

TABLE I: Unified summary of WAMs in Sec. IV, grouped by application domain. Category labels: **VAM** Video-Action Models, **Uni** Unified WAMs (joint observation and action generation), **WM-Plan** WMs-guided Planning (predictions from WMs and action generated by external planner), **RL** WM-RL (imagination-based RL), **3D** 3D/Geometry-based.

Method	Paradigm	Backbone	Eval / Note	Method	Paradigm	Backbone	Eval / Note
Manipulation							
GR-1 [13]	VAM	GPT	LIBERO, CALVIN	EVA [65]	RL	video WM	RoboTwin 2.0
GR-2 [38]	VAM	GPT, 38M	LIBERO, CALVIN	RISE [154]	RL	latent	contact-rich
PAD [103]	Uni	DiT	LIBERO	VLA-MBPO [155]	RL	video WM	LIBERO-Plus
UWM [19]	Uni	4-mode DiT	LIBERO	GPC [156]	RL	rollout WM	Push-T
VPP [20]	VAM	video pred.	dexterous	WMPO [130]	RL	pixel WM	LIBERO
Seer [60]	VAM	GPT+mask	100K demos	RoboHorizon [47]	RL	LLM+WM	sparse reward
VidMan [67]	VAM	vid. diff.+adapt.	Bridge	World-VLA-Loop [64]	RL	feedback	ManiSkill
Cosmos Policy [21]	VAM	Cosmos-P2	bimanual	IWS [55]	RL	video WM	10-min @15FPS
ViPRA [97]	VAM	flow-match	100-200 demos	ManiGaussian++ [133]	3D	Gaussian	bimanual
GigaWorld-Policy [25]	Uni	action-cent.	DROID	ParticleFormer [132]	DROID	particles	deformable
Fast-WAM [22]	VAM	video pred.	LIBERO, RoboTwin	GWM [135]	3D	VAE+diff	MetaWorld
WorldVLA [39]	Uni	Chameleon AR	LIBERO	PointWorld [157]	3D	point cloud	2M+ traj
Motus [99]	Uni	MoT	RoboTwin	PIN-WM [158]	3D	physics-inf.	non-prehensile
LingBot-VA [24]	Uni	causal-diff MoT	AgiBot	VT-WM [143]	3D	visuo-tactile	contact
World2Act [159]	Uni	LLM+skill	RoboCasa	Cloth-WM [160]	3D	DreamerV2	cloth
Act2Goal [102]	VAM	goal-cond.	RoboTwin 2.0	Audio-WM [142]	3D	acoustic	liquid pour
VLA-JEPA [115]	VAM	JEPA+flow	LIBERO	World4RL [66]	RL	diff. WM	frozen env
LAWM [161]	VAM	latent action	LIBERO	DreamPlan [162]	RL	video WM	VLA finetune
HiLAM [163]	VAM	hier. latent	long-horizon	VAMPO [164]	RL	GRPO/denoise	LIBERO, CALVIN
PRWM [165]	RL	contrastive RL	DROID				
Navigation							
Pathdreamer [2]	WM-Plan	panoramic gen.	R2R	MonoDream [166]	WM-Plan	monocular	R2R-CE, RxR-CE
DreamWalker [167]	WM-Plan	latent+MCTS	R2R, VLN-CE	X-MOBILITY [168]	RL	latent WM env	mobile robot
NWM [169]	WM-Plan	multi-domain	Habitat, Ego4D	InternVLA-N1 [170]	VAM	dual-system latent	R2R-CE
perceptive FDM [171]	WM-Plan	GRU	ANYmal, quadruped	RWM [172]	RL	GRU	A1, quadruped
WMP [173]	RL	RSSM	Isaac				
Autonomous Driving							
DriveDreamer [174]	WM-Plan	video gen.	nuScenes	ADriver-I [175]	Uni	AR img+act	nuScenes
GAIA-1 [68]	WM-Plan	text-video	large-scale	DrivingGPT [176]	Uni	large AR video	NAVSIM
DriveDreamer-2 [177]	WM-Plan	multi-cam. gen.	nuScenes	Doe-1 [178]	Uni	autoregressive	nuScenes
GAIA-2 [179]	WM-Plan	multi-cam. scaled	OpenDV	OccLLaMA [180]	Uni	occ.+language	nuScenes-QA
DrivingWorld [181]	WM-Plan	long-horizon AR	nuScenes	ProphetDWM [182]	Uni	long-horizon	NAVSIM, nuPlan
DrivingGen [183]	WM-Plan	policy-cont. eval	nuScenes	VaViM/VaVAM [184]	Uni	video AR+action AR	OpenDV
OccWorld [59]	WM-Plan	voxel tokens	nuScenes-Occ	DriveLaW [185]	Uni	language-action	nuScenes
Drive-OccWorld [134]	WM-Plan	multi-step occ.	nuScenes-Occ	AutoMoT [186]	Uni	async. MoT	Bench2Drive
DreamerAD [187]	RL	Dreamer-AD	nuScenes, nuPlan	World4Drive [188]	RL	virtual env RL	NAVSIM, nuPlan