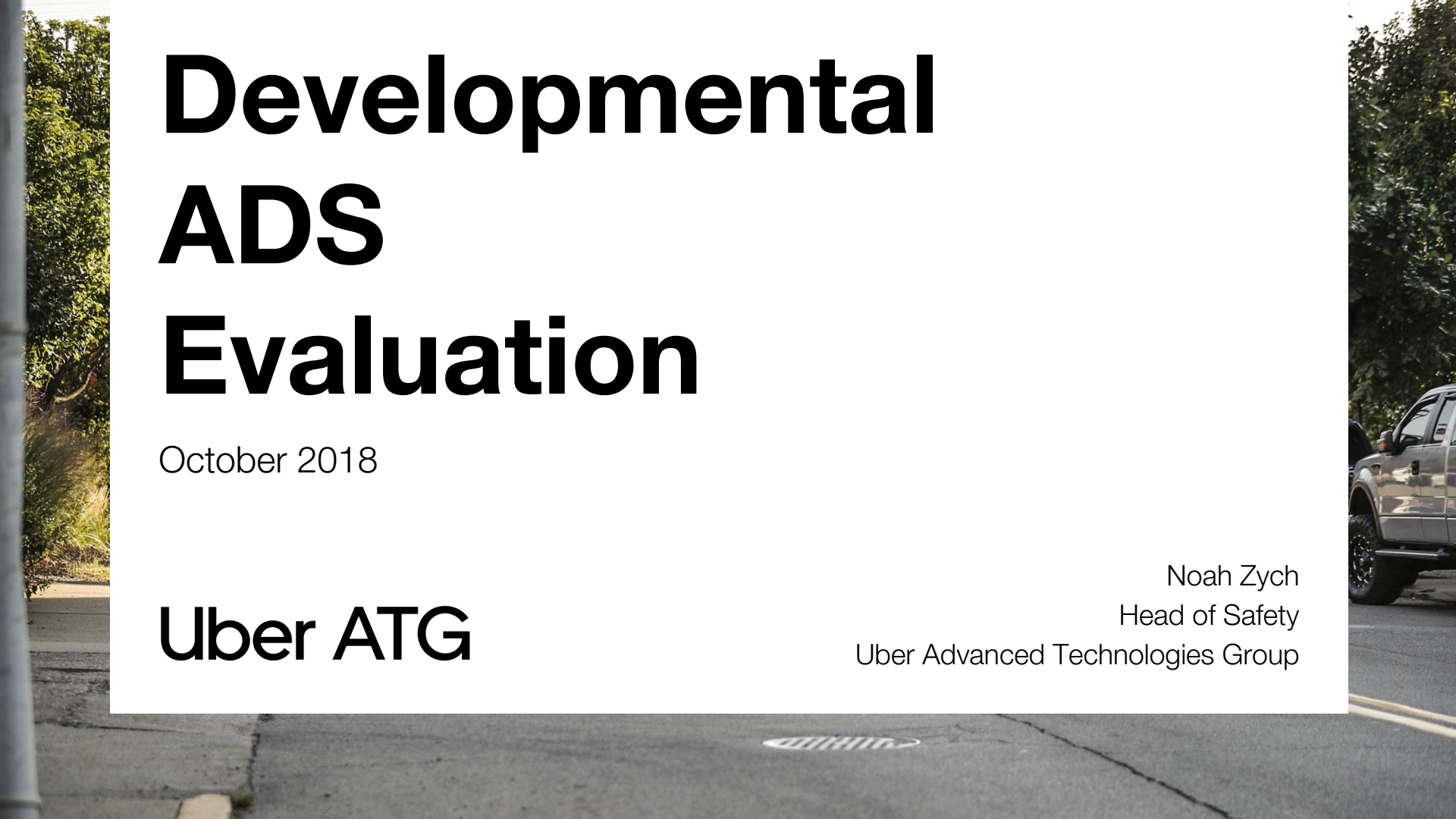


Developmental ADS Evaluation

October 2018

Uber ATG

Noah Zych
Head of Safety
Uber Advanced Technologies Group

Abstract

Uber’s approach to developmental system evaluation:

- Is anchored on a precisely defined operational design domain
- Entails traceable test coverage to required capabilities and anticipated conflict scenarios
- Treats virtual to track to road testing as accumulative, not as independent regimes

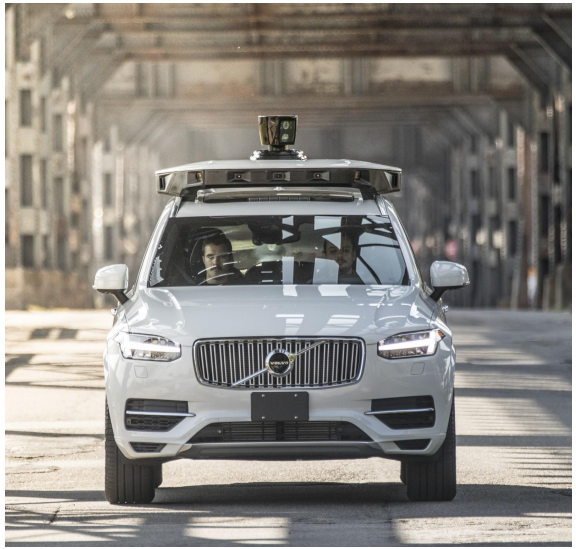
Approach to Development

Specify Operational Domain

Don’t try to tackle a whole city. Start small and precisely define requirements for handling infrastructure, environment, actors, and interactions

Develop Test Strategy

Identify capabilities required and types of interactions expected for the identified ODD. Create virtual and physical test coverage.



Confirm in the Real World

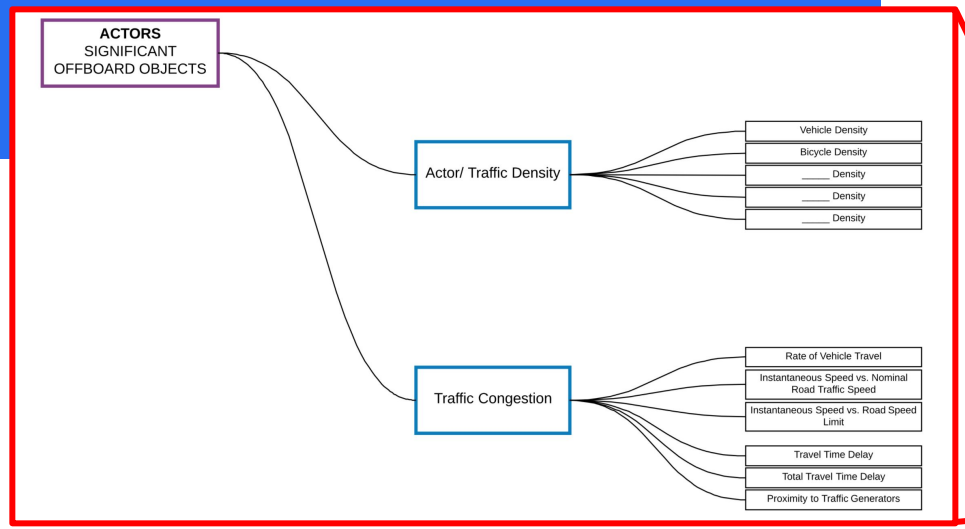
Close the loop with any unexpected behaviors observed under real life conditions → advance development and refine the tests upstream.

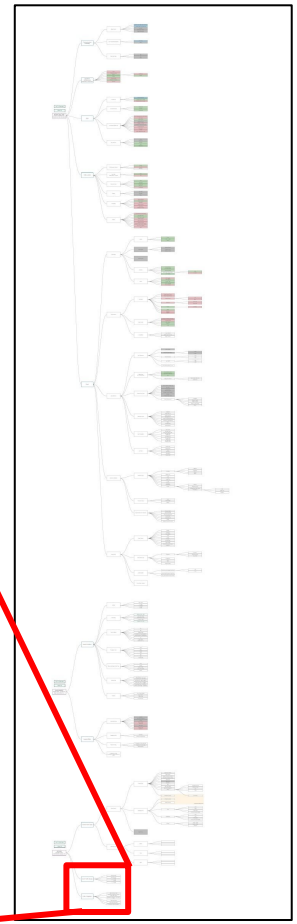
Operational Design Domain

Need a precise and extensive lexicon for describing an ODD

- **Infrastructure**: Signage, signals, road/lane, traffic law and commands
- **Environment**: Natural conditions (weather, geography) and induced effects
- **Actors**: Human/ vehicle/ animal actors, significant objects

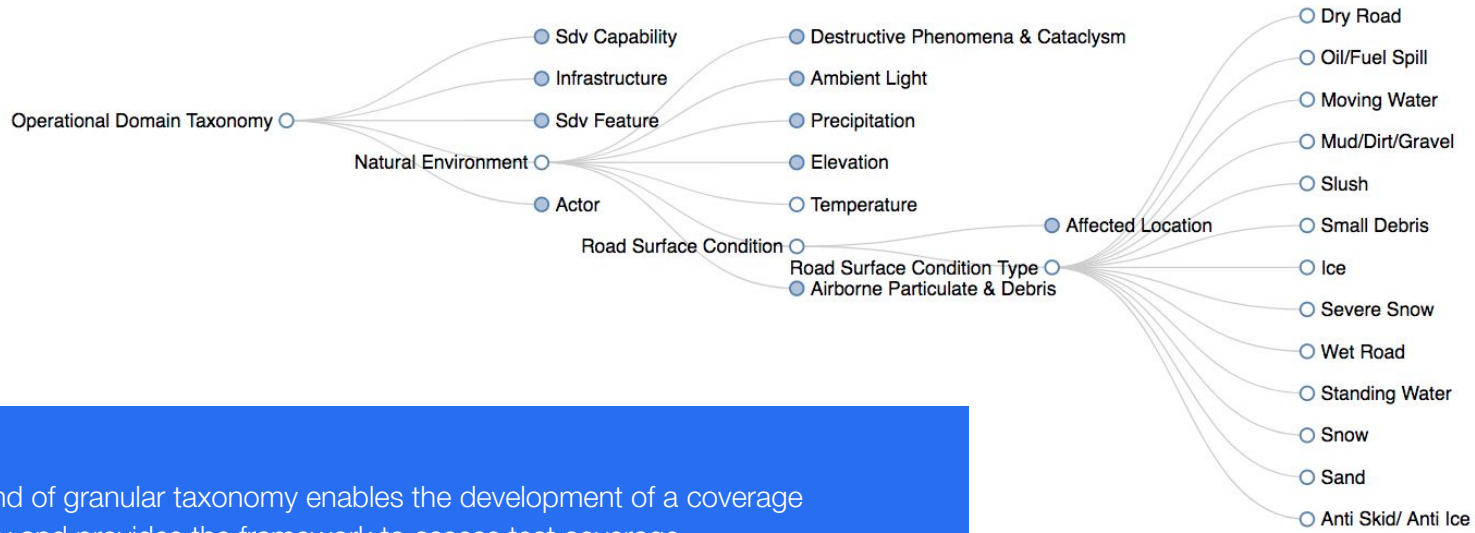
→ **Self-driving performance evaluation scope must be tied to these granular characterizations**





Uber

Operational Design Domain



This kind of granular taxonomy enables the development of a coverage strategy and provides the framework to assess test coverage

Uber

Test Coverage

Test coverage corresponds to ODD

- Perform strategic testing in simulation, on the track, and in supervised road operations → ensure appropriate sampling of rare and common scenarios to test system behavior

Likely Conflict Scenarios
Derived from Euro NCAP, PROSPECT, and our past self-driving experience

Nominal Capabilities
Competencies required for everyday driving: e.g., vehicle following, obeying traffic lights

Domain-Specific Capabilities
Competencies required for driving in a certain ODD: e.g., occluded intersection navigation, rain

Ex: Past Experience

Lessons learned from observed situations on the road are incorporated into tests.

Test & Evaluation Strategy

Accumulative Validation:
Offline | Simulation | Track | Road

- **Offline**: run unit tests prior to integration in the codebase, and additional automated release tests
- **Simulation**: rapidly test myriad variations of challenging scenarios, and test close calls without the physical risk
- **Track**: Execute fault-injection, performance, and field-derived tests on the track
- **Road**: Assess data from supervised driving to derive root causes for takeovers and harsh performance events, and augment test cases based on observed deficiencies

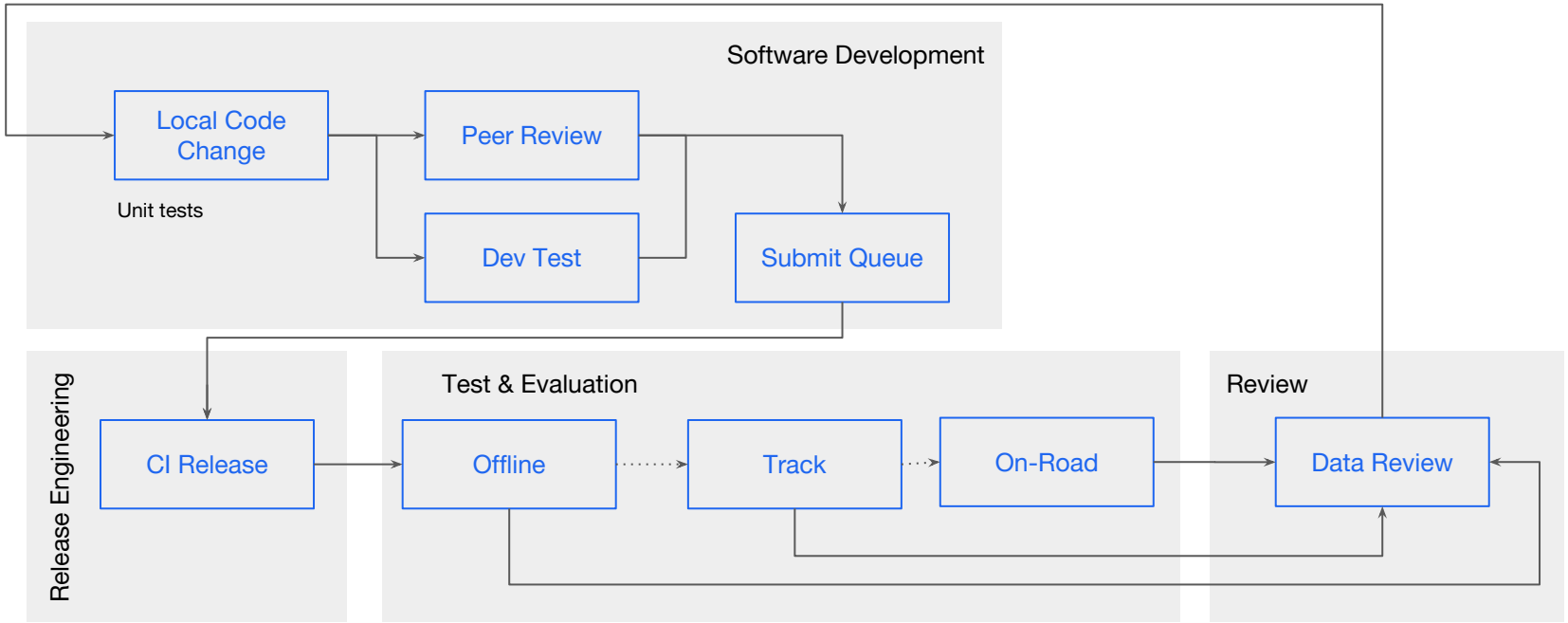
Progressive stages of testing should demonstrate consistency with prior stages

Sim **Track**	Developmental testing is accomplished by engineers, technicians, test personnel in a controlled environment to facilitate failure analysis. ● Verify status of technical progress ● Verify that design risks are minimized ● Certify readiness for initial operational test
Road	Operational testing is accomplished in the real world to facilitate analysis of effectiveness and suitability. ● Verify degree to which the system accomplishes its missions when employed in a realistic scenario ● Evaluate natural environmental effects and impacts, reliability, human factors

[definitions adapted from US DoD Test and Evaluation Management Guide]

Uber

Software Release Process



Track Verification Testing

Broad test suite with established scenario coverage, scoring methodology, and corresponding simulated tests

TVT

Highly repeatable, standardized tests that are regularly run at the test track with clear scoring criteria

Process Gate

Result is binary: Pass/fail. Fail blocks release.

Example TVT for a given ODD

96

Scenario Families

17+

Autonomy Capabilities

Ex: Off-map_Right_Bike_Cut-off_Decel

A bike to the front-right of the SDV in an unmapped lane of travel performs an unprotected right turn maneuver into the SDV lane, then decelerates to a stop.

Success Criteria

(C007) maintain d ≥ 0.2 meters from cyclist (Collision)

(C008) maintain d ≥ 1.2 meters from cyclist (Safety)

(C018) maintain d ≥ 0 meters from any painted lane boundary marking or d ≥ 9 inches from road center or road edge if no lane boundary markings present

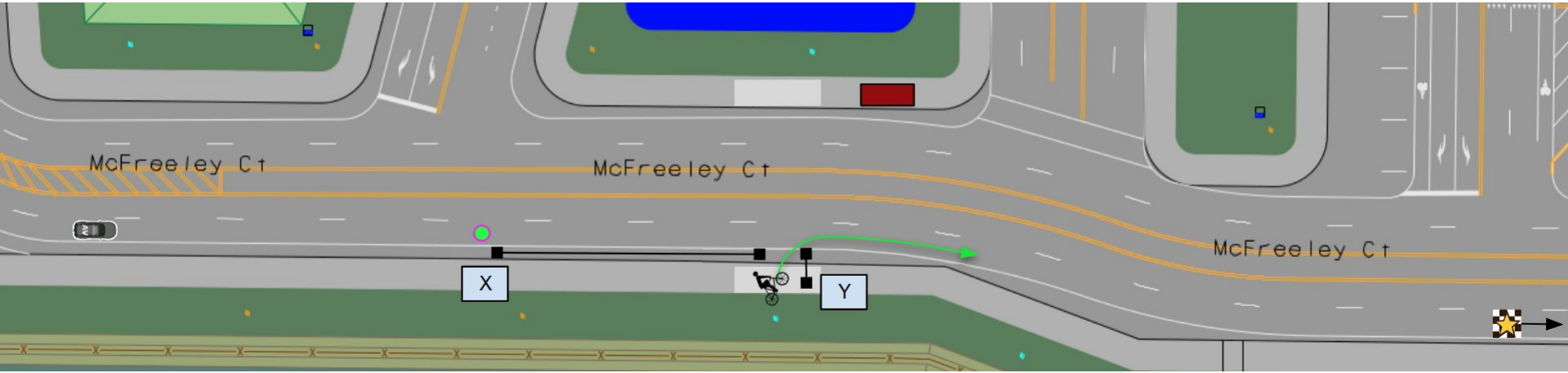
(C028) Acceleration -3.0m/s/s < a < 3.0m/s/s, Jerk -2.0m/s/s/s < j < 2.0m/s/s/s

Track Test Case

Parameter	Value	Tolerance	Units
AV Initial Speed	35	3.5	mph
Trigger distance from actor **X**	40	4	m
Actor front distance from AV lane **Y**	4	0.4	m
Actor initial acceleration	2	0.2	m/s2
Actor max speed	10	1	mph
Actor time at max speed	2	0.2	s
Actor deceleration rate to stop	3	0.3	m/s2

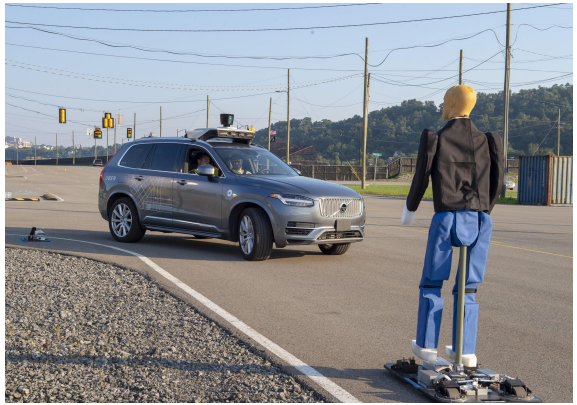
Variation Details

Variation Parameter	Min	Max	Step	Units
AV Initial Speed	35	35	none	mph
Trigger distance from actor **X**	30	50	4	m
Actor front distance from AV lane **Y**	3	8	1	m
Actor initial acceleration	2	2	none	m/s2
Actor max speed	6	12	1	mph
Actor time at max speed	2	2	none	s
Actor deceleration rate to stop	2	4	1	m/s2



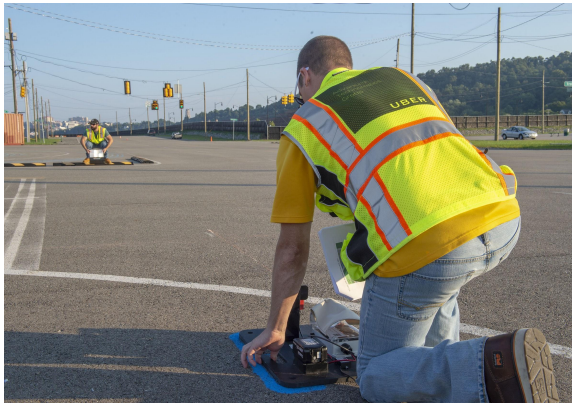
Uber

Ensuring Repeatability



GPS-Guided Actors

Robotic pedestrians and bikes provide safe, repeatable actors for track scenarios.



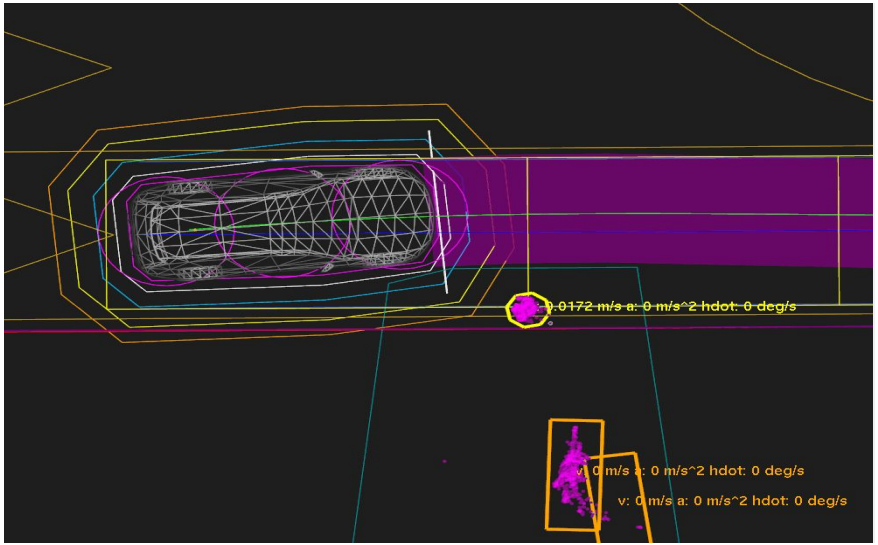
Laser Triggers

Laser break-beam triggers allow actors to move with split second timing.

Vehicles

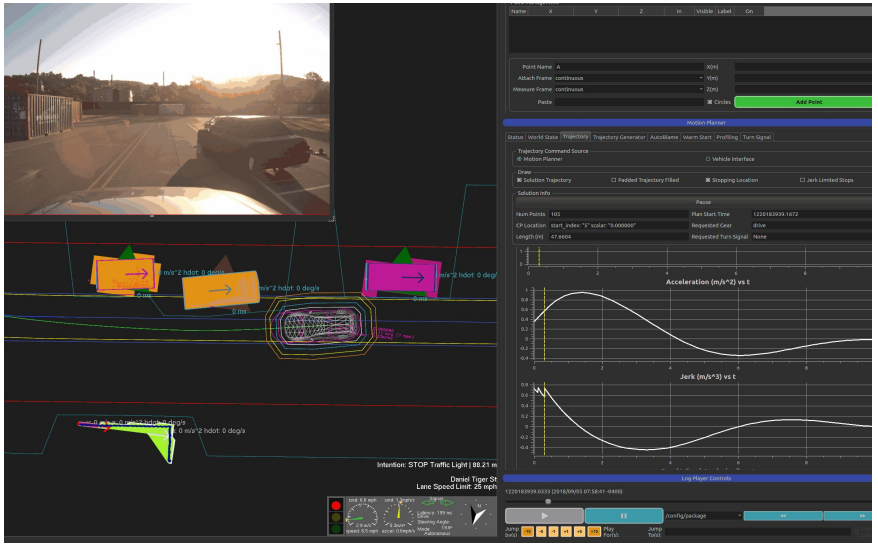
The large fleet of actor vehicles can replicate nearly anything found on public roads.

Scoring Consistently



Safety

We evaluate each scenario test for collisions and buffer violations to make sure the SDV behaves safely around all other road users.



Behavior

We also evaluate acceleration and jerk profiles to curtail suboptimal performance even without safety violations.

Uber