A Survey of Autonomous Driving Testing Practices and Challenges

Part 1 - Background Investigation

*What type of organization are you working for?

- ◯ Technology companies
- ◯ Automobile manufacturing companies
- ◯ Research Institutes
- ◯ Governments
- ◯ Other (Please specify)

*What is your role in the organization?

- ◯ Software Engineer
- ◯ Software Architect / Tech Lead
- ◯ Quality Assurance / Software Tester
- ◯ Research Scientist
- ◯ Director / Manager
- ◯ Safety Engineer
- ◯ Self-employed
- ◯ Other (Please specify)

*Which gender identity do you most identify with?

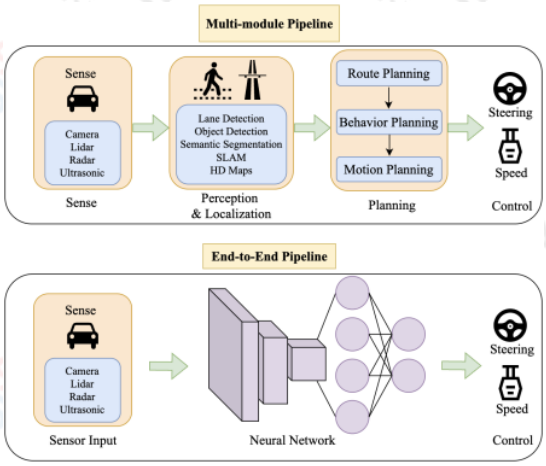
- ◯ Female
- ◯ Male
- ◯ Prefer not to say
- ◯ Other (Please specify)

*How long have you been conducting research in the Autonomous Driving field?

- ◯ Less than 1 year
- ◯ 1 ~ 5 years
- ◯ 5 ~ 10 years
- ◯ More than 10 years

Part 2 – Current Practice for Autonomous Driving System (ADS)

*What kind of ADS are you working on? (If your answer is multiple, please choose the field that is most familiar to you.)



- ◯ Multi-module structure
- ◯ End-to-end (E2E) system
- ◯ Other (Please specify)

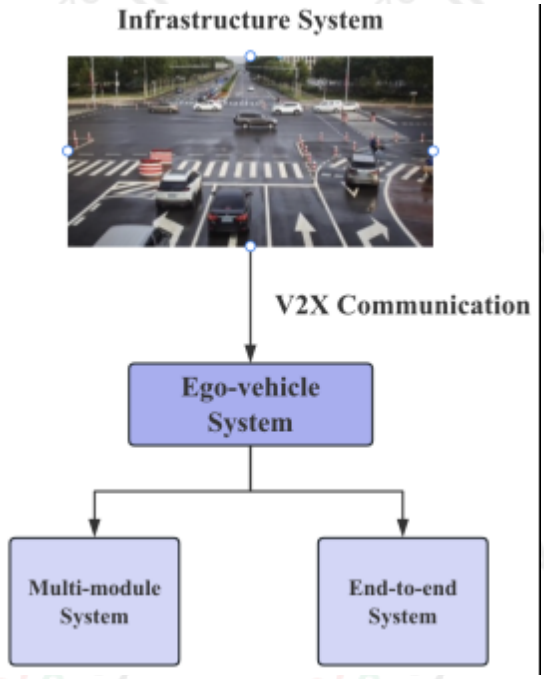
*What is your ADS based on? (check all that apply)

- ☐ Machine Learning
- ☐ Deep Learning
- ☐ Reinforcement Learning
- ☐ Other (Please specify)

*What sensors is your ADS equipped with? (check all that apply)

- ☐ Camera
- ☐ LiDAR
- ☐ Radar
- ☐ GPS & IMU
- ☐ Other (Please specify)

*Do you use Vehicle-to-everything (V2X) communication?



- ( ) Yes, I use V2X on Multi-module: Leveraging V2X communication for a single task (e.g., object detection).
- ( ) Yes, I use V2X on E2E (e.g., UniV2X): Seamlessly integrating all key driving modules (e.g., perception, planning) across diverse views (including other vehicles' views) into a unified network.
- ( ) Yes, I use V2X on the ADSs other than E2E and multi-module.
- ( ) No.

*What kind of data do you use for V2X Transmission? (check all that apply)

- [ ] Raw data (e.g., LiDAR point clouds, RGB images)
- [ ] Intermediate-level data (e.g., BEV features)
- [ ] Other (Please specify)

*Have you ever conducted testing on the V2X system?

- ( ) Yes
- ( ) No

*What kind of systems have you tested? (check all that apply)

- [ ] V2X perception systems
- [ ] V2X planning systems
- [ ] Other (Please specify)

*What sensors do you conduct V2X testing on? (check all that apply)

- [ ] LiDAR (e.g., Point Clouds)
- [ ] Camera (e.g., RGB images)
- [ ] Radar (e.g., Point Clouds)
- [ ] Other (Please specify)

*What testing techniques do you use on the V2X system? (check all that apply)

- [ ] Black-box (query access to the model, i.e., only know inputs and outputs)
- [ ] White-box (with information, e.g., data, parameter, or model structure)
- [ ] Grey-box (with partial information)
- [ ] Unknown
- [ ] Other (Please specify)

*What testing methods do you use on the V2X system? (check all that apply)

- [ ] Knowledge-based approaches (e.g., metamorphic relation, ontology, adversarial)
- [ ] Search-based approaches (e.g., driving pattern, classification model, specification)
- [ ] Data driven approaches (scenarios are identified, extracted, created or clustered data)
- [ ] Unknown
- [ ] Other (Please specify)

*Do you use public or private datasets? (check all that apply)

- [ ] Public datasets
- [ ] Private datasets
- [ ] Other (Please specify)

*Public datasets you have used:

*What testing scenarios do you use for testing ADS? (check all that apply)

- [ ] Simulation testing (testing on virtual scenarios in a simulator, e.g., CARLA)
- [ ] Real-world testing (testing on real-world scenarios)
- [ ] Other (Please specify)

*What simulation platforms are you use? (check all that apply)

- [ ] BeamNG
- [ ] CARLA
- [ ] Baidu Apollo
- [ ] AWSIM (Autoware)
- [ ] Other (Please specify)

*For Testing ADS, which type of testing do you use? (check all that apply)

- [ ] Online Testing (conduct testing within a real-time environment)
- [ ] Offline Testing (conduct testing in pre-acquired/fixed scenarios)
- [ ] Other (Please specify)

Multi-module Structure

What level do you conduct testing in ADS? (check all that apply)

- [ ] Module-level (conduct testing on each module separately, e.g., perception, prediction, planning, control)
- [ ] System-level (conduct testing on the whole ADS)

Multi-module: Module-level Testing

*Which units are you testing? (check all that apply)

- [ ] Perception Module
- [ ] Planning Module
- [ ] Control Module
- [ ] Other (Please specify)

Module-level Testing: Perception Module

*What is your Perception module based on? (check all that apply)

- [ ] Machine Learning
- [ ] Deep Learning
- [ ] Reinforcement Learning
- [ ] Other (Please specify)

*In the Perception module, what specific tasks are you working on? (check all that apply)

- [ ] Semantic Segmentation
- [ ] Objective Detection
- [ ] Objective Tracking
- [ ] Environmental Understanding
- [ ] Other (Please specify)

*In the Perception module, what sensors do you conduct testing on? (check all that apply)

- [ ] LiDAR (e.g., Point Clouds)
- [ ] Camera (e.g., RGB images)
- [ ] Radar (e.g., Point Clouds)
- [ ] Other (Please specify)

*What testing techniques do you use for testing Perception module? (check all that apply)

- [ ] Black-box (query access to the model, i.e., only know inputs and outputs)
- [ ] White-box (with full information, e.g., data, parameter, or model structure)
- [ ] Grey-box (with partial information)
- [ ] Unknown
- [ ] Other (Please specify)

*What testing methods do you use for testing Perception module? (check all that apply)

- [ ] Knowledge-based approaches (e.g., metamorphic relation, ontology, adversarial)
- [ ] Search-based approaches (e.g., driving pattern, classification model, specification)
- [ ] Data driven approaches (scenarios are identified, extracted, created or clustered data)
- [ ] Unknown
- [ ] Other (Please specify)

*The current methods you use for testing the Perception module can meet the testing requirements. (Strong disagree to strong agree)

Module-level Testing: Planning Module

*What is your Planning module based on? (check all that apply)

- [ ] Machine Learning
- [ ] Deep Learning
- [ ] Reinforcement Learning
- [ ] Other (Please specify)

*In the Planning module, what specific tasks are you working on? (check all that apply)

- [ ] Path Planning
- [ ] Behavior Planning
- [ ] Scenario Planning
- [ ] Other (Please specify)

*What testing techniques do you use for testing Planning module? (check all that apply)

- [ ] Black-box (query access to the model, i.e., only know inputs and outputs)
- [ ] White-box (with full information, e.g., data, parameter, or model structure)
- [ ] Grey-box (with partial information)
- [ ] Unknown
- [ ] Other (Please specify)

*What testing methods do you use for testing Planning module? (check all that apply)

- [ ] Knowledge-based approaches (e.g., metamorphic relation, ontology, adversarial)
- [ ] Search-based approaches (e.g., driving pattern, classification model, specification)
- [ ] Data driven approaches (scenarios are identified, extracted, created or clustered data)
- [ ] Unknown
- [ ] Other (Please specify)

*The current methods you use for testing the Planning module can meet the testing requirements. (Strong disagree to strong agree)

Module-level Testing: Control Module

*What is your Control module based on? (check all that apply)

- [ ] Machine Learning
- [ ] Deep Learning
- [ ] Reinforcement Learning
- [ ] Other (Please specify)

*In Control module, what specific tasks are you working on? (check all that apply)

- [ ] Steering Control
- [ ] Speed Control
- [ ] Stability Control
- [ ] Other (Please specify)

*What testing techniques do you use for testing Control module? (check all that apply)

- [ ] Black-box (query access to the model, i.e., only know inputs and outputs)
- [ ] White-box (with full information, e.g., data, parameter, or model structure)
- [ ] Grey-box (with partial information)
- [ ] Unknown
- [ ] Other (Please specify)

*What testing methods do you use for testing Control module? (check all that apply)

- [ ] Knowledge-based approaches (e.g., metamorphic relation, ontology, adversarial)
- [ ] Search-based approaches (e.g., driving pattern, classification model, specification)
- [ ] Data driven approaches (scenarios are identified, extracted, created or clustered data)
- [ ] Unknown
- [ ] Other (Please specify)

*The current methods you use for testing the Control module can meet the testing requirements. (Strong disagree to strong agree)

Module-level Testing: Other Module(s)

*What module(s) are you testing?

*What is your other module(s) based on? (check all that apply)

- [ ] Machine Learning
- [ ] Deep Learning
- [ ] Reinforcement Learning
- [ ] Other (Please specify)

*In other module(s), what specific tasks are you working on?

*What testing techniques do you use for testing other module(s)? (check all that apply)

- [ ] Black-box (query access to the model, i.e., only know inputs and outputs)
- [ ] White-box (with full information, e.g., data, parameter, or model structure)
- [ ] Grey-box (with partial information)
- [ ] Unknown
- [ ] Other (Please specify)

*What testing methods do you use for testing other module(s)? (check all that apply)

- [ ] Knowledge-based approaches (e.g., metamorphic relation, ontology, adversarial)
- [ ] Search-based approaches (e.g., driving pattern, classification model, specification)
- [ ] Data driven approaches (scenarios are identified, extracted, created or clustered data)
- [ ] Unknown
- [ ] Other (Please specify)

*The current methods you use for testing the other module(s) can meet the testing requirements. (Strong disagree to strong agree)

Multi-module Testing: LLMs/VFMs/Others Investigation

*Have you ever used Large Language Models (LLMs: e.g., GPT, BERT), Vision Foundation Models (VFMs: e.g., CLIP, DINO) for multi-module ADS? (check all that apply)

- [ ] Yes, I have used LLMs for testing multi-module ADS
- [ ] Yes, I have used VFMs for testing multi-module ADS
- [ ] Yes, I have used LLM-based multi-module ADS
- [ ] Yes, I have used VFM-based multi-module ADS
- [ ] Other (I use other emerging models for multi-module ADS)
- [ ] None of the above

*What other emerging models do you use for ADS?

LLMs for Multi-module Testing

*Which module(s) in ADS do you use the LLMs for testing? (check all that apply)

- [ ] Perception Module
- [ ] Planning Module
- [ ] Control Module
- [ ] Other (Please specify)

*How do you use LLMs for testing? (check all that apply)

- [ ] Generate critical scenarios
- [ ] Retraining ADS
- [ ] Other (Please specify)

*Compared with other testing methods, the introduction of LLMs into ADS testing improves the testing capability. (Strong disagree to strong agree)

VFMs for Multi-module Testing

Which module(s) in ADS do you use the VFMs for testing? (check all that apply)

- [ ] Perception Module
- [ ] Planning Module
- [ ] Control Module
- [ ] Other (Please specify)

*How do you use VFMs for testing? (check all that apply)

- [ ] Generate critical scenarios
- [ ] Retraining ADS
- [ ] Other (Please specify)

*Compared with other testing methods, the introduction of VFMs into ADS testing improves the testing capability. (Strong disagree to strong agree)

LLMs-based Multi-module ADS

*Which module(s) in ADS do you use the LLMs? (check all that apply)

- [ ] Perception Module
- [ ] Planning Module
- [ ] Control Module
- [ ] Other (Please specify)

*Compared with other ADS, the introduction of LLMs-based ADS improves the performance. (Strong disagree to strong agree)

VFMs-based Multi-module ADS

*Which module(s) in ADS do you use the VFMs? (check all that apply)

- [ ] Perception Module
- [ ] Planning Module
- [ ] Control Module
- [ ] Other (Please specify)

*Compared with other ADS, the introduction of VFMs-based ADS improves the performance. (Strong disagree to strong agree)

Other Model(s) for ADS

*Please simply specify the base structure of the model(s) you use.

*How do you use this emerging model(s)? (check all that apply)

- [ ] For ADS testing
- [ ] For emeging-model(s)-based ADS
- [ ] Other (Please specify)

*Which module(s) in ADS do you use the model(s)? (check all that apply)

- [ ] Perception Module
- [ ] Planning Module
- [ ] Control Module
- [ ] Other (Please specify)

*Compared with other models, the introduction of the emerging improves the performance. (Strong disagree to strong agree)

Multi-module: System-level Testing

*At system-level testing, what testing techniques do you use? (check all that apply)

- [ ] Black-box (query access to the model, i.e., only know inputs and outputs)
- [ ] White-box (with information, e.g., data, parameter, or model structure)
- [ ] Grey-box (with partial information)
- [ ] Unknown

*At system-level testing, what testing methods do you use? (check all that apply)

- [ ] Knowledge-based approaches (e.g., metamorphic relation, ontology, adversarial)
- [ ] Search-based approaches (e.g., driving pattern, classification model, specification)
- [ ] Data driven approaches (scenarios are identified, extracted, created or clustered data)
- [ ] Unknown
- [ ] Other (Please specify)

*The current methods you use for system-level testing can meet the testing requirements. (Strong disagree to strong agree)

*At system-level testing for ADS, have you ever used Large Language Models (LLMs: e.g., GPT, BERT), Vision Foundation Models (VFMs: e.g., CLIP, DINO)? (check all that apply)

- [ ] Yes, I have used LLMs for testing ADS
- [ ] Yes, I have used VFMs for testing ADS
- [ ] Yes, I have used LLM-based ADS
- [ ] Yes, I have used VFM-based ADS
- [ ] Other (I use other emerging models for ADS)
- [ ] None of the above

*What other emerging models have you used for system-level testing? (Please simply specify the base structure)

*At system-level testing, how do you use LLMs for testing? (check all that apply)

- [ ] Generate critical scenarios
- [ ] Retraining ADS
- [ ] Other (Please specify)

*At system-level testing, how do you use VFMs for testing? (check all that apply)

- [ ] Generate critical scenarios
- [ ] Retraining ADS
- [ ] Other (Please specify)

*Compared with other testing methods, the introduction of LLMs into ADS testing improves the system-level testing capability. (Strong disagree to strong agree)

*Compared with other testing methods, the introduction of VFMs into ADS testing improves the system-level testing capability. (Strong disagree to strong agree)

*Compared with other ADS, the introduction of LLMs-based ADS improves the performance. (Strong disagree to strong agree)

*Compared with other ADS, the introduction of VFMs-based ADS improves the performance. (Strong disagree to strong agree)

*Compared with other ADS, the introduction of other emerging model(s)-based ADS improves the performance. (Strong disagree to strong agree)

End-to-end (E2E) System

*In your E2E system, what are the output control actions? (check all that apply)

- [ ] Steering
- [ ] Speed (e.g, Brake, Throttle)
- [ ] Trajectory
- [ ] Location
- [ ] Other (Please specify)

*In the E2E system, what sensors do you conduct testing on? (check all that apply)

- [ ] LiDAR (e.g., Point Clouds)
- [ ] Camera (e.g., RGB images)
- [ ] Radar (e.g., Point Clouds)
- [ ] Other (Please specify)

*In your E2E system, what testing techniques do you use? (check all that apply)

- [ ] Black-box (query access to the model, i.e., only know inputs and outputs)
- [ ] White-box (with information, e.g., data, parameter, or model structure)
- [ ] Grey-box (with partial information)
- [ ] Unknown
- [ ] Other (Please specify)

*In your E2E system, what testing methods do you use? (check all that apply)

- [ ] Knowledge-based approaches (e.g., metamorphic relation, ontology, adversarial)
- [ ] Search-based approaches (e.g., driving pattern, classification model, specification)
- [ ] Data driven approaches (scenarios are identified, extracted, created or clustered data)
- [ ] Unknown
- [ ] Other (Please specify)

*The current methods you use for testing E2E system can meet the testing requirements. (Strong disagree to strong agree)

*In your E2E system, have you ever used Large Language Models (LLMs: e.g., GPT, BERT), Vision Foundation Models (VFMs: e.g., CLIP, DINO) for ADS? (check all that apply)

- [ ] Yes, I have used LLMs for testing ADS
- [ ] Yes, I have used VFMs for testing ADS
- [ ] Yes, I have used LLM-based ADS
- [ ] Yes, I have used VFM-based ADS
- [ ] Other (I use other emerging models for ADS)
- [ ] None of the above

*What other emerging models do you use for ADS? (Please simply specify the base structure)

*In the E2E system, how do you use LLMs for testing? (check all that apply)

- [ ] Generate critical scenarios
- [ ] Retraining ADS
- [ ] Other (Please specify)

*In the E2E system, how do you use VFMs for testing? (check all that apply)

- [ ] Generate critical scenarios
- [ ] Retraining ADS
- [ ] Other (Please specify)

*Compared with other testing methods, the introduction of LLMs into ADS testing improves the E2E system testing capability. (Strong disagree to strong agree)

*Compared with other testing methods, the introduction of VFMs into ADS testing improves the E2E system testing capability. (Strong disagree to strong agree)

*Compared with other ADS, the introduction of LLMs-based ADS improves the performance. (Strong disagree to strong agree)

*Compared with other ADS, the introduction of VFMs-based ADS improves the performance. (Strong disagree to strong agree)

*Compared with other ADS, the introduction of other emerging model(s)-based ADS improves the performance. (Strong disagree to strong agree)

Other System

*Could you please briefly describe your ADS? (e.g., How is your system different from E2E and multi-module system?)

*In the other system, what are the output control actions of your system? (check all that apply)

- [ ] Steering
- [ ] Speed (brake, throttle)
- [ ] Trajectory
- [ ] Location
- [ ] Environment
- [ ] Other (Please specify)

*In other system, what sensors do you conduct testing on? (check all that apply)

- [ ] LiDAR (e.g., Point Clouds)
- [ ] Camera (e.g., RGB images)
- [ ] Radar (e.g., Point Clouds)
- [ ] Other (Please specify)

*In the other system, what testing techniques do you use? (check all that apply)

- [ ] Black-box (query access to the model, i.e., only know inputs and outputs)
- [ ] White-box (with information, e.g., data, parameter, or model structure)
- [ ] Grey-box (with partial information)
- [ ] Unknown
- [ ] Other (Please specify)

*In the other system, what testing methods do you use for testing the system? (check all that apply)

- [ ] Knowledge-based approaches (e.g., metamorphic relation, ontology, adversarial)
- [ ] Search-based approaches (e.g., driving pattern, classification model, specification)
- [ ] Data driven approaches (scenarios are identified, extracted, created or clustered data)
- [ ] Unknown
- [ ] Other (Please specify)

*The current methods you use for testing the system can meet the testing requirements. (Strong disagree to strong agree)

*In the other system, have you ever used Large Language Models (LLMs: e.g., GPT, BERT), Vision Foundation Models (VFMs: e.g., CLIP, DINO) for ADS? (check all that apply)

- [ ] Yes, I have used LLMs for testing ADS
- [ ] Yes, I have used VFMs for testing ADS
- [ ] Yes, I have used LLM-based ADS
- [ ] Yes, I have used VFM-based ADS
- [ ] Other (I use other emerging models for ADS)
- [ ] None of the above

*What other emerging models do you use for ADS? (Please simply specify the base structure)

*In other system, how do you use LLMs for testing? (check all that apply)

- [ ] Generate critical scenarios
- [ ] Retraining ADS
- [ ] Other (Please specify)

*In other system, how do you use VFMs for testing? (check all that apply)

- [ ] Generate critical scenarios
- [ ] Retraining ADS
- [ ] Other (Please specify)

*Compared with other testing methods, the introduction of LLMs into ADS testing improves the other system testing capability. (Strong disagree to strong agree)

*Compared with other testing methods, the introduction of VFMs into ADS testing improves the other system testing capability. (Strong disagree to strong agree)

*Compared with other testing methods, the introduction of LLMs-based ADS improves the performance. (Strong disagree to strong agree)

*Compared with other ADS, the introduction of VFMs-based ADS improves the performance. (Strong disagree to strong agree)

*Compared with other ADS, the introduction of other emerging model(s)-based ADS improves the performance. (Strong disagree to strong agree)

Part 3 - Security Testing Investigation

*Which type of attacks have you used for testing? (check all that apply)

- [ ] Adversarial Attacks
- [ ] Cyber Attacks
- [ ] Inference Attacks (e.g., membership inference, attribute inference, model stealing, model inversion attacks)
- [ ] Replay Attacks
- [ ] Sensor Attacks (e.g., jamming, spoofing)
- [ ] Physical Attacks (e.g., sensor damage)
- [ ] Other (Please specify)

*What defense methods have you used? (check all that apply)

- [ ] Adversary training or model regularization
- [ ] Encrypted communication, security protocols, identity authentication, or network control
- [ ] Differential privacy or model obfuscation
- [ ] Timestamp and sequence number verification or dynamic key management and authentication
- [ ] Multi-sensor fusion or anomaly detection
- [ ] Physical security design (e.g., set upphysical shields) or monitoring and intrusion detection
- [ ] Other (Please specify)

*What defense tools have you used? (check all that apply)

- [ ] CleverHans, Advbox, or TF-Defense
- [ ] Wireshark, Metasploit, or Snort
- [ ] Differential privacy libraries
- [ ] TLS/SSL or CANsec
- [ ] Anomaly detection systems
- [ ] Anti-jamming materials
- [ ] Other (Please specify)

*Which attacks do you think still pose a threat, but are currently poorly defended? (check all that apply)

- [ ] Adversarial Attacks
- [ ] Cyber Attacks
- [ ] Inference Attacks (e.g., membership inference, attribute inference, model stealing, model inversion attacks)
- [ ] Replay Attacks
- [ ] Sensor Attacks (e.g., jamming, spoofing)
- [ ] Physical Attacks (e.g., sensor damage)
- [ ] Other (Please specify)

Part 4 - Follow Up

If you want to receive the study results, please enter your contact information (email address):

Can we contact you via this email address for clarifications?

- ◯ Yes
- ◯ No

Are you willing to participate in a follow-up questionnaire?

- ◯ Yes
- ◯ No