

YOUDrive: Driving Style as a Composable Direction for End-to-End Autonomous Driving

Guantian Zheng^{1,3} Lidong Yu³ Tianyu Gao⁴ Jiaying Chen² Jiashu Li⁴ Chen Lv^{1,†}

¹AutoMan Lab, Nanyang Technological University

²School of Vehicle and Mobility, Tsinghua University

³Neolix

⁴City University of Hong Kong

gzheng004@e.ntu.edu.sg

Abstract

End-to-end driving policies can achieve strong task performance while offering limited control over their driving style. Driving style describes how a vehicle accelerates, brakes, and maintains distance. Policies trained without explicit preference conditioning lack a direct mechanism for adjusting these behaviors at inference. We propose YOUDrive, a vision–language–action planner for continuous and composable driving-style control. Each teacher’s behavior is distilled into a task vector, a learned parameter update that induces a style direction in the planner’s representation. These directions can be scaled and combined at inference without retraining the backbone. Conditioned on these directions, a flow-matching decoder models a multimodal trajectory distribution and generates kinematically feasible motion alternatives. We introduce the Style Alignment Score (SAS) to describe human-referenced kinematic behavior after safety and task gating. On NAVSIM navtest, a single coefficient monotonically controls assertiveness across six teacher-derived personas, while PDMS remains above 0.90 across the tested individual and blended configurations, reaching 0.954. Further analyses reveal systematic policy-specific kinematics after controlling for scene context, whereas the evaluated discrete style labels provide limited additional explanatory power. Code, checkpoints, and the metric will be released.

1. Introduction

End-to-end driving models learn to predict trajectories from sensor inputs using large-scale driving datasets that pool demonstrations from multiple drivers [3, 9, 12]. These data reflect not only the demands of each scene but also how individual drivers execute their maneuvers. Their patterns of acceleration, braking, and following distance characterize driving style. A policy trained on these demonstrations acquires its own driving behavior, but learning from different

drivers does not by itself give users control over that behavior. The challenge is to make driving style adjustable at inference while preserving the policy’s ability to respond to traffic and complete the driving task.

One approach is to condition a driving policy on discrete style labels, such as aggressive, normal, and conservative [6]. However, the distribution of scenes and motion states can differ substantially across these categories. When conservative labels are concentrated in low-speed or near-stationary clips, they reflect scene conditions more than driver preferences (Sec. 4.4). This limits their ability to distinguish driver preferences from scene-dependent behavior. For style control, we instead seek a behavioral reference that can be examined under comparable driving conditions. We treat each trained teacher policy as a driver and characterize its style through the trajectories it produces across a common set of scenes.

Building on this policy-level view, we propose YOUDrive to transfer teacher behaviors to a shared vision–language–action (VLA) planner and control their expression at inference. For each teacher, we distill its driving behavior into a task vector, a learned parameter update relative to a common base planner [11]. For a given scene, the difference between the updated and base representations defines the teacher’s style direction. A coefficient controls the strength of this direction, allowing interpolation between the two representations. Directions from multiple teachers can also be combined through weighted addition in the same representation space, without retraining the shared planner. A flow-matching decoder translates the resulting style condition into motion by generating control sequences and integrating them through a constrained motion model. This produces multiple kinematically feasible trajectory candidates under the chosen style condition. To characterize the resulting behavior, we introduce the Style Alignment Score (SAS), which measures interpretable, human-referenced kinematic features after safety and task gating. Reporting SAS alongside the unchanged PDMS allows style changes to be examined separately from planning performance. Our NAVSIM evaluation

[†]Corresponding author.

covers both single-teacher control and multi-teacher composition, demonstrating systematic changes in assertiveness while maintaining high planning performance. Our contributions are summarized as follows:

- We introduce a task-vector formulation of driving-style control in a shared VLA planner, coupled with a style-conditioned flow-matching decoder. Teacher-derived directions can be scaled and combined at inference without retraining the backbone.
- We propose the Style Alignment Score (SAS), which characterizes human-referenced kinematic behavior after safety and task gating, allowing driving style to be reported separately from planning performance.
- Our NAVSIM study identifies policy-specific kinematic patterns and demonstrates monotonic assertiveness control across six teacher-derived personas. PDMS remains above 0.90 across the tested individual and blended configurations, reaching 0.954.

2. Related Work

End-to-end and VLA planning. End-to-end driving methods integrate scene understanding and trajectory planning within a learned policy. TransFuser combines camera and LiDAR features for imitation-based driving [3]. UniAD coordinates perception and prediction tasks around planning, while VAD uses vectorized scene representations [9, 12]. VLA approaches use pretrained vision-language models to connect visual observations and language instructions with action generation. EMMA represents driving inputs and outputs in a unified language format [10], whereas AutoVLA encodes trajectories as discrete feasible actions and predicts them autoregressively [27]. For multimodal trajectory generation, DiffusionDrive uses truncated diffusion with multimodal anchors [19], while GoalFlow employs goal-conditioned flow matching [25]. YOUDrive builds on a VLA planner and replaces discrete action decoding with a continuous flow-matching head in a kinematically constrained control space. We focus on steering the resulting trajectory distribution with teacher-derived behavior directions.

Personalized autonomous driving. Personalized driving has been studied through learning driving styles from demonstrations and modeling social preferences in vehicle interactions [13, 23]. In end-to-end driving, StyleDrive introduces scenario-aware aggressive, normal, and conservative annotations and corresponding style-conditioned baselines [6]. Its style-modulated PDMS adjusts comfort, progress, and time-to-collision terms to assess style alignment. Although these labels enable categorical control, unequal scene and motion-state distributions across categories complicate the interpretation of individual preferences, as examined in our analysis (Sec. 4.4). StyleVLA conditions trajectory gener-

ation on driving-style instructions with physics-informed training objectives [7]. TrajHF fine-tunes generative trajectory models using human feedback to align their outputs with driving preferences [14]. YOUDrive focuses on deriving reusable behavior directions through teacher-policy distillation. Each teacher provides a behavioral reference across scenes, and its induced direction can be scaled and combined with other teacher directions within a shared VLA planner. We also retain PDMS and report human-referenced kinematic behavior through SAS after safety and task gating, so behavioral changes can be examined separately from planning performance.

Task vectors and activation steering. Task arithmetic constructs task vectors by subtracting pretrained weights from their fine-tuned counterparts, allowing learned adaptations to be scaled and combined [11]. Model Soups averages fine-tuned model weights [24], while TIES-Merging addresses interference between parameter updates during merging [26]. Activation steering instead intervenes in hidden representations during inference. ITI shifts selected attention-head activations [15], while Persona Vectors identifies activation directions associated with behavioral traits [2]. YOUDrive links parameter updates to representation-level control through teacher-policy distillation. The learned task vector induces a scene-dependent direction in the VLA planning representation. We scale and compose these directions to condition trajectory generation, rather than directly merging weights at inference.

3. Method

3.1. Overview

At each planning step, YOUDrive takes multi-view images I , ego state e , and a navigation command g as input and predicts a trajectory $\tau = ((x_k, y_k, \psi_k))_{k=1}^T$, where each pose contains a planar position and heading. We use $T=10$ poses at intervals of $\Delta t=0.5$ s.

A shared VLA backbone encodes the inputs into content tokens. Each teacher’s behavior is distilled into a task vector, whose induced change in the content representation defines a scene-dependent style direction. The selected direction is scaled, or several directions are combined, to condition a flow-matching decoder (Fig. 1). The decoder generates control sequences that are integrated through a constrained motion model to produce trajectory candidates, from which the final output is selected.

3.2. VLA backbone

YOUDrive uses Qwen2.5-VL-3B [1] as a shared scene encoder. Multi-view camera images I are encoded as visual tokens, while ego state e and navigation command g are provided as text tokens. After base-policy training,

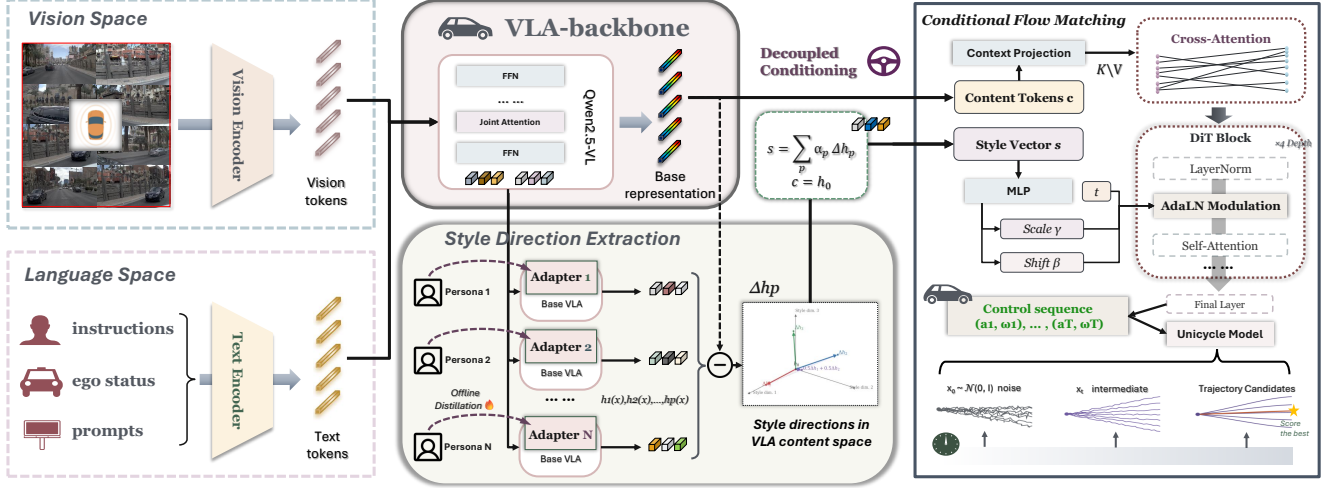


Figure 1. YOUDrive pipeline. The VLA backbone provides content tokens, style is a task-vector interpolation applied to those tokens, and a flow-matching decoder samples several feasible candidates before returning one.

the backbone maps $x = (I, e, g)$ to a content sequence $h_0(x) = f_{\theta_0}(x) \in \mathbb{R}^{L \times d}$. Here θ_0 denotes the trained backbone parameters; the trajectory decoder has separate parameters. This sequence provides scene and navigation context for trajectory generation and serves as the common reference for the teacher-specific representation changes defined in Sec. 3.3.

3.3. Style as a direction in VLA content space

For each teacher policy p , we train a LoRA adapter [8] on its NAVTRAIN trajectories while keeping the base backbone and trajectory decoder fixed. The training objective is the same control-space flow-matching objective used for the base, with teacher rather than human trajectories as targets (Sec. 3.4). Following task arithmetic [11], we treat the adapter’s parameter update ΔW_p relative to the shared backbone θ_0 as a task vector. Applied to the same scene x , the base and teacher-adapted backbones produce $h_0(x)$ and $h_p(x) = f_{\theta_0 + \Delta W_p}(x)$, respectively. Their difference defines the teacher-induced style direction

$$\Delta h_p(x) = h_p(x) - h_0(x). \quad (1)$$

The parameter update ΔW_p and the representation change $\Delta h_p(x)$ are distinct: the latter depends on the current scene and is the quantity used to control trajectory generation.

Scaling and composition. For a single teacher, the decoder receives the interpolated content

$$c_{p,\alpha}(x) = h_0(x) + \alpha \Delta h_p(x), \quad \alpha \in [0, 1], \quad (2)$$

where $\alpha=0$ recovers the base policy and $\alpha=1$ uses the teacher-adapted content. For a blend, we first combine the task vectors with weights β_p and evaluate the adapted back-

bone:

$$\Delta W_\beta = \sum_p \beta_p \Delta W_p, \quad (3)$$

$$c_{\alpha,\beta}(x) = h_0(x) + \alpha(f_{\theta_0 + \Delta W_\beta}(x) - h_0(x)).$$

The β_p used for the reported blends sum to one; the overall strength α remains independently adjustable. Because the backbone is nonlinear, the blend-induced representation change need not equal $\sum_p \beta_p \Delta h_p(x)$. We test the resulting behavior under scaling and composition in Sec. 4.5.

Why the content space. The VLA content sequence already provides the decoder with scene and navigation context. Interpolating the base and adapted content therefore lets a teacher update change the representation consumed by the same trajectory generator, without training a separate decoder or restricting style to a fixed label.

3.4. Flow-matching trajectory decoder

Our decoder is a flow-matching *action head for a VLA*. Generative planners such as DiffusionDrive and GoalFlow model multiple plausible trajectories [19, 25]. YOUDrive retains this capability while using teacher-derived directions as continuous, composable conditions for trajectory generation. The DiT/AdaLN-based head [22] cross-attends to the style-conditioned VLA content sequence c of Eqs. (2) and (3) and generates multiple candidates. In the content-interpolation configuration, the separate AdaLN style input is kept neutral; the teacher direction acts through the content sequence.

We decode in a control space: writing $u_k = (a_k, \omega_k)$ for longitudinal acceleration and yaw rate, each candidate corresponds to a control sequence $z = (u_1, \dots, u_T) \in \mathbb{R}^{T \times 2}$. The training target z_1 is obtained by inverting the target trajectory through the unicycle model and standardizing the

two control channels. Base training uses human trajectories; persona distillation uses the corresponding teacher trajectories. The initial speed for target inversion is estimated from the first target displacement.

Rectified-flow objective. Following rectified flow and flow matching [20, 21], for $z_0 \sim \mathcal{N}(0, I)$ and target z_1 we construct

$$z_t = (1 - t)z_0 + tz_1, \quad t \in [0, 1], \quad (4)$$

with target velocity $z_1 - z_0$. The decoder v_ϕ , with parameters ϕ separate from the backbone θ_0 , regresses that velocity conditioned on the content sequence c through cross-attention:

$$\mathcal{L}_{\text{FM}} = \mathbb{E}_{t, z_0, z_1} \|v_\phi(z_t, t; c) - (z_1 - z_0)\|^2. \quad (5)$$

The training objective also includes a lightly weighted control-jerk penalty ($\lambda_j=0.01$) on the estimated clean controls, weighted by t for flow times $t>0.5$. The same objective is used for base training and persona distillation.

Sampling. At inference we draw $N=16$ independent initial samples and integrate $\dot{z}_t = v_\phi(z_t, t; c)$ using $K=100$ Euler steps:

$$z_{t+\frac{1}{K}} = z_t + \frac{1}{K} v_\phi(z_t, t; c). \quad (6)$$

The sampled set can represent distinct maneuvers under the same scene and style condition instead of averaging them into a single point estimate.

After reversing control normalization, we cap acceleration and yaw rate as $\bar{a}_k = \text{clip}(a_k, -a_{\max}, a_{\max})$ and $\bar{\omega}_k = \text{clip}(\omega_k, -\omega_{\max}, \omega_{\max})$. We use $a_{\max}=5 \text{ m/s}^2$, $\omega_{\max}=1 \text{ rad/s}$, and $v_{\max}=25 \text{ m/s}$. From initial ego speed v_0 (provided by e at inference), heading $\psi_0 = 0$, and position $p_0 = (0, 0)$, the unicycle model integrates the controls as

$$v_k = \text{clip}\left(v_0 + \Delta t \sum_{i=1}^k \bar{a}_i, 0, v_{\max}\right), \quad \psi_k = \psi_{k-1} + \bar{\omega}_k \Delta t, \quad (7)$$

$$p_k = p_{k-1} + v_k(\cos \psi_k, \sin \psi_k) \Delta t,$$

yielding candidates within the specified kinematic bounds. Finally, a prediction-based PDM selector [4, 5] ranks the N candidates using the current scene, map context, and constant-velocity forecasts of surrounding agents, without using their logged future trajectories. The selected candidate is the planner output on which style and planning performance are evaluated.

4. Experiments

4.1. Setup

Our planner uses Qwen2.5-VL-3B [1] as its scene encoder: multi-view images, ego status, and the navigation command are tokenized into content tokens summarizing both the scene and the requested maneuver. On top we run the flow-matching decoder of Sec. 3.4, conditioned on the

content tokens through cross-attention; it flows in a control space of longitudinal acceleration and yaw rate under a unicycle model, integrated with K steps and sampled $N=16$ times per frame. We warm-start from the released AutoVLA_PDMS_89 checkpoint [27] and fine-tune the language backbone jointly with the decoder on NAVSIM navtrain, vision encoder frozen and the human trajectory as target; the result is our “base” ($\alpha=0$). Each persona adapter is distilled from a teacher planner—GTRS [18], HydraMDP [17], WoTe [16], TransFuser [3], GoalFlow [25], DiffusionDriveV2 [19]—and applied through Eq. (2). Each is a LoRA [8] module ($r=64$, $\alpha_{\text{LoRA}}=128$, dropout 0.05) on the attention and MLP projections; the base VLM (including the vision tower) and the flow head are frozen, so only the LoRA weights train per persona. We evaluate on NAVSIM [5] navtest (about 12k scenarios, 8 shards) and report scenario-weighted PDMS (v1), alongside our Style Alignment Score (SAS), which places each policy on an assertiveness axis H and a proximity axis P relative to the human (Sec. 4.2).

4.2. SAS: Style Alignment Score

Gate, then score. Safety and driving style answer different questions and should not share one axis, so we evaluate a policy in two stages. First a *gate*: a trajectory must pass NAVSIM’s safety and task terms (no-collision, drivable-area compliance, time-to-collision, ego progress); style is scored only for trajectories that clear the gate, and style never modulates the safety verdict. Second, we ask how the passing motion aligns in style with a target driver. SAS and PDMS are thus reported on separate axes: we leave PDMS unchanged and report SAS beside it. This separation is one of reporting—neither score modulates the other—not a claim of statistical independence; indeed a persona that improves ego-progress can also raise PDMS.

Atoms and human-referenced normalization. We summarize a trajectory by physically interpretable kinematic atoms: dynamics $\{a^+, a^-, j_\ell, j_t\}$ (peak accel/decel, longitudinal and lateral jerk) and proximity $\{\text{thw}, \text{ttc}, d_{\min}, d_{\text{vru}}\}$ (min time-headway, time-to-collision, object and VRU gap). For a policy d and atom f we take the per-scene value and its median $m_f(d)$ over the split. To compare heterogeneous units against the human driver, we apply a signed z -score

$$z_f(d) = \frac{m_f(d) - m_f(\text{human})}{\sigma_f}, \quad \sigma_f = \text{std}_{d'} m_f(d'), \quad (8)$$

where σ_f is taken over the evaluated policies (the human, our base, and the six teachers); thus $z=0$ is human, $z>0$ is “more than human”, and one unit is one cross-policy standard deviation. SAS is therefore a *within-study, relative* diagnostic: it measures how far apart policies sit on a fixed evaluation set, not a portable benchmark. Its validity rests on independent evidence (per-model separability and monotone response to α), not on a universal scale.

Two style axes. We aggregate the atoms into a primary *Assertiveness* axis and a secondary *Proximity* axis,

$$H(d) = \frac{1}{|\mathcal{D}|} \sum_{f \in \mathcal{D}} z_f(d), \quad P(d) = \frac{1}{|\mathcal{P}|} \sum_{f \in \mathcal{P}} (-z_f(d)), \quad (9)$$

with $\mathcal{D}$ the dynamics atoms and $\mathcal{P}$ the proximity atoms (equal weight). $H > 0$ denotes harder, jerkier driving than the human; $P > 0$ denotes tighter gaps. SAS reports style as (H, P) with H as the primary, validated axis.

What makes SAS valid. A style metric must (i) separate drivers and (ii) respond monotonically to a style control. We show that H is strongly per-model separable (Sec. 4.4) and is driven monotonically by the single strength α (Sec. 4.5), and we report a per-atom variance ratio η^2 next to the axes so the metric states its own discriminative power. On NAVTEST the proximity atoms vary little across policies; we therefore keep P secondary and return to it as a limitation.

4.3. Style and safety on navtest (v1)

Table 1 reports PDMS with its sub-scores on navtest together with the SAS assertiveness axis H , for the base policy and, for each of the six teachers, the native teacher planner and our distilled persona. Every persona is applied at full strength ($\alpha=1$); the continuous behavior as α varies is the subject of Sec. 4.5, and persona blends, which reach operating points beyond any single teacher, are shown in Tab. 5b. At full strength the personas span a wide assertiveness range, from smooth teachers close to the human ($H \approx 0.05$ – 0.15 for GTRS, Hydra-MDP) to an aggressive persona (DiffusionDriveV2, $H \approx 3.3$), while our personas keep PDMS in a narrow band (0.90 – 0.94) around the base. The sub-scores localize the change. The hard-safety components that gate the score (no-collision NC and drivable-area compliance DAC) stay essentially flat, whereas ego-progress (EP), time-to-collision (TTC), and comfort (Comf) move with H . Each distilled persona lands near its native teacher’s H , so the adapter transfers the teacher’s assertiveness rather than a generic average: where a native planner sits at one fixed point on the assertiveness axis, the matching persona sweeps a controllable range around it (Sec. 4.5), and the complementary proximity axis is deferred to Sec. 4.7.

4.4. Style is a per-model signal, not a label

A style axis is only meaningful if style is a real, systematic signal rather than scene noise. Here a *driver* is a behaviorally self-consistent policy—a model, in the sense of Sec. 1—so its style is the kinematic signature it produces across scenes. On a fixed set of tokens we collect trajectories for 16 such drivers (the human reference, external planners, and our base) and run a one-way variance decomposition of per-token kinematics with driver identity as the factor (Tab. 2). Driver identity explains a mean η^2 of 0.27, up to 0.46 for lateral jerk, so how a driver accelerates, brakes, and jerks

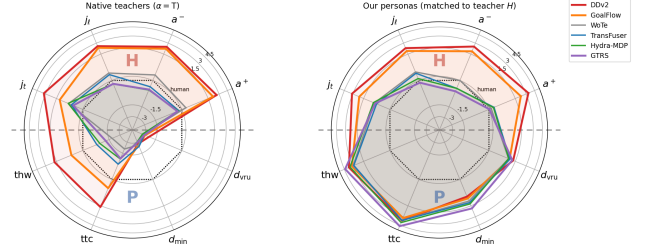


Figure 2. Style geometry in atom space (human-referenced signed z ; the dotted ring marks the human reference, radius grows with assertiveness on the dynamics half H and with gap-tightening on the proximity half P). Left: native teachers ($\alpha=T$); right: our personas. Both fan out along the dynamics atoms (the primary axis H); on the proximity atoms the cross-policy spread is small, so P is near-degenerate on navtest and secondary, with a teacher/persona sign mismatch we treat as an open problem (Sec. 4.7).

is strongly driver-specific. The one exception is average speed ($\eta^2=0.005$), set by road and traffic—hence a weak style coordinate on navtest. This supports treating style as a continuous axis organized by model-as-driver rather than coarse labels. Figure 2 makes the split visible in atom space—dynamics atoms fan out across policies while proximity atoms collapse near the human—so we take assertiveness H as primary and keep proximity P secondary (Sec. 4.7).

Discrete labels do not encode style beyond speed. If model-as-driver is the right unit for style, the natural alternative—hand-assigned categorical labels—should be tested with the same kinematic atoms. We take the aggressive/normal/conservative (ANC) labels released by StyleDrive [6] for the navtest scenes and evaluate them with *our* kinematic atoms, on the human log-replay trajectory of each scene. This is the fair test: StyleDrive assigns ANC from each clip’s own ego kinematics, so measuring the labels against those same kinematics asks exactly what the label is supposed to encode.

Grouping the human trajectories by ANC, the raw between-group effect size on the assertiveness atoms is near zero ($\epsilon^2=0.006$ for peak acceleration, 0.012 for peak deceleration, ≤ 0.05 for jerk, 0.063 aggregated to H); the only atom ANC separates is average speed ($\epsilon^2=0.088$), and the proximity atoms it appears to move follow directly from speed. As a positive control, grouping the same trajectories by a speed band recovers large effects on exactly these atoms ($\epsilon^2=0.38$ – 0.40 for peak accel/decel), so ANC’s near-zero effect is a true null, not a dead test. The residual signal also vanishes once environment is accounted for: regressing each atom on the scene environment (scenario, turn, lead, road), ANC adds almost nothing ($\Delta R^2(\text{ANC} \mid \text{env})=0.037$ on H , 0.000 on P ; Tab. 3), whereas the same regression leaves driver identity a strong organizer ($\Delta R^2=0.232$ on H , over six times the label’s).

What ANC does track is speed: its composition follows a six-bin speed partition (Cramér’s $V=0.34$), and conser-

config	α	PDMS $\uparrow$	PDMS sub-scores					SAS
			NC	DAC	EP	TTC	Comf	H
human anchor	—	—	—	—	—	—	—	0.00
base (ours)	0	0.912	98.8	96.6	86.7	96.3	100.0	0.48
<i>each teacher's native style ($\alpha=T$) and our persona steered to it ($\alpha=1$)</i>								
GTRS	T	0.899	99.3	97.6	83.2	96.1	100.0	−0.01
↪ ours	1.0	0.903	97.5	96.6	87.6	93.6	100.0	0.05
Hydra-MDP	T	0.865	98.3	96.0	78.7	94.6	100.0	+0.05
↪ ours	1.0	0.907	98.0	96.6	87.7	94.4	100.0	0.13
WoTe	T	0.883	98.5	96.8	81.9	94.9	99.9	0.16
↪ ours	1.0	0.905	98.4	95.6	86.9	95.5	100.0	0.17
TransFuser	T	0.840	97.7	92.8	79.2	92.8	100.0	0.06
↪ ours	1.0	0.900	98.6	95.7	85.8	95.5	100.0	0.10
GoalFlow	T	0.903	98.4	98.3	85.0	94.6	100.0	1.93
↪ ours	1.0	0.935	99.4	97.9	89.0	98.0	100.0	1.99
DiffusionDriveV2	T	0.881	98.2	96.2	82.2	94.7	100.0	2.72
↪ ours	1.0	0.941	99.5	98.5	89.5	98.1	100.0	3.26

Table 1. PDMS with its sub-scores (no-collision NC, drivable-area compliance DAC, ego-progress EP, time-to-collision TTC, comfort Comf) and the SAS assertiveness axis H (Eq. (9)) on NAVSIM navtest. For each teacher we list its native behavior ($\alpha=T$) and, below it ($\hookrightarrow$ ours), our persona steered to that teacher at full strength ($\alpha=1$); base is the same model with no persona ($\alpha=0$), and the human driver sits at $H=0$. Native-teacher PDMS are the reported navtest v1 numbers from each planner’s paper [3, 16, 17, 19, 25]; GTRS reports v2 EPDMS only, so its v1 row is our own evaluation.

kinematic feature	$\eta^2(\text{driver})\uparrow$	ICC(1)	F
avg. speed	0.005	0.005	20
peak accel	0.275	0.288	1631
peak decel	0.287	0.300	1735
long. jerk	0.445	0.461	3451
lat. jerk	0.460	0.476	3663
lat. accel (max)	0.167	0.176	862
mean	0.273	0.286	—

Table 2. Variance decomposition of per-token kinematics with driver identity as the factor (16 drivers, common tokens). Control features are driver-dominated, while average speed is scene-locked. In contrast, the coarse ANC label of [6] carries little style signal once environment is accounted for (Sec. 4.4).

vative is largely a stationary-scene detector—median speed 0.45 m/s, 77.5% of its scenes near-stationary (< 2 m/s), and 87% of fully stopped scenes labelled conservative. A plain speed partition with no style design separates our atoms as well as or better than ANC on nearly every atom. The label thus re-describes driving *context* (how fast, what scene), not an independent per-model style—the failure mode that motivates model-as-driver.

4.5. Controllable, composable style

Sweeping α moves the assertiveness axis H and its kinematics from the base toward each teacher, so one coefficient spans a range rather than switching between presets (Tab. 4 and Fig. 4; decoded trajectories in Fig. 3). For the DiffusionDriveV2 persona, raising α from 0 to 1 drives H monotonically to 3.26 while PDMS stays high and peaks near $\alpha=0.75$ —strong style expression at no cost to safety;

	ANC label	driver identity
raw effect size, peak accel	0.006	0.275
raw effect size, peak decel	0.012	0.287
raw effect size, assertiveness H	0.063	0.270
$\Delta R^2(\cdot \mid \text{env})$ on H	0.037	0.232
$\Delta R^2(\cdot \mid \text{env})$ on P	0.000	0.065

Table 3. Two complementary diagnostics of what organizes driving style, on the navtest atoms. The ANC column groups the *human log-replay trajectories* by StyleDrive’s label; the driver column groups the *16 policies’ trajectories* by identity. Top: raw between-group effect size (ϵ^2 , rank-based, for ANC; η^2 , one-way ANOVA, for driver). Bottom: incremental *adjusted* R^2 after regressing out the scene environment, penalizing the differing degrees of freedom (3 ANC classes vs. 16 drivers). After control, model identity remains a strong organizer, unlike the label; P is weak for both (Sec. 4.7).

the dose–response is smooth and monotone across the whole persona set (Fig. 4). Multi-modality is common enough to matter on navtest: among scenes with sufficient driver coverage, about 7.4% are laterally bimodal and 13.9% longitudinally bimodal (brake or go). There a single-mode regressor returns a path between the modes, whereas the sampling decoder draws distinct maneuvers and commits to one, so style survives where averaging would erase it.

Task-vector properties of the persona direction. A construction that is linear in content space need not stay linear or additive at the measured output, so we test these two task-vector properties directly on the style axis. *Linearity* we have already seen as the dose–response of Sec. 4.5, where H moves monotonically with α (Fig. 4 and Tab. 4); we quantify it as the fit R^2 of H against α over the full sweep. *Compo-*

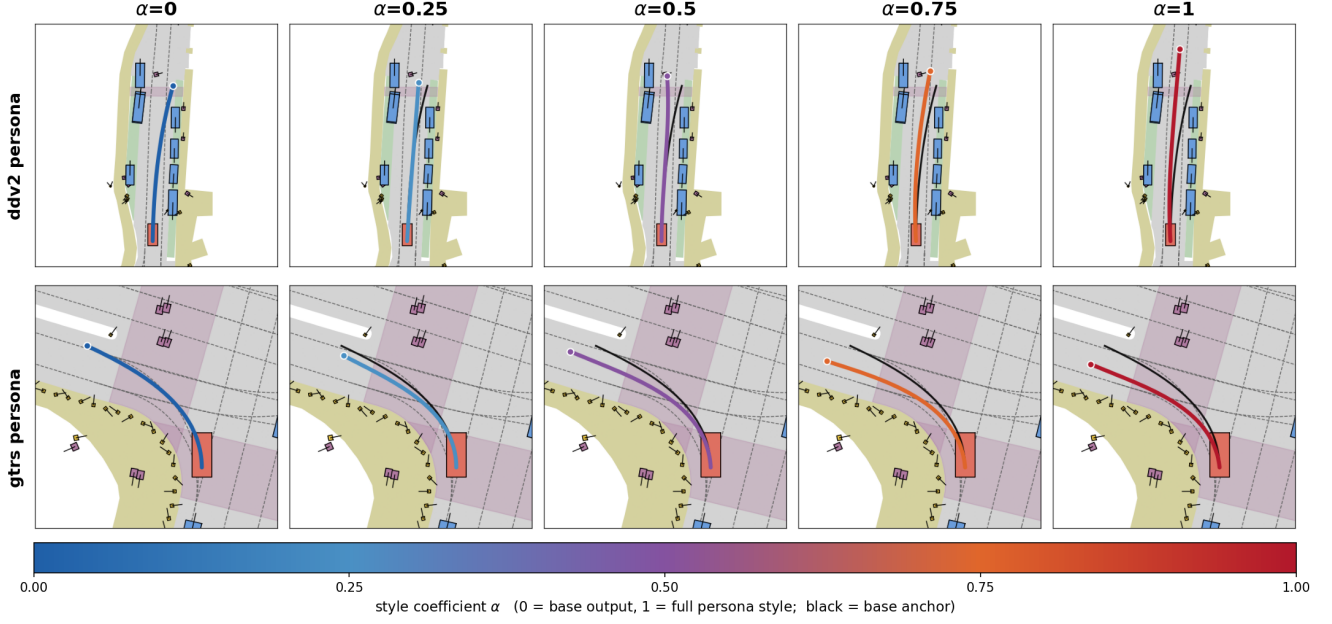


Figure 3. Style controllability in BEV. Each row shows one scene decoded at increasing style strength α (columns, $0 \rightarrow 1$): an assertive persona (DiffusionDriveV2, top) and a cautious persona (GTRS, bottom). Trajectory color encodes α (0 is the base output, 1 is full persona style); the black curve is the fixed base anchor. A single coefficient sweeps the decoded trajectory continuously, and every α stays safe (gate-passing). Ego is outlined in red, other agents in blue.

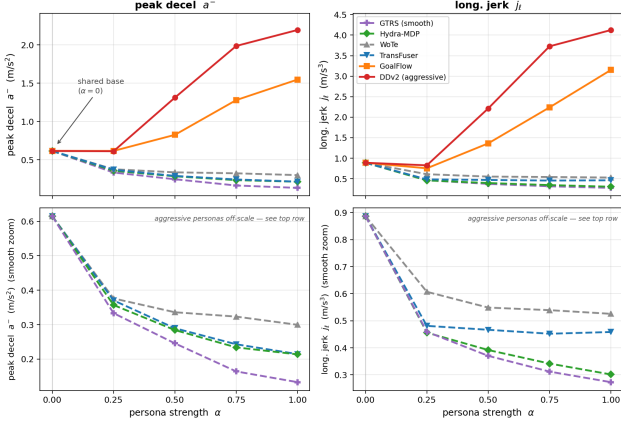


Figure 4. Dose-response: as the single strength α rises, the dynamics atoms (peak deceleration and longitudinal jerk shown) move monotonically from the base toward each teacher’s operating point, so α is a continuous style coefficient rather than a discrete selector. The smoother personas (lower panel, zoomed) separate cleanly at small magnitudes.

sition asks whether a blend $\sum_i w_i \Delta_i$ matches the weighted sum of the single-persona shifts (Tab. 5a), letting personas reach new operating points beyond any single teacher.

Composition is not just a property but a capability: a convex blend yields a new operating point at no extra training cost and reaches places no single teacher does. Table 5b lists measured blends at $\alpha=1$: the strongest, Dif-

α	PDMS $\uparrow$	H
0.00	0.912	0.48
0.25	0.926	0.53
0.50	0.941	1.52
0.75	0.947	2.81
1.00	0.941	3.26

(a) DiffusionDriveV2 (aggressive): (b) GTRS (smooth): H falls as driving turns more conservative.

Table 4. Style following as α increases (full sweep in Tab. 7). One coefficient α steers the base toward each teacher in *opposite* directions on the assertiveness axis H —the aggressive DiffusionDriveV2 persona (left) rises while the smooth GTRS persona (right) falls—with PDMS high throughout.

fusionDriveV2/GTRS, attains PDMS=0.954, above every single persona (0.941, Tab. 1), while blends without DiffusionDriveV2 occupy a milder band ($H \approx 1.05$), so blends sweep a continuous style range at uniformly high safety. Same-direction blends track the weighted-sum prediction to within $\sim 10\%$, whereas aggressive+smooth blends are *super-additive*: the aggressive direction dominates, so measured H exceeds the linear prediction rather than averaging toward the milder teacher.

4.6. Ablation Study

We ablate the two design choices that most directly shape controllable style: *how* trajectories are decoded and *how much* sampling inference uses. Unless noted, each variant is

property	measurement	result
linearity in α	R^2 of H vs. α	0.93
additive composition	error vs. weighted sum (same-dir.)	6%

(a) Task-vector properties: linearity is the R^2 of H against α over the aggressive teachers (where H has room to move); composition is the relative error between a measured blend and the weighted sum of single-persona shifts, over the same-direction blends of (b).

persona blend (weights)	PDMS $\uparrow$	H	P
ddv2 + gtrs (60/40)	0.954	+2.69	+2.45
ddv2 + gf + hy (40/40/20)	0.950	+2.42	+2.37
ddv2 + gf (60/40)	0.949	+2.96	+2.27
ddv2 + gf + tf + gtrs (35/35/15/15)	0.940	+2.13	+2.46
gf + gtrs (50/50)	0.930	+1.05	+2.35
tf + gf (50/50)	0.921	+1.06	+2.33

(b) Representative blends at $\alpha=1$, a convex mixture $\sum_i w_i \Delta_i$ of single-persona directions (codes: ddv2 = DiffusionDriveV2, gf = GoalFlow, tf = TransFuser, gtrs = GTRS, hy = Hydra-MDP). No retraining; the best, ddv2 + gtrs, hits 0.954 vs 0.941 for the strongest single persona (Tab. 1). H/P are the SAS axes (human at 0); full set (all α) in Tab. 8.

Table 5. Persona directions behave as reusable, composable task vectors.

trained and evaluated on the same NAVTEST split at matched capacity; we report PDMS with its safety sub-scores (no-collision NC, drivable-area compliance DAC) next to the reachable *style span* ΔH —the range of the assertiveness axis a single model covers across its personas and $\alpha \in [0, 1]$ (per-persona values in Sec. B, full sweep in Sec. A).

Decoder family. We replace the flow-matching decoder with the backbone’s discrete action codebook, keeping conditioning and training data fixed. The codebook preserves safety but quantizes trajectories onto a fixed set of anchors, which caps the reachable style range and leaves only coarse, quantized style control (Tab. 6a); the flow-matching decoder recovers a smooth, wider style axis at equal safety.

Samples and steps. Sweeping the two inference knobs, both PDMS and the style span ΔH saturate by $N=16$, $K=20$ (Tab. 6b); controllability thus lives in the persona direction, not the sampler, so we default to $N=16$, $K=100$.

4.7. Limitations

SAS is validated along assertiveness (H); proximity (P) is reported only as a secondary, preliminary axis. On NAVTEST the proximity atoms vary far less across policies than the dynamics atoms (cross-policy std 0.14–0.29 vs 0.57–1.52), so social style is near-degenerate on this split, dominated by open-road, low-interaction scenes. Within this narrow range every native teacher loosens gaps ($P<0$) while every persona tightens them ($P>0$, Fig. 2), so the persona direction does not reproduce the teacher’s sign on P ; with the signal this close to degenerate, whether P is independently controllable remains open. Resolving it requires interaction-dense, reactive closed-loop evaluation (e.g. nuPlan reactive, CARLA), where headway and gap have room to vary; we

decoder	PDMS $\uparrow$	NC $\uparrow$	DAC $\uparrow$	EP $\uparrow$	TTC $\uparrow$	Comf $\uparrow$	ΔH
discrete codebook	0.891	0.984	0.956	0.818	0.980	0.999	0.8
flow matching (ours)	0.912	0.988	0.966	0.867	0.963	1.000	3.2

(a) Decoder family (conditioning fixed). Both keep hard safety (NC/DAC) and PDMS close, but the discrete codebook caps the reachable style span ($\Delta H=0.8$ vs 3.2) with only coarse, quantized control; flow matching yields a smooth, wider style axis at equal safety.

setting	PDMS $\uparrow$	ΔH
<i>candidates N (steps $K=100$)</i>		
$N=4$	0.883	2.5
$N=8$	0.899	3.1
$N=16$ (ours)	0.912	3.2
$N=64$	0.913	3.2
<i>steps K (candidates $N=16$)</i>		
$K=5$	0.864	2.3
$K=20$	0.910	3.2
$K=100$ (ours)	0.912	3.2

(b) Inference budget. PDMS and ΔH saturate by $N=16$, $K=20$ and narrow only when the sampler is starved ($N=4$, $K=5$), so the default $N=16$, $K=100$ leaves little on the table.

Table 6. Ablations: decoder family (a) and inference budget (b). Sub-scores as in Tab. 1.

leave this to future work.

5. Conclusion

We presented YOUDrive, which treats driving style as a measurable, composable control axis on a shared backbone—each driver a reusable direction—rather than an average in the data or a label pasted onto it. The framework rests on three parts: a flow-matching decoder that commits to a feasible trajectory rather than averaging, and a low-rank task vector per driver that one coefficient scales and composes without retraining (*expression*); the Style Alignment Score (SAS), which reports style on its own terms after gating hard safety (*measurement*); and, on NAVSIM, evidence that this axis is a systematic per-model signal, unlike a coarse label, with a single coefficient tracing a wide style range while PDMS stays above 0.90 (*validation*). Its main limit: SAS is validated along assertiveness alone; the proximity axis is near-degenerate on navtest, so whether social style is independently controllable remains open, calling for interaction-dense, reactive closed-loop evaluation. Validating style against human preference is the next step.

References

- [1] Shuai Bai et al. Qwen2.5-VL technical report. *arXiv preprint arXiv:2502.13923*, 2025. 2, 4
- [2] Runjin Chen, Andy Ardit, Henry Sleight, Owain Evans, and Jack Lindsey. Persona vectors: Monitoring and controlling character traits in language models. *arXiv preprint arXiv:2507.21509*, 2025. 2
- [3] Kashyap Chitta, Aditya Prakash, Bernhard Jaeger, Zehao Yu, Katrin Renz, and Andreas Geiger. TransFuser: Imitation

- with transformer-based sensor fusion for autonomous driving. *IEEE TPAMI*, 2023. 1, 2, 4, 6
- [4] Daniel Dauner, Marcel Hallgarten, Andreas Geiger, and Kashyap Chitta. Parting with misconceptions about learning-based vehicle motion planning. In *Conf. Robot Learn. (CoRL)*, 2023. 4
- [5] Daniel Dauner et al. NAVSIM: Data-driven non-reactive autonomous vehicle simulation and benchmarking. In *NeurIPS*, 2024. 4
- [6] Ruiyang Gao et al. StyleDrive: Towards driving-style aware benchmarking of end-to-end autonomous driving. *arXiv preprint arXiv:2506.23982*, 2025. 1, 2, 5, 6
- [7] Yuan Gao, Dengyuan Hua, Mattia Piccinini, Finn Rasmus Schäfer, Korbinian Moller, Lin Li, and Johannes Betz. StyleVLA: Driving style-aware vision language action model for autonomous driving. *arXiv preprint arXiv:2603.09482*, 2026. 2
- [8] Edward J. Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. LoRA: Low-rank adaptation of large language models. In *ICLR*, 2022. 3, 4
- [9] Yihan Hu, Jiazhi Yang, Li Chen, Keyu Li, Chonghao Sima, Xizhou Zhu, Siqi Chai, Senyao Du, Tianwei Lin, Wenhai Wang, et al. Planning-oriented autonomous driving. In *CVPR*, 2023. 1, 2
- [10] Jyh-Jing Hwang, Runsheng Xu, Hubert Lin, Wei-Chih Hung, Jingwei Ji, Kristy Choi, Di Huang, Tong He, Paul Covington, Benjamin Sapp, et al. EMMA: End-to-end multimodal model for autonomous driving. *arXiv preprint arXiv:2410.23262*, 2024. 2
- [11] Gabriel Ilharco, Marco Tulio Ribeiro, Mitchell Wortsman, Ludwig Schmidt, Hannaneh Hajishirzi, and Ali Farhadi. Editing models with task arithmetic. In *ICLR*, 2023. 1, 2, 3
- [12] Bo Jiang, Shaoyu Chen, Qing Xu, Bencheng Liao, Jiajie Chen, Helong Zhou, Qian Zhang, Wenyu Liu, Chang Huang, and Xinggang Wang. VAD: Vectorized scene representation for efficient autonomous driving. In *ICCV*, 2023. 1, 2
- [13] Markus Kuderer, Shilpa Gulati, and Wolfram Burgard. Learning driving styles for autonomous vehicles from demonstration. In *IEEE Int. Conf. Robot. Autom. (ICRA)*, 2015. 2
- [14] Derun Li, Changye Li, Yue Wang, Jianwei Ren, Xin Wen, Pengxiang Li, Leimeng Xu, Kun Zhan, Peng Jia, Xianpeng Lang, Ningyi Xu, and Hang Zhao. Learning personalized driving styles via reinforcement learning from human feedback. *arXiv preprint arXiv:2503.10434*, 2025. 2
- [15] Kenneth Li, Oam Patel, Fernanda Viégas, Hanspeter Pfister, and Martin Wattenberg. Inference-time intervention: Eliciting truthful answers from a language model. In *NeurIPS*, 2023. 2
- [16] Yingyan Li et al. End-to-end driving with online trajectory evaluation via BEV world model. In *ICCV*, 2025. 4, 6
- [17] Zhiyu Li et al. Hydra-MDP: End-to-end multimodal planning with multi-target hydra-distillation. In *CVPRW*, 2024. 4, 6
- [18] Zhiyu Li et al. Generalized trajectory scoring for end-to-end multimodal planning. In *ICCV*, 2025. 4
- [19] Bencheng Liao et al. DiffusionDrive: Truncated diffusion model for end-to-end autonomous driving. In *CVPR*, 2025. 2, 3, 4, 6
- [20] Yaron Lipman, Ricky T. Q. Chen, Heli Ben-Hamu, Maximilian Nickel, and Matt Le. Flow matching for generative modeling. In *ICLR*, 2023. 4
- [21] Xingchao Liu, Chengyue Gong, and Qiang Liu. Flow straight and fast: Learning to generate and transfer data with rectified flow. In *ICLR*, 2023. 4
- [22] William Peebles and Saining Xie. Scalable diffusion models with transformers. In *ICCV*, 2023. 3
- [23] Wilko Schwarting, Alyssa Pierson, Javier Alonso-Mora, Sertac Karaman, and Daniela Rus. Social behavior for autonomous vehicles. *Proc. Natl. Acad. Sci. (PNAS)*, 116(50): 24972–24978, 2019. 2
- [24] Mitchell Wortsman, Gabriel Ilharco, Samir Ya Gadre, Rebecca Roelofs, Raphael Gontijo-Lopes, Ari S. Morcos, Hongseok Namkoong, Ali Farhadi, Yair Carmon, Simon Kornblith, and Ludwig Schmidt. Model soups: Averaging weights of multiple fine-tuned models improves accuracy without increasing inference time. In *Int. Conf. Mach. Learn. (ICML)*, 2022. 2
- [25] Zebin Xing et al. GoalFlow: Goal-driven flow matching for multimodal trajectory generation in end-to-end autonomous driving. In *CVPR*, 2025. 2, 3, 4, 6
- [26] Prateek Yadav, Derek Tam, Leshem Choshen, Colin Raffel, and Mohit Bansal. TIES-Merging: Resolving interference when merging models. In *NeurIPS*, 2023. 2
- [27] Zewei Zhou et al. AutoVLA: A vision-language-action model for end-to-end autonomous driving. *arXiv preprint arXiv:2506.13757*, 2025. 2, 4

A. Full Style Control and Composition Results

Tabs. 7 and 8 report the complete results for individual persona sweeps and persona blends, respectively. All results are scenario-weighted means over eight NAVTEST shards. Alongside PDMS and the SAS axes H and P , we include the individual kinematic and proximity measurements, including dimensions with weak separation on this split. The `scope` column distinguishes human and base references (`anchor`), individual persona sweeps (`sweep`), and blends (`combo`). In `sweep`, α varies the strength of a single persona and $\alpha=T$ denotes the native teacher. In `combo`, the configuration name gives the mixing weights of the teacher task vectors, while α controls their overall strength.

The full individual sweeps in Tab. 7 extend the selected examples in Tab. 4. For all six personas, H changes monotonically with α , though in different directions. DiffusionDriveV2 raises H from 0.48 to 3.26, while GTRS lowers it from 0.48 to 0.05. The corresponding peak acceleration and longitudinal jerk rise from 0.53 and 0.89 to 2.66 and 4.12 for DiffusionDriveV2, but fall to 0.24 and 0.27 for GTRS. Average speed varies less than these dynamics measurements, consistent with its weak driver-specific signal in Sec. 4.4. PDMS stays above 0.90 across the tested persona strengths, but does not change monotonically with α . The proximity axis P also does not follow the native teachers: it is positive

for the distilled personas and negative for the native teachers. We keep this as a limitation, as discussed in Sec. 4.7.

The blend results in Tab. 8 give a fuller view of the composition reported in Tab. 5b. Changing the mixing weights can change both H and PDMS. For DiffusionDriveV2/TransFuser at $\alpha=1$, increasing the DiffusionDriveV2 weight from 30% to 70% changes H from 0.74 to 2.92 and PDMS from 0.901 to 0.948. A fixed blend can also be scaled: DiffusionDriveV2/GTRS (60/40) moves from $H=0.48$ at $\alpha=0$ to 0.92 at $\alpha=0.5$ and 2.69 at $\alpha=1$, with PDMS reaching 0.954. These output changes are not necessarily additive, consistent with the nonlinear blended representation in Sec. 3.3. All tested blends retain PDMS above 0.90.

B. Per-Persona Ablation Breakdown

Tab. 9 gives the per-persona assertiveness scores behind the decoder comparison in Tab. 6a. The style span ΔH is the difference between the largest and smallest measured H values across the evaluated persona settings. We show the base ($\alpha=0$) and each persona at full strength ($\alpha=1$), together with the resulting span.

Flow matching covers a wider assertiveness range than the discrete codebook ($\Delta H=3.2$ versus 0.8). Most of this difference comes from the more assertive personas: DiffusionDriveV2 reaches $H=3.3$ with flow matching but 0.8 with the codebook, while GoalFlow reaches 2.0 and 0.7, respectively. The smoother personas remain close to the human reference under both decoders. Thus, the wider range in Tab. 6a mainly reflects the decoder’s ability to express the stronger teacher-induced dynamics, rather than a similar increase for every persona.

scope	teacher	α	PDMS $\uparrow$	H	P	$\bar{v}$	a^+	a^-	j_ℓ	j_t	thw	ttc	lead	$d_{\min}$	d_{vru}
anchor	human	–	0.9480	+0.00	+0.00	4.6181	0.2446	0.3059	0.2898	0.1026	3.7850	4.3915	12.9060	4.5340	12.0630
anchor	base (ours)	–	0.9120	+0.48	+0.08	4.6051	0.5282	0.6151	0.8865	0.4750	3.7640	4.3710	12.8275	4.5030	12.0780
sweep	gtrs	0	0.9120	+0.48	+0.08	4.6051	0.5282	0.6151	0.8865	0.4750	3.7640	4.3710	12.8275	4.5030	12.0780
sweep	gtrs	0.25	0.9187	+0.27	+1.75	4.8300	0.3713	0.3331	0.4590	0.5258	3.2140	3.5045	11.6580	4.4240	11.9630
sweep	gtrs	0.5	0.9230	+0.18	+2.03	4.9096	0.3341	0.2454	0.3701	0.4722	3.1210	3.3810	11.5660	4.3940	11.9520
sweep	gtrs	0.75	0.9158	+0.10	+2.38	4.9346	0.2753	0.1638	0.3118	0.4139	3.0620	3.2390	11.1695	4.3610	11.8570
sweep	gtrs	1	0.9030	+0.05	+3.01	4.9977	0.2449	0.1326	0.2728	0.3721	2.8625	3.0440	10.8545	4.2760	11.7870
sweep	gtrs	T	0.899	-0.01	-2.02	4.4350	0.2160	0.1840	0.2520	0.2160	4.0270	4.6615	14.1960	4.9570	12.7700
sweep	hydra_mdp	0	0.9120	+0.48	+0.08	4.6051	0.5282	0.6151	0.8865	0.4750	3.7640	4.3710	12.8275	4.5030	12.0780
sweep	hydra_mdp	0.25	0.9191	+0.28	+1.69	4.8250	0.3668	0.3568	0.4559	0.5304	3.2820	3.5350	11.7510	4.3980	11.9590
sweep	hydra_mdp	0.5	0.9232	+0.21	+1.78	4.8747	0.3171	0.2840	0.3916	0.5004	3.2250	3.4960	11.6775	4.4050	11.9580
sweep	hydra_mdp	0.75	0.9198	+0.16	+2.12	4.8833	0.2932	0.2337	0.3416	0.4722	3.1280	3.3625	11.3740	4.3740	11.9080
sweep	hydra_mdp	1	0.9067	+0.13	+2.40	4.9509	0.2756	0.2133	0.3013	0.4412	3.0530	3.2420	11.2290	4.3500	11.8650
sweep	hydra_mdp	T	0.8650	+0.05	-1.99	4.5170	0.2250	0.1700	0.2510	0.3690	3.9170	4.6515	14.1240	4.9690	12.8480
sweep	wote	0	0.9120	+0.48	+0.08	4.6051	0.5282	0.6151	0.8865	0.4750	3.7640	4.3710	12.8275	4.5030	12.0780
sweep	wote	0.25	0.9164	+0.32	+1.64	4.8201	0.4228	0.3759	0.6068	0.5135	3.2790	3.5840	11.6820	4.4490	11.8890
sweep	wote	0.5	0.9192	+0.29	+1.70	4.8399	0.3552	0.3358	0.5485	0.5483	3.2160	3.4620	11.6325	4.4530	11.9950
sweep	wote	0.75	0.9152	+0.23	+1.88	4.8764	0.3007	0.3231	0.5390	0.4679	3.2080	3.4450	11.4520	4.3930	11.9430
sweep	wote	1	0.9051	+0.17	+2.09	4.8702	0.2523	0.2993	0.5254	0.4028	3.1610	3.3650	11.4460	4.3900	11.8680
sweep	wote	T	0.8830	+0.16	-2.41	4.3269	0.2693	0.3496	0.4682	0.3516	4.1780	4.9650	14.3820	4.8985	12.8190
sweep	ddv2	0	0.9120	+0.48	+0.08	4.6051	0.5282	0.6151	0.8865	0.4750	3.7640	4.3710	12.8275	4.5030	12.0780
sweep	ddv2	0.25	0.9262	+0.53	+1.55	4.7682	0.5200	0.6112	0.8234	0.6351	3.2960	3.6320	11.8680	4.4540	11.9040
sweep	ddv2	0.5	0.9406	+1.52	+1.48	4.8019	1.4282	1.3115	2.2109	1.0144	3.2270	3.6350	12.1850	4.4710	12.0180
sweep	ddv2	0.75	0.9467	+2.81	+2.30	4.8995	2.5549	1.9862	3.7266	1.8909	3.0130	3.3460	11.8520	4.4440	11.7650
sweep	ddv2	1	0.9407	+3.26	+2.33	4.8513	2.6606	2.1951	4.1235	2.5209	3.0275	3.2980	12.0275	4.4460	11.7610
sweep	ddv2	T	0.8810	+2.72	-0.42	4.8577	2.0381	1.9625	4.1628	1.9892	3.4170	3.9205	13.5680	4.9230	12.6070
sweep	goalflow	0	0.9120	+0.48	+0.08	4.6051	0.5282	0.6151	0.8865	0.4750	3.7640	4.3710	12.8275	4.5030	12.0780
sweep	goalflow	0.25	0.9235	+0.50	+1.42	4.7899	0.4476	0.6154	0.7507	0.6462	3.3200	3.6700	11.8880	4.4930	11.9030
sweep	goalflow	0.5	0.9191	+0.86	+1.67	4.7600	0.7489	0.8253	1.3588	0.8111	3.3030	3.6760	11.8095	4.4290	11.7950
sweep	goalflow	0.75	0.9256	+1.50	+1.73	4.8250	1.3531	1.2791	2.2369	1.0610	3.2000	3.5730	11.7940	4.4070	11.9710
sweep	goalflow	1	0.9354	+1.99	+2.01	4.8689	1.7530	1.5476	3.1498	1.2741	3.1210	3.4700	11.6940	4.4230	11.8470
sweep	goalflow	T	0.9030	+1.93	-1.26	4.6061	1.6850	1.7739	3.7930	0.7634	3.7250	4.3440	13.8280	4.8980	12.7450
sweep	transfuser	0	0.9120	+0.48	+0.08	4.6051	0.5282	0.6151	0.8865	0.4750	3.7640	4.3710	12.8275	4.5030	12.0780
sweep	transfuser	0.25	0.9153	+0.29	+1.58	4.8027	0.3722	0.3703	0.4809	0.5427	3.3120	3.5900	11.8520	4.4420	11.9100
sweep	transfuser	0.5	0.9158	+0.20	+1.64	4.8546	0.3340	0.2893	0.4664	0.4407	3.2710	3.5230	11.7300	4.4420	11.9550
sweep	transfuser	0.75	0.9112	+0.14	+1.97	4.8675	0.3081	0.2427	0.4518	0.3705	3.2220	3.4175	11.6560	4.3890	11.8780
sweep	transfuser	1	0.9004	+0.10	+2.14	4.8629	0.2834	0.2142	0.4582	0.3248	3.1855	3.3540	11.3445	4.3690	11.8400
sweep	transfuser	T	0.8400	+0.06	-1.58	4.5679	0.2320	0.2474	0.4006	0.2626	3.8750	4.5360	13.9250	4.8690	12.7710

Table 7. Full v1 NAVTEST results for per-teacher α sweeps, aggregated across eight shards with weights proportional to the number of scenarios. The human serves as the SAS reference ($H = P = 0$), while the base is the shared policy at $\alpha = 0$; $\alpha = T$ denotes the native teacher. H and P are the SAS axes defined in Sec. 4.2. Sub-metrics: $\bar{v}$ mean speed; a^+/a^- peak acceleration/deceleration; j_ℓ/j_t longitudinal/lateral jerk; thw/ttc minimum headway/time-to-collision; lead/ $d_{\min}/d_{\text{vru}}$ minimum lead/object/VRU distances.

scope	teacher	α	PDMS $\uparrow$	H	P	$\bar{v}$	a^+	a^-	j_ℓ	j_t	thw	ttc	lead	$d_{\min}$	d_{vru}
combo	ddv2+goalflow 50/50	0	0.9120	+0.48	+0.08	4.6051	0.5282	0.6151	0.8865	0.4750	3.7640	4.3710	12.8275	4.5030	12.0780
combo	ddv2+goalflow 50/50	0.5	0.9520	+1.06	+1.73	4.8950	1.0565	0.9788	1.7424	0.7410	3.3390	3.7605	12.0920	4.4030	11.6645
combo	ddv2+goalflow 50/50	1	0.9500	+2.64	+2.45	4.8309	2.4876	1.8331	3.6637	1.7195	3.0040	3.4865	12.5640	4.3570	11.6665
combo	ddv2+transfuser 70/30	0	0.9120	+0.48	+0.08	4.6051	0.5282	0.6151	0.8865	0.4750	3.7640	4.3710	12.8275	4.5030	12.0780
combo	ddv2+transfuser 70/30	0.5	0.9437	+1.06	+1.76	4.7260	1.0065	0.9937	1.6054	0.8264	3.3110	3.9040	12.8625	4.2290	11.8385
combo	ddv2+transfuser 70/30	1	0.9480	+2.92	+2.76	4.8494	2.5248	2.0674	3.7133	2.1054	3.1130	3.3615	12.1140	4.1460	11.7070
combo	ddv2+transfuser 50/50	0	0.9120	+0.48	+0.08	4.6051	0.5282	0.6151	0.8865	0.4750	3.7640	4.3710	12.8275	4.5030	12.0780
combo	ddv2+transfuser 50/50	0.5	0.9456	+0.71	+1.61	4.8358	0.7557	0.6825	1.1434	0.6745	3.2300	3.8975	12.5060	4.4110	11.7720
combo	ddv2+transfuser 50/50	1	0.9475	+2.11	+1.86	4.9691	2.0326	1.5691	2.8817	1.3920	3.0725	3.6615	11.9565	4.4330	11.8725
combo	ddv2+transfuser 30/70	0	0.9120	+0.48	+0.08	4.6051	0.5282	0.6151	0.8865	0.4750	3.7640	4.3710	12.8275	4.5030	12.0780
combo	ddv2+transfuser 30/70	0.5	0.9089	+0.46	+2.04	4.7204	0.5644	0.4788	0.7054	0.5833	3.1725	3.4600	11.2675	4.3330	11.9235
combo	ddv2+transfuser 30/70	1	0.9008	+0.74	+2.74	4.9728	0.7749	0.5997	1.1111	0.7917	3.0730	3.4105	11.2915	4.2370	11.5775
combo	transfuser+goalflow 50/50	0	0.9120	+0.48	+0.08	4.6051	0.5282	0.6151	0.8865	0.4750	3.7640	4.3710	12.8275	4.5030	12.0780
combo	transfuser+goalflow 50/50	0.5	0.9248	+0.55	+1.92	4.6543	0.5435	0.6262	0.8534	0.6315	3.3145	3.8360	11.8075	4.2920	11.6380
combo	transfuser+goalflow 50/50	1	0.9206	+1.06	+2.33	4.7809	1.0321	0.9105	1.5680	0.8838	3.2060	3.6705	11.0475	4.2370	11.5960
combo	ddv2+gtrs 60/40	0	0.9120	+0.48	+0.08	4.6051	0.5282	0.6151	0.8865	0.4750	3.7640	4.3710	12.8275	4.5030	12.0780
combo	ddv2+gtrs 60/40	0.5	0.9521	+0.92	+1.79	4.8853	0.8642	0.8730	1.4049	0.7874	3.1620	3.5695	11.8470	4.4390	11.9000
combo	ddv2+gtrs 60/40	1	0.9543	+2.69	+2.45	5.0393	2.4151	1.9628	3.5283	1.8284	2.9325	3.2810	11.6675	4.3950	11.8510
combo	ddv2+hydra 60/40	0	0.9120	+0.48	+0.08	4.6051	0.5282	0.6151	0.8865	0.4750	3.7640	4.3710	12.8275	4.5030	12.0780
combo	ddv2+hydra 60/40	0.5	0.9499	+0.92	+1.93	4.8883	0.8822	0.8653	1.4252	0.7692	3.1735	3.5015	11.7565	4.4490	11.8000
combo	ddv2+hydra 60/40	1	0.9519	+2.68	+2.60	5.0141	2.3923	1.9392	3.4845	1.8634	2.8890	3.2330	11.8015	4.3970	11.7950
combo	ddv2+goalflow 60/40	0	0.9120	+0.48	+0.08	4.6051	0.5282	0.6151	0.8865	0.4750	3.7640	4.3710	12.8275	4.5030	12.0780
combo	ddv2+goalflow 60/40	0.5	0.9452	+1.24	+1.65	4.8364	1.1836	1.1556	1.8826	0.8407	3.1810	3.5640	11.9270	4.4530	11.9990
combo	ddv2+goalflow 60/40	1	0.9494	+2.96	+2.27	4.9343	2.6060	2.0955	3.9844	2.0162	3.0575	3.3890	11.8915	4.4170	11.7530
combo	ddv2+goalflow+hydra 40/40/20	0	0.9120	+0.48	+0.08	4.6051	0.5282	0.6151	0.8865	0.4750	3.7640	4.3710	12.8275	4.5030	12.0780
combo	ddv2+goalflow+hydra 40/40/20	0.5	0.9387	+0.91	+1.98	4.9062	0.8805	0.8821	1.4372	0.7460	3.1440	3.5290	11.6810	4.4320	11.7880
combo	ddv2+goalflow+hydra 40/40/20	1	0.9496	+2.42	+2.37	5.0126	2.3134	1.7531	3.3477	1.5544	2.9870	3.3260	11.6890	4.4140	11.7960
combo	ddv2+goalflow+wote 45/45/10	0	0.9120	+0.48	+0.08	4.6051	0.5282	0.6151	0.8865	0.4750	3.7640	4.3710	12.8275	4.5030	12.0780
combo	ddv2+goalflow+wote 45/45/10	0.5	0.9319	+1.03	+1.96	4.8539	0.9849	0.9920	1.6205	0.7741	3.1700	3.5400	11.6570	4.4300	11.7750
combo	ddv2+goalflow+wote 45/45/10	1	0.9490	+2.61	+2.41	4.9875	2.3998	1.8981	3.5852	1.6949	2.9790	3.2710	11.9215	4.4540	11.7420
combo	ddv2+goalflow+gtrs 40/40/20	0	0.9120	+0.48	+0.08	4.6051	0.5282	0.6151	0.8865	0.4750	3.7640	4.3710	12.8275	4.5030	12.0780
combo	ddv2+goalflow+gtrs 40/40/20	0.5	0.9465	+0.91	+1.98	4.9022	0.8959	0.8772	1.4354	0.7268	3.1600	3.5270	11.7180	4.4330	11.7760
combo	ddv2+goalflow+gtrs 40/40/20	1	0.9459	+2.46	+2.42	5.0072	2.3323	1.7972	3.4049	1.5534	2.9580	3.2970	11.5530	4.4290	11.7800
combo	ddv2+goalflow+transfuser+gtrs 35/35/15/15	0	0.9120	+0.48	+0.08	4.6051	0.5282	0.6151	0.8865	0.4750	3.7640	4.3710	12.8275	4.5030	12.0780
combo	ddv2+goalflow+transfuser+gtrs 35/35/15/15	0.5	0.9377	+0.81	+1.97	4.8538	0.7870	0.8131	1.2800	0.7112	3.1810	3.4970	11.6360	4.4230	11.7960
combo	ddv2+goalflow+transfuser+gtrs 35/35/15/15	1	0.9402	+2.13	+2.46	5.0196	2.0699	1.5952	3.0159	1.3398	3.0240	3.3060	11.4245	4.4000	11.7130
combo	goalflow+gtrs 50/50	0	0.9120	+0.48	+0.08	4.6051	0.5282	0.6151	0.8865	0.4750	3.7640	4.3710	12.8275	4.5030	12.0780
combo	goalflow+gtrs 50/50	0.5	0.9246	+0.52	+1.94	4.8540	0.5334	0.5611	0.8342	0.6375	3.2320	3.5480	11.5610	4.4120	11.7490
combo	goalflow+gtrs 50/50	1	0.9299	+1.05	+2.35	4.9193	0.9743	0.9576	1.6231	0.8446	3.1040	3.4010	11.2400	4.3550	11.7250

Table 8. Full v1 NAVTEST results for persona blends. Each configuration specifies the teacher task vectors and their mixing weights, with α controlling the overall strength. We report PDMS, SAS axes H and P , and the individual kinematic and proximity measurements, following the same definitions as Tab. 7. All configurations recover the shared base at $\alpha=0$.

persona	discrete codebook	flow (ours)
base ($\alpha=0$)	0.5	0.5
ddv2	0.8	3.3
goalflow	0.7	2.0
wote	0.1	0.2
transfuser	0.1	0.1
hydra.mdp	0.0	0.1
gtrs	0.0	0.1
ΔH	0.8	3.2

Table 9. Per-persona assertiveness H for the decoder ablation in Tab. 6a. Both decoders use the same style conditioning. The base is evaluated at $\alpha=0$ and each persona at $\alpha=1$. The bottom row reports the measured assertiveness span ΔH .