, T., Ko, W.J., Chen, W., Viola, F., Shivam, K., Wang, L., Elish, M.C., Popa, R.A., Pereira, S., Liu, J., Koster, R., Kim, D., Zhang, G., Ebrahimi, S., Talukdar, P., Zheng, Y., Poklukar, P., Mikhalap, A., Johnson, D., Vijayakumar, A., Omernick, M., Dibb, M., Dubey, A., Hu, Q., Suman, A., Aggarwal, V., Kornakov, I., Xia, F., Lowe, W., Kolganov, A., Xiao, T., Nikolaev, V., Hemingray, S., Li, B., Iljazi, J., Rybiński, M., Sandhu, B., Lu, P., Luong, T., Jenatton, R., Govindaraj, V., Hui, Li, Dulac-Arnold, G., Park, W., Wang, H., Modi, A., Pouget-Abadie, J., Greller, K., Gupta, R., Berry, R., Ramachandran, P., Xie, J., McCafferty, L., Wang, J., Gupta, K., Lim, H., Bratanič, B., Brock, A., Akolzin, I., Sproch, J., Karliner, D., Kim, D., Goedeckemeyer, A., Shazeer, N., Schmid, C., Calandriello, D., Bhatia, P., Choromanski, K., Montgomery, C., Dua, D., Ramalho, A., King, H., Gao, Y., Nguyen, L., Lindner, D., Pitta, D., Johnson, O., Salama, K., Ardila, D., Han, M., Farnese, E., Odoom, S., Wang, Z., Ding, X., Rink, N., Smith, R., Lehri, H.T., Cohen, E., Vats, N., He, T., Gopavarapu, P., Paszke, A., Patel, M., Gansbeke, W.V., Loher, L., Castro, L., Voitovich, M., von Glehn, T., George, N., Niklaus, S., Eaton-Rosen, Z., Rakićević, N., Jue, E., Perel, S., Zhang, C., Bahat, Y., Pouget, A., Xing, Z., Huot, F., Shenoy, A., Bos, T., Coriou, V., Richter, B., Noy, N., Wang, Y., Ontanon, S., Qin, S., Makarchuk, G., Hassabis, D., Li, Z., Sharma, M., Venkatesan, K., Kemaev, I., Daniel, R., Huang, S., Shah, S., Ponce, O., Warren, Chen, Faruqui, M., Wu, J., Andačić, S., Payrits, S., McDuff, D., Hume, T., Cao, Y., Tessler, M., Wang, Q., Wang, Y., Rendulic, I., Agustsson, E., Johnson, M., Lando, T., Howard, A., Padmanabhan, S.G.S., Daswani, M., Banino, A., Kilgore, M., Heek, J., Ji, Z., Caceres, A., Li, C., Kassner, N., Vlaskin, A., Liu, Z., Grills, A., Hou, Y., Sukkerd, R., Cheon, G., Shetty, N., Markeeva, L., Stanczyk, P., Iyer, T., Gong, Y., Gao, S., Gopalakrishnan, K., Blyth, T., Reynolds, M., Bhoopchand, A., Bilenko, M., Gharibian, D., Zayats, V., Faust, A., Singh, A., Ma, M., Jiao, H., Vijayanarasimhan, S., Aroyo, L., Yadav, V., Chakera, S., Kakarla, A., Meshram, V., Gregor, K., Botea, G., Senter, E., Jia, D., Kovacs, G., Sharma, N., Baur, S., Kang, K., He, Y., Zhuo, L., Kostelac, M., Laish, I., Peng, S., O'Bryan, L., Kasenberg, D., Rao, G.R., Leurent, E., Zhang, B., Stevens, S., Salazar, A., Zhang, Y., Lobov, I., Walker, J., Porter, A., Redshaw, M., Ke, H., Rao, A., Lee, A., Lam, H., Moffitt, M., Kim, J., Qiao, S., Koo, T., Dadashi, R., Song, X., Sundararajan, M., Xu, P., Kawamoto, C., Zhong, Y., Barbu, C., Reddy, A., Verzetti, M., Li, L., Papamakarios, G., Klimczak-Plucińska, H., Cassin, M., Kavukcuoglu, K., Swavely, R., Vaucher, A., Zhao, J., Hemsley, R., Tschannen, M., Ge, H., Menghani, G., Yu, Y., Ha, N., He, W., Wu, X., Song, M., Sterneck, R., Zinke, S., Calian, D.A., Marsden, A., Ruiz, A.C., Hessel, M., Gueta, A., Lee, B., Farris, B., Gupta, M., Li, Y., Saleh, M., Misra, V., Xiao, K., Mendolicchio, P., Buttimore, G., Krayvanova, V., Nayakanti, N., Wiethoff, M., Pande, Y., Mirhoseini, A., Lao, N., Liu, J., Hua, Y., Chen, A., Malkov, Y., Kalashnikov, D., Gupta, S., Audhkhasi, K., Zhai, Y., Kopalle, S., Jain, P., Ofek, E., Meyer, C., Baatarsukh, K., Strejček, H., Qian, J., Freedman, J., Figueira, R., Sokolik, M., Bachem, O., Lin, R., Kharrat, D., Hidey, C., Xu, P., Duan, D., Li, Y., Ersoy, M., Everett, R., Cen, K., Santamaria-Fernandez, R., Taubenfeld, A., Mackinnon, I., Deng, L., Zablotskaia, P., Viswanadha, S., Goel, S., Yates, D., Deng, Y., Choy, P., Chen, M., Sinha, A., Mossin, A., Wang, Y., Szlam, A., Hao, S., Rubenstein, P.K., Toksoz-Exley, M., Aperghis, M., Zhong, Y., Ahn, J., Isard, M., Lacombe, O., Luisier, F., Anastasiou,

C., Kalley, Y., Prabhu, U., Dunleavy, E., Bijwadia, S., Mao-Jones, J., Chen, K., Pasumarthi, R., Wood, E., Dostmohamed, A., Hurley, N., Simsa, J., Parrish, A., Pajarskas, M., Harvey, M., Skopek, O., Kochinski, Y., Rey, J., Rieser, V., Zhou, D., Lee, S.J., Acharya, T., Li, G., Jiang, J., Zhang, X., Gipson, B., Mahintorabi, E., Gelmi, M., Khajehnouri, N., Yeh, A., Lee, K., Matthey, L., Baker, L., Pham, T., Fu, H., Pak, A., Gupta, P., Vasconcelos, C., Sadovsky, A., Walker, B., Hsiao, S., Zochbauer, P., Marzoca, A., Velan, N., Zeng, J., Baechler, G., Driess, D., Jain, D., Huang, Y., Tao, L., Maggs, J., Levine, N., Schneider, J., Gemzer, E., Petit, S., Han, S., Fisher, Z., Zelle, D., Biles, C., Ie, E., Fadeeva, A., Liu, C., Franco, J.V., Collister, A., Zhang, H., Wang, R., Zhao, R., Kieliger, L., Shuster, K., Zhu, R., Gong, B., Chan, L., Sun, R., Basu, S., Zimmermann, R., Hayes, J., Bapna, A., Snoek, J., Yang, W., Datta, P., Abdallah, J.A., Kilgour, K., Li, L., Mah, S., Jun, Y., Rivière, M., Karmarkar, A., Spalink, T., Huang, T., Gonzalez, L., Tran, D.H., Nowak, A., Palowitch, J., Chadwick, M., Talius, E., Mehta, H., Sellam, T., Fränken, P., Nicosia, M., He, K., Kini, A., Amos, D., Basu, S., Jobe, H., Shaw, E., Xu, Q., Evans, C., Ikeda, D., Yan, C., Jin, L., Wang, L., Yadav, S., Labzovsky, I., Sampath, R., Ma, A., Schumann, C., Siddhant, A., Shah, R., Youssef, J., Agarwal, R., Dabney, N., Tonioni, A., Ambar, M., Li, J., Guyon, I., Li, B., Soergel, D., Fang, B., Karadzhov, G., Udrescu, C., Trinh, T., Raunak, V., Noury, S., Guo, D., Gupta, S., Finkelstein, M., Petek, D., Liang, L., Billock, G., Sun, P., Wood, D., Song, Y., Yu, X., Matejovicova, T., Cohen, R., Andra, K., D'Ambrosio, D., Deng, Z., Nallatamby, V., Songhori, E., Dangovski, R., Lampinen, A., Botadra, P., Hillier, A., Cao, J., Baddi, N., Kuncoro, A., Yoshino, T., Bhagatwala, A., Ranzato, M., Schaeffer, R., Liu, T., Ye, S., Sarvana, O., Nham, J., Kuang, C., Gao, I., Baek, J., Mittal, S., Wahid, A., Gergely, A., Ni, B., Feldman, J., Muir, C., Lamblin, P., Macherey, W., Dyer, E., Kilpatrick, L., Campos, V., Bhutani, M., Fort, S., Ahmad, Y., Severyn, A., Chatziprimou, K., Ferludin, O., Dimarco, M., Kusupati, A., Heyward, J., Bahir, D., Villela, K., Millican, K., Marcus, D., Bahargam, S., Unlu, C., Roth, N., Wei, Z., Gopal, S., Ghoshal, D., Lee, E., Lin, S., Lees, J., Lee, D., Hosseini, A., Fan, C., Neel, S., Wu, M., Altun, Y., Cai, H., Piqueras, E., Woodward, J., Bissacco, A., Haykal, S., Bordbar, M., Sundaram, P., Hodkinson, S., Toyama, D., Polovets, G., Myers, A., Sinha, A., Levinboim, T., Krishnakumar, K., Chhaparia, R., Sholokhova, T., Gundavarapu, N.B., Jawahar, G., Qureshi, H., Hu, J., Momchev, N., Rahtz, M., Wu, R., S, A.P., Dhamdhere, K., Guo, M., Gupta, U., Eslami, A., Schain, M., Blokzijl, M., Welling, D., Orr, D., Bolelli, L., Perez-Nieves, N., Sirotenko, M., Prasad, A., Kar, A., Pigem, B.D.B., Terzi, T., Weisz, G., Ghosh, D., Mavalankar, A., Madeka, D., Daugaard, K., Adam, H., Shah, V., Berman, D., Tran, M., Baker, S., Andrejczuk, E., Chole, G., Raboshchuk, G., Mirzazadeh, M., Kagohara, T., Wu, S., Schallhart, C., Orlando, B., Wang, C., Rrustemi, A., Xiong, H., Liu, H., Vezer, A., Ramsden, N., yiin Chang, S., Mudgal, S., Li, Y., Vieillard, N., Hoshen, Y., Ahmad, F., Slone, A., Hua, A., Potikha, N., Rossini, M., Stritar, J., Prakash, S., Wang, Z., Dong, X., Nazari, A., Nehoran, E., Tekelioglu, K., Li, Y., Badola, K., Funkhouser, T., Li, Y., Yerram, V., Ganeshan, R., Formoso, D., Langner, K., Shi, T., Li, H., Yamamori, Y., Panda, A., Saade, A., Scarpati, A.S., Breaux, C., Carey, C., Zhou, Z., Hsieh, C.J., Bridgers, S., Butryna, A., Gupta, N., Tulsyan, V., Woo, S., Eltyshev, E., Grathwohl, W., Parks, C., Benjamin, S., Panigrahy, R., Dodhia, S., Freitas, D.D., Sauer, C., Song, W., Alet, F., Tolins, J., Paduraru, C., Zhou, X., Albert, B., Zhang, Z., Shu, L., Bansal, M., Nguyen, S., Globerson, A., Xiao, O., Manyika, J., Hennigan, T., Rong, R., Matak, J., Bakalov, A., Sharma, A., Sinopalnikov, D., Pierson, A., Roller, S., Brown, G., Gao, M.,

Fukuzawa, T., Ghafouri, A., Vassigh, K., Barr, I., Wang, Z., Korsun, A., Jayaram, R., Ren, L., Zaman, T., Khan, S., Lunts, Y., Deutsch, D., Uthus, D., Katz, N., Samsikova, M., Khalifa, A., Sethi, N., Sun, J., Tang, L., Alon, U., Luo, X., Yu, D., Nayyar, A., Petrini, B., Truong, W., Hellendoorn, V., Chinaev, N., Alberti, C., Wang, W., Hu, J., Mirrokni, V., Balashankar, A., Aharon, A., Mehta, A., Iscen, A., Kready, J., Manning, L., Mohananey, A., Chen, Y., Tripathi, A., Wu, A., Petrovski, I., Hwang, D., Baeuml, M., Chandrakaladharan, S., Liu, Y., Coaguila, R., Chen, M., Ma, S., Tafti, P., Tatineni, S., Spitz, T., Ye, J., Vicol, P., Rosca, M., Puigdomènech, A., Yahav, Z., Ghemawat, S., Lin, H., Kirk, P., Nabulsi, Z., Brin, S., Bohnet, B., Caluwaerts, K., Veerubhotla, A.S., Zheng, D., Dai, Z., Petrov, P., Xu, Y., Mehran, R., Xu, Z., Zintgraf, L., Choi, J., Hombaiah, S.A., Thoppilan, R., Reddi, S., Lew, L., Li, L., Webster, K., Sawhney, K., Lamprou, L., Shakeri, S., Lunayach, M., Chen, J., Bagri, S., Salcianu, A., Chen, Y., Donchev, Y., Magister, C., Nørly, S., Rodrigues, V., Izo, T., Noga, H., Zou, J., Köppe, T., Zhou, W., Lee, K., Long, X., Eisenbud, D., Chen, A., Schenck, C., To, C.M., Zhong, P., Taropa, E., Truong, M., Levy, O., Martins, D., Zhang, Z., Semturs, C., Zhang, K., Yakubovich, A., Moreno, P., McConnaughey, L., Lu, D., Redmond, S., Weerts, L., Bitton, Y., Refice, T., Lacasse, N., Conmy, A., Tallec, C., Odell, J., Forbes-Pollard, H., Socala, A., Hoech, J., Kohli, P., Walton, A., Wang, R., Sazanovich, M., Zhu, K., Kapishnikov, A., Galt, R., Denton, M., Murdoch, B., Sikora, C., Mohamed, K., Wei, W., First, U., McConnell, T., Cobo, L.C., Qin, J., Avrahami, T., Balle, D., Watanabe, Y., Louis, A., Kraft, A., Ariafar, S., Gu, Y., Rives, E., Yoon, C., Rusu, A., Cobon-Kerr, J., Hahn, C., Luo, J., Yuvein, Zhu, Ahuja, N., Benenson, R., Kaufman, R.L., Yu, H., Hightower, L., Zhang, J., Ni, D., Hendricks, L.A., Wang, G., Yona, G., Jain, L., Barrio, P., Bhupatiraju, S., Velusamy, S., Dafoe, A., Riedel, S., Thomas, T., Yuan, Z., Bellaiche, M., Panthaplackel, S., Kloboves, K., Jauhari, S., Akbulut, C., Davchev, T., Gladchenko, E., Madras, D., Chuklin, A., Hill, T., Yuan, Q., Madhavan, M., Leonhard, L., Scandinaro, D., Chen, Q., Niu, N., Douillard, A., Damoc, B., Onoe, Y., Pedregosa, F., Bertsch, F., Leichner, C., Pagadora, J., Malmaud, J., Ponda, S., Twigg, A., Duzhyi, O., Shen, J., Wang, M., Garg, R., Chen, J., Evci, U., Lee, J., Liu, L., Kojima, K., Yamaguchi, M., Rajendran, A., Piergiovanni, A., Rajendran, V.K., Fornoni, M., Ibagon, G., Ragan, H., Khan, S.M., Blitzer, J., Bunner, A., Sun, G., Kosakai, T., Lundberg, S., Elue, N., Guu, K., Park, S., Park, J., Narayanaswamy, A., Wu, C., Mudigonda, J., Cohn, T., Mu, H., Kumar, R., Graesser, L., Zhang, Y., Killam, R., Zhuang, V., Giménez, M., Jishi, W.A., Ley-Wild, R., Zhai, A., Osawa, K., Cedillo, D., Liu, J., Upadhyay, M., Sieniek, M., Sharma, R., Paine, T., Angelova, A., Addepalli, S., Parada, C., Majumder, K., Lamp, A., Kumar, S., Deng, X., Myaskovsky, A., Sabolić, T., Dudek, J., York, S., de Chaumont Quitry, F., Nie, J., Cattle, D., Gunjan, A., Piot, B., Khawaja, W., Bang, S., Wang, S., Khodadadeh, S., R, R., Rawlani, P., Powell, R., Lee, K., Griesser, J., Oh, G., Magalhaes, C., Li, Y., Tokumine, S., Vogel, H.N., Hsu, D., BC, A., Jindal, D., Cohen, M., Yang, Z., Yuan, J., de Cesare, D., Bruguier, T., Xu, J., Roy, M., Jacovi, A., Belov, D., Arya, R., Meadowlark, P., Cohen-Ganor, S., Ye, W., Morris-Suzuki, P., Banzal, P., Song, G., Ponnuramu, P., Zhang, F., Scrivener, G., Zaiem, S., Rochman, A.R., Han, K., Ghazi, B., Lee, K., Drath, S., Suo, D., Girgis, A., Shenoy, P., Nguyen, D., Eck, D., Gupta, S., Yan, L., Carreira, J., Gulati, A., Sang, R., Mirylenka, D., Cooney, E., Chou, E., Ling, M., Fan, C., Coleman, B., Tubone, G., Kumar, R., Baldridge, J., Hernandez-Campos, F., Lazaridou, A., Besley, J., Yona, I., Bulut, N., Wellens, Q., Pierigiovanni, A., George, J., Green, R., Han, P., Tao, C., Clark, G., You, C., Abdolmaleki, A., Fu,

J., Chen, T., Chaugule, A., Chandorkar, A., Rahman, A., Thompson, W., Koanantakool, P., Bernico, M., Ren, J., Vlasov, A., Vassilvitskii, S., Kula, M., Liang, Y., Kim, D., Huang, Y., Ye, C., Lepikhin, D., Helmholz, W.: Gemini 2.5: Pushing the frontier with advanced reasoning, multimodality, long context, and next generation agentic capabilities (2025), https://arxiv.org/abs/2507.06261

19. Cui, C., Huang, J., Wang, S., Zheng, L., Kong, Q., Zeng, Z.: Agentic reward modeling: Verifying gui agent via online proactive interaction (2026), https://arxiv.org/abs/2602.00575
20. Dai, Y., Ramakrishnan, K., Gu, J., Fernandez, M., Luo, Y., Prabhu, V., Hu, Z., Savarese, S., Xiong, C., Chen, Z., Xu, R.: Scuba: Salesforce computer use benchmark (2025), https://arxiv.org/abs/2509.26506
21. Dai, Y., Ramakrishnan, K., Gu, J., Fernandez, M., Luo, Y., Prabhu, V., Hu, Z., Savarese, S., Xiong, C., Chen, Z., Xu, R.: Scuba: Salesforce computer use benchmark (2025), https://arxiv.org/abs/2509.26506
22. Deitke, M., Clark, C., Lee, S., Tripathi, R., Yang, Y., Park, J.S., Salehi, M., Muennighoff, N., Lo, K., Soldaini, L., Lu, J., Anderson, T., Bransom, E., Ehsani, K., Ngo, H., Chen, Y., Patel, A., Yatskar, M., Callison-Burch, C., Head, A., Hendrix, R., Bastani, F., VanderBilt, E., Lambert, N., Chou, Y., Chheda, A., Sparks, J., Skjonsberg, S., Schmitz, M., Sarnat, A., Bischoff, B., Walsh, P., Newell, C., Wolters, P., Gupta, T., Zeng, K.H., Borchardt, J., Groeneveld, D., Nam, C., Lebrecht, S., Wittlif, C., Schoenick, C., Michel, O., Krishna, R., Weihs, L., Smith, N.A., Hajishirzi, H., Girshick, R., Farhadi, A., Kembhavi, A.: Molmo and pixmo: Open weights and open data for state-of-the-art vision-language models (2024), https://arxiv.org/abs/2409.17146
23. Deng, X., Gu, Y., Zheng, B., Chen, S., Stevens, S., Wang, B., Sun, H., Su, Y.: Mind2web: Towards a generalist agent for the web. In: Advances in Neural Information Processing Systems (NeurIPS) (2023), https://arxiv.org/abs/2306.06070
24. Fang, T., Zhang, H., Zhang, Z., Ma, K., Yu, W., Mi, H., Yu, D.: Webevolver: Enhancing web agent self-improvement with coevolving world model (2025), https://arxiv.org/abs/2504.21024
25. Fu, T., Su, A., Zhao, C., Wang, H., Wu, M., Yu, Z., Hu, F., Shi, M., Dong, W., Wang, J., Chen, Y., Yu, R., Peng, S., Li, M., Huang, N., Wei, H., Yu, J., Xin, Y., Zhao, X., Gu, K., Jiang, P., Zhou, S., Wang, S.: Mano technical report (2025), https://arxiv.org/abs/2509.17336
26. Gao, L., Zhang, L., Gao, P., Liu, W., Luan, J., Xu, M.: Gui-shift: Enhancing vlm-based gui agents through self-supervised reinforcement learning (2025), https://arxiv.org/abs/2505.12493
27. Guo, L., Zhu, B., Tao, Q., Liu, K., Zhao, X., Qin, X., Gao, J., Hao, G.: Agentic lybic: Multi-agent execution system with tiered reasoning and orchestration (2025), https://arxiv.org/abs/2509.11067
28. Hong, W., Wang, W., Lv, Q., Xu, J., Yu, W., Ji, J., Wang, Y., Wang, Z., Dong, Y., Ding, M., Tang, J.: Cogagent: A visual language model for gui agents. In: 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 14281–14290 (2024). https://doi.org/10.1109/CVPR52733.2024.01354
29. Huang, K.H., Qiu, H., Dai, Y., Xiong, C., Wu, C.S.: Gui-kv: Efficient gui agents via kv cache with spatio-temporal awareness (2025), https://arxiv.org/abs/2510.00536
30. Huang, X., Zhou, H., Han, K.: PruneVid: Visual token pruning for efficient video large language models. In: Che, W., Nabende, J., Shutova, E., Pilehvar, M.T. (eds.) Findings of the Association for Computational Linguistics: ACL 2025.

pp. 19959–19973. Association for Computational Linguistics, Vienna, Austria (Jul 2025). https://doi.org/10.18653/v1/2025.findings-acl.1024, https://aclanthology.org/2025.findings-acl.1024/
31. Ilie, A.G.: Uipath screen agent. https://www.uipath.com/ai/research/screen-agent (October 2025), accessed: February 14, 2026
32. Koh, J.Y., Lo, R., Jang, L., Duvvur, V., Lim, M., Huang, P.Y., Neubig, G., Zhou, S., Salakhutdinov, R., Fried, D.: VisualWebArena: Evaluating multimodal agents on realistic visual web tasks. In: Ku, L.W., Martins, A., Srikumar, V. (eds.) Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). pp. 881–905. Association for Computational Linguistics, Bangkok, Thailand (Aug 2024). https://doi.org/10.18653/v1/2024.acl-long.50, https://aclanthology.org/2024.acl-long.50/
33. Li, B., Zhang, H., Zhang, K., Guo, D., Zhang, Y., Zhang, R., Li, F., Liu, Z., Li, C.: Llava-next: What else influences visual instruction tuning beyond data? (May 2024), https://llava-vl.github.io/blog/2024-05-25-llava-next-ablations/
34. Li, K., Meng, Z., Lin, H., Luo, Z., Tian, Y., Ma, J., Huang, Z., Chua, T.S.: Screenspot-pro: Gui grounding for professional high-resolution computer use (2025), https://arxiv.org/abs/2504.07981
35. Li, X., Yu, P., Zhou, C., Schick, T., Levy, O., Zettlemoyer, L., Weston, J., Lewis, M.: Self-alignment with instruction backtranslation (2024), https://arxiv.org/abs/2308.06259
36. Lin, H., Tan, X., Qin, Y., Xu, Z., Shi, Y., Li, Z., Li, G., Cai, S., Cai, S., Fu, C., Li, K., Sun, X.: Cuarewardbench: A benchmark for evaluating reward models on computer-using agent (2025), https://arxiv.org/abs/2510.18596
37. Lin, K.Q., Li, L., Gao, D., Yang, Z., Wu, S., Bai, Z., Lei, S.W., Wang, L., Shou, M.Z.: Showui: One vision-language-action model for gui visual agent. In: Proceedings of the Computer Vision and Pattern Recognition Conference. pp. 19498–19508 (2025)
38. Liu, X., Qin, B., Liang, D., Dong, G., Lai, H., Zhang, H., Zhao, H., Iong, I.L., Sun, J., Wang, J., Gao, J., Shan, J., Liu, K., Zhang, S., Yao, S., Cheng, S., Yao, W., Zhao, W., Liu, X., Liu, X., Chen, X., Yang, X., Yang, Y., Xu, Y., Yang, Y., Wang, Y., Xu, Y., Qi, Z., Dong, Y., Tang, J.: Autoglm: Autonomous foundation agents for guis (2024), https://arxiv.org/abs/2411.00820
39. Liu, Z., Xie, J., Ding, Z., Li, Z., Yang, B., Wu, Z., Wang, X., Sun, Q., Liu, S., Wang, W., Ye, S., Li, Q., Dong, X., Yu, Y., Lu, C., Mo, Y., Yan, Y., Tian, Z., Zhang, X., Huang, Y., Liu, Y., Su, W., Luo, G., Yue, X., Qi, B., Chen, K., Zhou, B., Qiao, Y., Chen, Q., Wang, W.: Scalecua: Scaling open-source computer use agents with cross-platform data. arXiv preprint arXiv:2509.15221 (2025), https://github.com/OpenGVLab/ScaleCUA, preprint
40. Lu, Y., Yang, J., Shen, Y., Awadallah, A.: Omniparser for pure vision based gui agent (2024), https://arxiv.org/abs/2408.00203
41. Lù, X.H., Kazemnejad, A., Meade, N., Patel, A., Shin, D., Zambrano, A., Stańczak, K., Shaw, P., Pal, C.J., Reddy, S.: Agentrewardbench: Evaluating automatic evaluations of web agent trajectories (2025), https://arxiv.org/abs/2504.08942
42. Murty, S., Zhu, H., Bahdanau, D., Manning, C.D.: Nnetnav: Unsupervised learning of browser agents through environment interaction in the wild (2025), https://arxiv.org/abs/2410.02907
43. OpenAI: Introducing GPT-5.2. https://openai.com/index/introducing-gpt-5-2/ (December 2025), accessed: 2026-02-16

44. OpenAI: Openai o3 and o4-mini system card. Tech. rep., OpenAI (April 2025), https://cdn.openai.com/pdf/2221c875-02dc-4789-800b-e7758f3722c1/o3-and-o4-mini-system-card.pdf
45. Ouyang, M., Lin, K.Q., Shou, M.Z., Ng, H.T.: Focusui: Efficient ui grounding via position-preserving visual token selection (2026), https://arxiv.org/abs/2601.03928
46. Pahuja, V., Lu, Y., Rosset, C., Gou, B., Mitra, A., Whitehead, S., Su, Y., Awadallah, A.: Explorer: Scaling exploration-driven web trajectory synthesis for multimodal web agents (2025), https://arxiv.org/abs/2502.11357
47. Pan, J., Zhang, Y., Tomlin, N., Zhou, Y., Levine, S., Suhr, A.: Autonomous evaluation and refinement of digital agents (2024), https://arxiv.org/abs/2404.06474
48. Qi, Z., Liu, X., Iong, I.L., Lai, H., Sun, X., Zhao, W., Yang, Y., Yang, X., Sun, J., Yao, S., Zhang, T., Xu, W., Tang, J., Dong, Y.: Webrl: Training llm web agents via self-evolving online curriculum reinforcement learning (2025), https://arxiv.org/abs/2411.02337
49. Qin, Y., Ye, Y., Fang, J., Wang, H., Liang, S., Tian, S., Zhang, J., Li, J., Li, Y., Huang, S., Zhong, W., Li, K., Yang, J., Miao, Y., Lin, W., Liu, L., Jiang, X., Ma, Q., Li, J., Xiao, X., Cai, K., Li, C., Zheng, Y., Jin, C., Li, C., Zhou, X., Wang, M., Chen, H., Li, Z., Yang, H., Liu, H., Lin, F., Peng, T., Liu, X., Shi, G.: Ui-tars: Pioneering automated gui interaction with native agents (2025), https://arxiv.org/abs/2501.12326
50. Rawles, C., Clinckemaillie, S., Chang, Y., Waltz, J., Lau, G., Fair, M., Li, A., Bishop, W.E., Li, W., Campbell-Ajala, F., Toyama, D.K., Berry, R.J., Tyamagundlu, D., Lillicrap, T.P., Riva, O.: Androidworld: A dynamic benchmarking environment for autonomous agents. In: The Thirteenth International Conference on Learning Representations (2025), https://openreview.net/forum?id=il5yUQsrjC
51. Ryoo, M.S., Zhou, H., Kendre, S., Qin, C., Xue, L., Shu, M., Park, J., Ranasinghe, K., Savarese, S., Xu, R., Xiong, C., Niebles, J.C.: xgen-mm-vid (blip-3-video): You only need 32 tokens to represent a video even in vlms (2025), https://arxiv.org/abs/2410.16267
52. Seed, B., :, Chen, J., Fan, T., Liu, X., Liu, L., Lin, Z., Wang, M., Wang, C., Wei, X., Xu, W., Yuan, Y., Yue, Y., Yan, L., Yu, Q., Zuo, X., Zhang, C., Zhu, R., An, Z., Bai, Z., Bao, Y., Bin, X., Chen, J., Chen, F., Chen, H., Chen, R., Chen, L., Chen, Z., Chen, J., Chen, S., Chen, K., Chen, Z., Chen, J., Chen, J., Chi, J., Dai, W., Dai, N., Dai, J., Dou, S., Du, Y., Du, Z., Duan, J., Dun, C., Fan, T.H., Feng, J., Feng, J., Feng, Z., Fu, Y., Fu, W., Fu, H., Ge, H., Guo, H., Han, M., Han, L., Hao, W., Hao, X., He, Q., He, J., He, F., Heng, W., Hong, Z., Hou, Q., Hu, L., Hu, S., Hu, N., Hua, K., Huang, Q., Huang, Z., Huang, H., Huang, Z., Huang, T., Huang, W., Jia, W., Jia, B., Jia, X., Jiang, Y., Jiang, H., Jiang, Z., Jiang, K., Jiang, C., Jiao, J., Jin, X., Jin, X., Lai, X., Li, Z., Li, X., Li, L., Li, H., Li, Z., Wan, S., Wang, Y., Li, Y., Li, C., Li, N., Li, S., Li, X., Li, X., Li, A., Li, Y., Liang, N., Liang, X., Lin, H., Lin, W., Lin, Y., Liu, Z., Liu, G., Liu, G., Liu, C., Liu, Y., Liu, G., Liu, J., Liu, C., Liu, D., Liu, K., Liu, S., Liu, Q., Liu, Y., Liu, K., Liu, G., Liu, B., Long, R., Lou, W., Lou, C., Luo, X., Luo, Y., Lv, C., Lv, H., Ma, B., Ma, Q., Ma, H., Ma, Y., Ma, J., Ma, W., Ma, T., Mao, C., Min, Q., Nan, Z., Ning, G., Ou, J., Pan, H., Pang, R., Peng, Y., Peng, T., Qian, L., Qian, L., Qiao, M., Qu, M., Ren, C., Ren, H., Shan, Y., Shen, W., Shen, K., Shen, K., Sheng, G., Shi, J., Shi, W., Shi, G., Cao, S.S., Song, Y., Song, Z., Su, J., Sun, Y., Sun, T., Sun, Z., Wan, B., Wang, Z., Wang, X., Wang, X., Wang, S., Wang, J., Wang, Q.,

Wang, C., Wang, S., Wang, Z., Wang, C., Wang, J., Wang, S., Wang, X., Wang, Z., Wang, Y., Wang, W., Wang, T., Wei, C., Wei, H., Wei, Z., Wei, S., Wu, Z., Wu, Y., Wu, Y., Wu, B., Wu, S., Wu, J., Wu, N., Wu, S., Wu, J., Xi, C., Xia, F., Xian, Y., Xiang, L., Xiang, B., Xiao, B., Xiao, Z., Xiao, X., Xiao, Y., Xin, C., Xin, S., Xiong, Y., Xu, J., Xu, Z., Xu, C., Xu, J., Xu, Y., Xu, W., Xu, Y., Xu, S., Yan, S., Yan, S., Yang, Q., Yang, X., Yang, T., Yang, Y., Yang, Y., Yang, X., Yang, Z., Yang, G., Yang, Y., Yao, X., Yi, B., Yin, F., Yin, J., Ying, Z., Yu, X., Yu, H., Yu, S., Yu, M., Yu, H., Yuan, S., Yuan, J., Zeng, Y., Zhan, T., Zhang, Z., Zhang, Y., Zhang, M., Zhang, W., Zhang, R., Zhang, Z., Zhang, T., Zhang, X., Zhang, Z., Zhang, S., Zhang, W., Zhang, X., Zhang, Y., Zhang, Y., Zhang, G., Zhang, H., Zhang, Y., Zheng, R., Zheng, N., Zheng, Z., Zheng, Y., Zheng, C., Zhi, X., Zhong, W., Zhong, C., Zhong, Z., Zhong, B., Zhou, X., Zhou, N., Zhou, H., Zhu, H., Zhu, D., Zhu, W., Zuo, L.: Seed1.5-thinking: Advancing superb reasoning models with reinforcement learning (2025), https://arxiv.org/abs/2504.13914

53. Shang, Y., Cai, M., Xu, B., Lee, Y.J., Yan, Y.: Llava-prumerge: Adaptive token reduction for efficient large multimodal models (2026), https://arxiv.org/abs/2403.15388
54. Shao, K., Tao, K., Qin, C., You, H., Sui, Y., Wang, H.: Holitom: Holistic token merging for fast video large language models (2025), https://arxiv.org/abs/2505.21334
55. Song, L., Dai, Y., Prabhu, V., Zhang, J., Shi, T., Li, L., Li, J., Savarese, S., Chen, Z., Zhao, J., Xu, R., Xiong, C.: Coact-1: Computer-using multi-agent system with coding actions (2026), https://arxiv.org/abs/2508.03923
56. Sun, Q., Cheng, K., Ding, Z., Jin, C., Wang, Y., Xu, F., Wu, Z., Jia, C., Chen, L., Liu, Z., Kao, B., Li, G., He, J., Qiao, Y., Wu, Z.: Os-genesis: Automating gui agent trajectory construction via reverse task synthesis. In: Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL). pp. 5555–5579 (2025), https://aclanthology.org/2025.acl-long.277/
57. Sun, Z., Liu, Z., Zang, Y., Cao, Y., Dong, X., Wu, T., Lin, D., Wang, J.: Seagent: Self-evolving computer use agent with autonomous learning from experience (2025), https://arxiv.org/abs/2508.04700
58. Team, G., Kamath, A., Ferret, J., Pathak, S., Vieillard, N., Merhej, R., Perrin, S., Matejovicova, T., Ramé, A., Rivière, M., Rouillard, L., Mesnard, T., Cideron, G., bastien Grill, J., Ramos, S., Yvinec, E., Casbon, M., Pot, E., Penchev, I., Liu, G., Visin, F., Kenealy, K., Beyer, L., Zhai, X., Tsitsulin, A., Busa-Fekete, R., Feng, A., Sachdeva, N., Coleman, B., Gao, Y., Mustafa, B., Barr, I., Parisotto, E., Tian, D., Eyal, M., Cherry, C., Peter, J.T., Sinopalnikov, D., Bhupatiraju, S., Agarwal, R., Kazemi, M., Malkin, D., Kumar, R., Vilar, D., Brusilovsky, I., Luo, J., Steiner, A., Friesen, A., Sharma, A., Sharma, A., Gilady, A.M., Goedeckemeyer, A., Saade, A., Feng, A., Kolesnikov, A., Bendebury, A., Abdagic, A., Vadi, A., György, A., Pinto, A.S., Das, A., Bapna, A., Miech, A., Yang, A., Paterson, A., Shenoy, A., Chakrabarti, A., Piot, B., Wu, B., Shahriari, B., Petrini, B., Chen, C., Lan, C.L., Choquette-Choo, C.A., Carey, C., Brick, C., Deutsch, D., Eisenbud, D., Cattle, D., Cheng, D., Paparas, D., Sreepathihalli, D.S., Reid, D., Tran, D., Zelle, D., Noland, E., Huizenga, E., Kharitonov, E., Liu, F., Amirkhanyan, G., Cameron, G., Hashemi, H., Klimczak-Plucińska, H., Singh, H., Mehta, H., Lehri, H.T., Hazimeh, H., Ballantyne, I., Szpektor, I., Nardini, I., Pouget-Abadie, J., Chan, J., Stanton, J., Wieting, J., Lai, J., Orbay, J., Fernandez, J., Newlan, J., yeong Ji, J., Singh, J., Black, K., Yu, K., Hui, K., Vodrahalli, K., Greff, K., Qiu, L., Valentine, M., Coelho, M., Ritter, M., Hoffman, M., Watson, M., Chaturvedi,

M., Moynihan, M., Ma, M., Babar, N., Noy, N., Byrd, N., Roy, N., Momchev, N., Chauhan, N., Sachdeva, N., Bunyan, O., Botarda, P., Caron, P., Rubenstein, P.K., Culliton, P., Schmid, P., Sessa, P.G., Xu, P., Stanczyk, P., Tafti, P., Shivanna, R., Wu, R., Pan, R., Rokni, R., Willoughby, R., Vallu, R., Mullins, R., Jerome, S., Smoot, S., Girgin, S., Iqbal, S., Reddy, S., Sheth, S., Põder, S., Bhatnagar, S., Panyam, S.R., Eiger, S., Zhang, S., Liu, T., Yacovone, T., Liechty, T., Kalra, U., Evci, U., Misra, V., Roseberry, V., Feinberg, V., Kolesnikov, V., Han, W., Kwon, W., Chen, X., Chow, Y., Zhu, Y., Wei, Z., Egyed, Z., Cotruta, V., Giang, M., Kirk, P., Rao, A., Black, K., Babar, N., Lo, J., Moreira, E., Martins, L.G., Sanseviero, O., Gonzalez, L., Gleicher, Z., Warkentin, T., Mirrokni, V., Senter, E., Collins, E., Barral, J., Ghahramani, Z., Hadsell, R., Matias, Y., Sculley, D., Petrov, S., Fiedel, N., Shazeer, N., Vinyals, O., Dean, J., Hassabis, D., Kavukcuoglu, K., Farabet, C., Buchatskaya, E., Alayrac, J.B., Anil, R., Dmitry, Lepikhin, Borgeaud, S., Bachem, O., Joulin, A., Andreev, A., Hardin, C., Dadashi, R., Hussenot, L.: Gemma 3 technical report (2025), https://arxiv.org/abs/2503.19786

59. Team, K., Du, A., Yin, B., Xing, B., Qu, B., Wang, B., Chen, C., Zhang, C., Du, C., Wei, C., Wang, C., Zhang, D., Du, D., Wang, D., Yuan, E., Lu, E., Li, F., Sung, F., Wei, G., Lai, G., Zhu, H., Ding, H., Hu, H., Yang, H., Zhang, H., Wu, H., Yao, H., Lu, H., Wang, H., Gao, H., Zheng, H., Li, J., Su, J., Wang, J., Deng, J., Qiu, J., Xie, J., Wang, J., Liu, J., Yan, J., Ouyang, K., Chen, L., Sui, L., Yu, L., Dong, M., Dong, M., Xu, N., Cheng, P., Gu, Q., Zhou, R., Liu, S., Cao, S., Yu, T., Song, T., Bai, T., Song, W., He, W., Huang, W., Xu, W., Yuan, X., Yao, X., Wu, X., Li, X., Zu, X., Zhou, X., Wang, X., Charles, Y., Zhong, Y., Li, Y., Hu, Y., Chen, Y., Wang, Y., Liu, Y., Miao, Y., Qin, Y., Chen, Y., Bao, Y., Wang, Y., Kang, Y., Liu, Y., Dong, Y., Du, Y., Wu, Y., Wang, Y., Yan, Y., Zhou, Z., Li, Z., Jiang, Z., Zhang, Z., Yang, Z., Huang, Z., Huang, Z., Zhao, Z., Chen, Z., Lin, Z.: Kimi-vl technical report (2025), https://arxiv.org/abs/2504.07491
60. Wang, H., Zou, H., Song, H., Feng, J., Fang, J., Lu, J., Liu, L., Luo, Q., Liang, S., Huang, S., Zhong, W., Ye, Y., Qin, Y., Xiong, Y., Song, Y., Wu, Z., Li, A., Li, B., Dun, C., Liu, C., Zan, D., Leng, F., Wang, H., Yu, H., Chen, H., Guo, H., Su, J., Huang, J., Shen, K., Shi, K., Yan, L., Zhao, P., Liu, P., Ye, Q., Zheng, R., Xin, S., Zhao, W.X., Heng, W., Huang, W., Wang, W., Qin, X., Lin, Y., Wu, Y., Chen, Z., Wang, Z., Zhong, B., Zhang, X., Li, X., Li, Y., Zhao, Z., Jiang, C., Wu, F., Zhou, H., Pang, J., Han, L., Liu, Q., Ma, Q., Liu, S., Cai, S., Fu, W., Liu, X., Wang, Y., Zhang, Z., Zhou, B., Li, G., Shi, J., Yang, J., Tang, J., Li, L., Han, Q., Lu, T., Lin, W., Tong, X., Li, X., Zhang, Y., Miao, Y., Jiang, Z., Li, Z., Zhao, Z., Li, C., Ma, D., Lin, F., Zhang, G., Yang, H., Guo, H., Zhu, H., Liu, J., Du, J., Cai, K., Li, K., Yuan, L., Han, M., Wang, M., Guo, S., Cheng, T., Ma, X., Xiao, X., Huang, X., Chen, X., Du, Y., Chen, Y., Wang, Y., Li, Z., Yang, Z., Zeng, Z., Jin, C., Li, C., Chen, H., Chen, H., Chen, J., Zhao, Q., Shi, G.: Ui-tars-2 technical report: Advancing gui agent with multi-turn reinforcement learning (2025), https://arxiv.org/abs/2509.02544
61. Wang, S., Fu, P., Zhang, R., Zhang, S., Xi, X., Yang, J., Qin, B., Huang, Y., Luo, Z., Luan, J.: Gaia: A data flywheel system for training gui test-time scaling critic models (2026), https://arxiv.org/abs/2601.18197
62. Wang, S., Jin, J., Wang, X., Song, L., Fu, R., Wang, H., Ge, Z., Lu, Y., Cheng, X.: Video-thinker: Sparking "thinking with videos" via reinforcement learning (2025), https://arxiv.org/abs/2510.23473
63. Wang, W., Gao, Z., Gu, L., Pu, H., Cui, L., Wei, X., Liu, Z., Jing, L., Ye, S., Shao, J., Wang, Z., Chen, Z., Zhang, H., Yang, G., Wang, H., Wei, Q., Yin, J.,

Li, W., Cui, E., Chen, G., Ding, Z., Tian, C., Wu, Z., Xie, J., Li, Z., Yang, B., Duan, Y., Wang, X., Hou, Z., Hao, H., Zhang, T., Li, S., Zhao, X., Duan, H., Deng, N., Fu, B., He, Y., Wang, Y., He, C., Shi, B., He, J., Xiong, Y., Lv, H., Wu, L., Shao, W., Zhang, K., Deng, H., Qi, B., Ge, J., Guo, Q., Zhang, W., Zhang, S., Cao, M., Lin, J., Tang, K., Gao, J., Huang, H., Gu, Y., Lyu, C., Tang, H., Wang, R., Lv, H., Ouyang, W., Wang, L., Dou, M., Zhu, X., Lu, T., Lin, D., Dai, J., Su, W., Zhou, B., Chen, K., Qiao, Y., Wang, W., Luo, G.: Internvl3.5: Advancing open-source multimodal models in versatility, reasoning, and efficiency (2025), https://arxiv.org/abs/2508.18265

64. Wang, X., Wang, B., Lu, D., Yang, J., Xie, T., Wang, J., Deng, J., Guo, X., Xu, Y., Wu, C.H., Shen, Z., Li, Z., Li, R., Li, X., Chen, J., Zheng, B., Li, P., Lei, F., Cao, R., Fu, Y., Shin, D., Shin, M., Hu, J., Wang, Y., Chen, J., Ye, Y., Zhang, D., Du, D., Hu, H., Chen, H., Zhou, Z., Yao, H., Chen, Z., Gu, Q., Wang, Y., Wang, H., Yang, D., Zhong, V., Sung, F., Charles, Y., Yang, Z., Yu, T.: Opencua: Open foundations for computer-use agents (2025), https://arxiv.org/abs/2508.09123
65. Wang, Y., Huzhang, G., Hu, Y., Xia, Y., Lu, S., Chen, Q.G., Xu, Z., Luo, W., Zhang, K., Zhang, L.: Building autonomous gui navigation via agentic-q estimation and step-wise policy optimization (2026), https://arxiv.org/abs/2602.13653
66. Wanyan, Y., Zhang, X., Xu, H., Liu, H., Wang, J., Ye, J., Kou, Y., Yan, M., Huang, F., Yang, X., Dong, W., Xu, C.: Look before you leap: A gui-critic-r1 model for pre-operative error diagnosis in gui automation (2025), https://arxiv.org/abs/2506.04614
67. Wu, Z., Xie, J., Li, Z., Yang, B., Sun, Q., Liu, Z., Liu, Z., Qiao, Y., Yue, X., Wang, Z., Ding, Z.: Os-oracle: A comprehensive framework for cross-platform gui critic models (2025), https://arxiv.org/abs/2512.16295
68. Xiao, H., Wang, G., Chai, Y., Lu, Z., Lin, W., He, H., Fan, L., Bian, L., Hu, R., Liu, L., Ren, S., Wen, Y., Chen, X., Zhou, A., Li, H.: Ui-genie: A self-improving approach for iteratively boosting mllm-based mobile gui agents (2025), https://arxiv.org/abs/2505.21496
69. Xie, T., Deng, J., Li, X., Yang, J., Wu, H., Chen, J., Hu, W., Wang, X., Xu, Y., Wang, Z., Xu, Y., Wang, J., Sahoo, D., Yu, T., Xiong, C.: Scaling computer-use grounding via user interface decomposition and synthesis (2025), https://arxiv.org/abs/2505.13227
70. Xie, T., Zhang, D., Chen, J., Li, X., Zhao, S., Cao, R., Hua, T.J., Cheng, Z., Shin, D., Lei, F., Liu, Y., Xu, Y., Zhou, S., Savarese, S., Xiong, C., Zhong, V., Yu, T.: Osworld: Benchmarking multimodal agents for open-ended tasks in real computer environments (2024)
71. Xu, Z., Zhou, B., Wang, Q., Feng, S., Xiao, J.: Spatio-temporal token pruning for efficient high-resolution gui agents (2026), https://arxiv.org/abs/2602.23235
72. Yang, A., Li, A., Yang, B., Zhang, B., Hui, B., Zheng, B., Yu, B., Gao, C., Huang, C., Lv, C., Zheng, C., Liu, D., Zhou, F., Huang, F., Hu, F., Ge, H., Wei, H., Lin, H., Tang, J., Yang, J., Tu, J., Zhang, J., Yang, J., Yang, J., Zhou, J., Zhou, J., Lin, J., Dang, K., Bao, K., Yang, K., Yu, L., Deng, L., Li, M., Xue, M., Li, M., Zhang, P., Wang, P., Zhu, Q., Men, R., Gao, R., Liu, S., Luo, S., Li, T., Tang, T., Yin, W., Ren, X., Wang, X., Zhang, X., Ren, X., Fan, Y., Su, Y., Zhang, Y., Zhang, Y., Wan, Y., Liu, Y., Wang, Z., Cui, Z., Zhang, Z., Zhou, Z., Qiu, Z.: Qwen3 technical report (2025), https://arxiv.org/abs/2505.09388
73. Yang, C., Su, S., Liu, S., Dong, X., Yu, Y., Su, W., Wang, X., Liu, Z., Zhu, J., Li, H., Wang, W., Qiao, Y., Zhu, X., Dai, J.: Zerogui: Automating online gui learning at zero human cost (2025), https://arxiv.org/abs/2505.23762

74. Yang, P., Ci, H., Shou, M.Z.: macosworld: A multilingual interactive benchmark for GUI agents. In: The Thirty-ninth Annual Conference on Neural Information Processing Systems (2025), https://openreview.net/forum?id=YJxGJP8feU
75. Yang, Y., Li, D., Dai, Y., Yang, Y., Luo, Z., Zhao, Z., Hu, Z., Huang, J., Saha, A., Chen, Z., Xu, R., Pan, L., Savarese, S., Xiong, C., Li, J.: Gta1: Gui test-time scaling agent (2025), https://arxiv.org/abs/2507.05791
76. Ye, J., Zhang, X., Xu, H., Liu, H., Wang, J., Zhu, Z., Zheng, Z., Gao, F., Cao, J., Lu, Z., Liao, J., Zheng, Q., Huang, F., Zhou, J., Yan, M.: Mobile-agent-v3: Fundamental agents for gui automation (2025), https://arxiv.org/abs/2508.15144
77. Ye, J., Wang, Z., Sun, H., Chandrasegaran, K., Durante, Z., Eyzaguirre, C., Bisk, Y., Niebles, J.C., Adeli, E., Fei-Fei, L., Wu, J., Li, M.: T*: Re-thinking temporal search for long-form video understanding (2025), https://arxiv.org/abs/2504.02259
78. Zhang, D., Zhang, S., Yang, Z., Zhu, Z., Zhao, Z., Cao, R., Chen, L., Yu, K.: Progrm: Build better gui agents with progress rewards (2025), https://arxiv.org/abs/2505.18121
79. Zhang, S., Dong, Y., Zhang, J., Kautz, J., Catanzaro, B., Tao, A., Wu, Q., Yu, Z., Liu, G.: Nemotron-research-tool-n1: Exploring tool-using language models with reinforced reasoning (2025), https://arxiv.org/abs/2505.00024
80. Zhang, Y., Tang, S., Li, Z., Han, Z., Tresp, V.: Webarbiter: A principle-guided reasoning process reward model for web agents (2026), https://arxiv.org/abs/2601.21872
81. Zheng, C., Mo, X., Ma, X., Lin, Q., Zhao, Y., Zhu, J., Lou, X., Wang, J., Wang, Z., Liu, W., Zhang, Z., Yu, Y., Zhang, W.: Adaptive milestone reward for gui agents (2026), https://arxiv.org/abs/2602.11524
82. Zheng, Y., Zhang, R., Zhang, J., Ye, Y., Luo, Z., Feng, Z., Ma, Y.: Llamafactory: Unified efficient fine-tuning of 100+ language models. In: Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 3: System Demonstrations). Association for Computational Linguistics, Bangkok, Thailand (2024), http://arxiv.org/abs/2403.13372
83. Zhou, S., Xu, F.F., Zhu, H., Zhou, X., Lo, R., Sridhar, A., Cheng, X., Ou, T., Bisk, Y., Fried, D., Alon, U., Neubig, G.: Webarena: A realistic web environment for building autonomous agents. In: The Twelfth International Conference on Learning Representations (2024), https://openreview.net/forum?id=oKn9c6ytLx