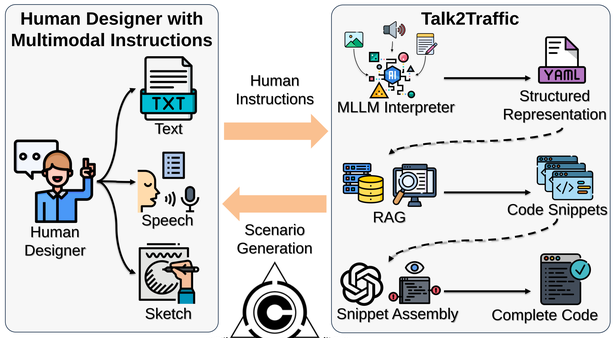
Talk2Traffic: Interactive and Editable Traffic Scenario Generation for Autonomous Driving with Multimodal Large Language Model

Zihao Sheng[1], Zilin Huang[1], Yansong Qu[2], Yue Leng[3], Sikai Chen[1] *

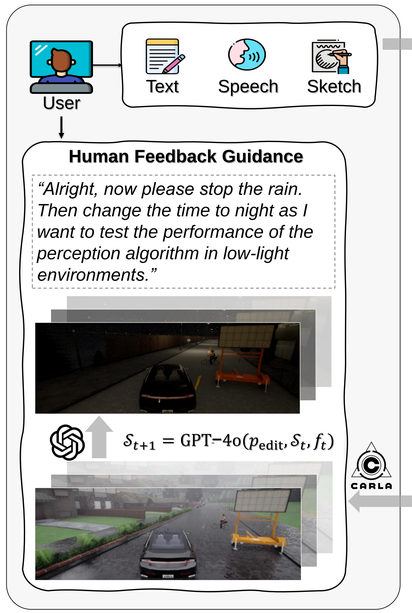
[1]University of Wisconsin-Madison, [2]Purdue University, [3]Google

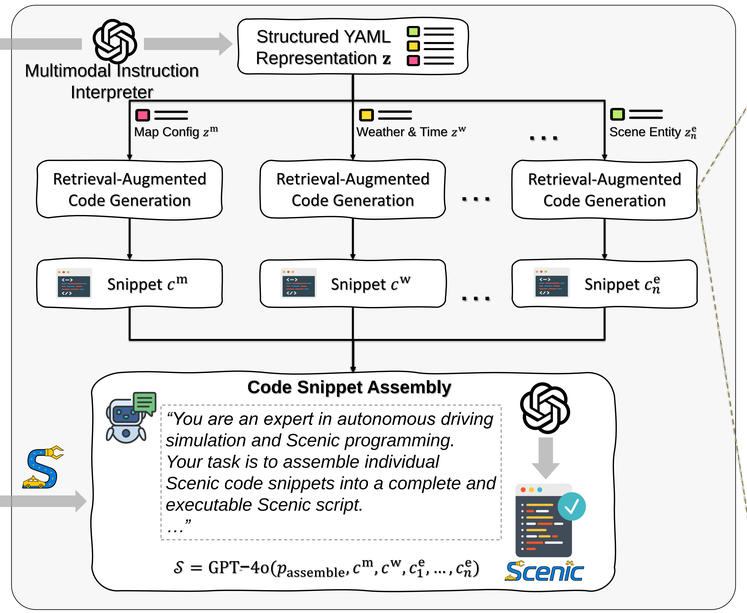
MOTIVATION

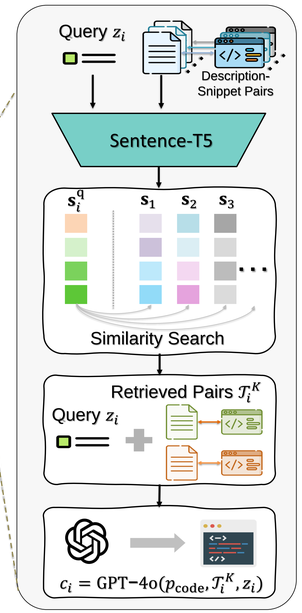
- **Safety validation** of autonomous vehicles (AVs) requires extensive testing in diverse and challenging scenarios.
- Collecting and annotating real-world data is prohibitively **expensive** and often misses **rare edge cases**.
- Rule-based methods offer **limited diversity** and need **expert knowledge** to model interactions.
- Data-driven approaches generate realistic scenarios but usually **lack interactivity and semantic control**.
- Most tools do not support intuitive, human-in-the-loop editing.
- **Multimodal Large Language Models (MLLMs)** provide a promising solution by enabling interactive, editable, and expressive scenario generation from text, speech, and sketches.



METHODOLOGY







Details:

- Multimodal Interpreter: Converts **text, speech, and sketches** into structured representations using MLLMs.
- Retrieval-Augmented Generation: Translates structured representations into **executable Scenic code** by retrieving relevant **verified code snippets**.
- Code Assembly: Combines snippets into complete scripts with **consistent variables and syntax**.
- Human Feedback Loop: Enables **natural language editing** for iterative scenario refinement without coding.

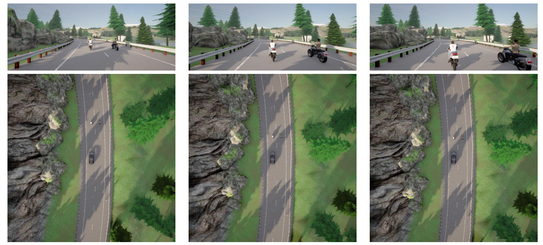
EXPERIMENTS

Metric	Model	Scenario Type: Straight Obstacle	Turning Obstacle	Lane Change	Vehicle Passing	Red Light Running	Unprotected Left Turn	Right Turn	Crossing Negotiation	Avg.
CR	LC	0.223	0.088	0.710	0.807	0.317	0.403	0.350	0.273	0.396
	AS	0.470	0.350	0.703	0.833	0.497	0.647	0.637	0.607	0.593
	AT	0.343	0.217	0.667	0.83	0.623	0.533	0.357	0.393	0.495
	CS	0.893	0.697	**0.950**	0.927	0.787	0.753	0.777	0.863	0.831
	T2T	**0.913**	**0.78**	0.893	**0.947**	**0.900**	**0.833**	**0.86**	**0.893**	**0.877**
OS	LC	0.809	0.849	0.566	0.508	0.804	0.753	0.681	0.748	0.715
	AS	0.697	0.708	0.563	0.504	0.711	0.628	0.54	0.575	0.616
	AT	0.747	0.776	0.583	0.507	0.643	0.684	0.68	0.684	0.663
	CS	**0.470**	0.522	**0.434**	0.440	0.537	0.56	0.474	**0.421**	0.482
	T2T	0.475	**0.481**	0.461	**0.437**	**0.496**	**0.539**	**0.449**	0.422	**0.471**

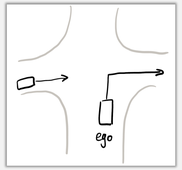
Model	Safety Level: CR	RR	SS	OR	Functionality Level: RF	Comp	TS	Etiquette Level: ACC	YV	LI	OS
LC	0.396	0.316	**0.150**	0.045	0.883	0.809	0.255	0.228	0.228	0.090	0.715
AS	0.593	0.301	0.148	0.041	0.884	0.742	0.255	0.242	0.226	0.094	0.616
AT	0.495	**0.315**	**0.150**	**0.052**	0.884	0.769	0.238	0.249	0.233	0.103	0.663
CS	0.831	0.179	0.143	0.035	**0.833**	0.544	0.223	**0.705**	**0.532**	**0.243**	0.482
T2T	**0.877**	0.226	0.148	0.013	0.895	**0.519**	**0.284**	0.322	0.262	0.060	**0.471**

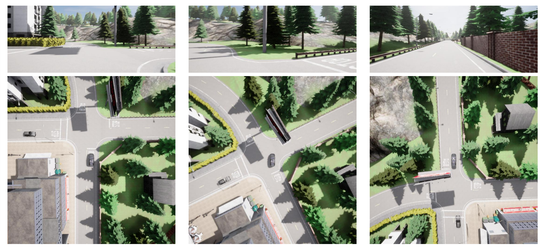
(a) Speech

Two slow moving motorcycles blocked the ego vehicle from moving forward under a clear dusk sky. The ego vehicle has to brake to ensure safety.

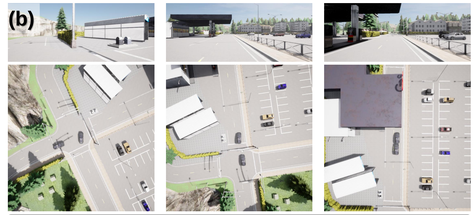


(b) Sketch





Human Feedback: Change the number of lanes to 3 and allow the ego vehicle to change lanes and overtake them.



Human Feedback: Increase the speed of the straight-moving car to simulate losing control of the throttle and accelerating into the ego car.