Trustworthy Decision-Making for Automated Driving

Jorge Villagra, CSIC

VEHITS, Benidorm (Spain)

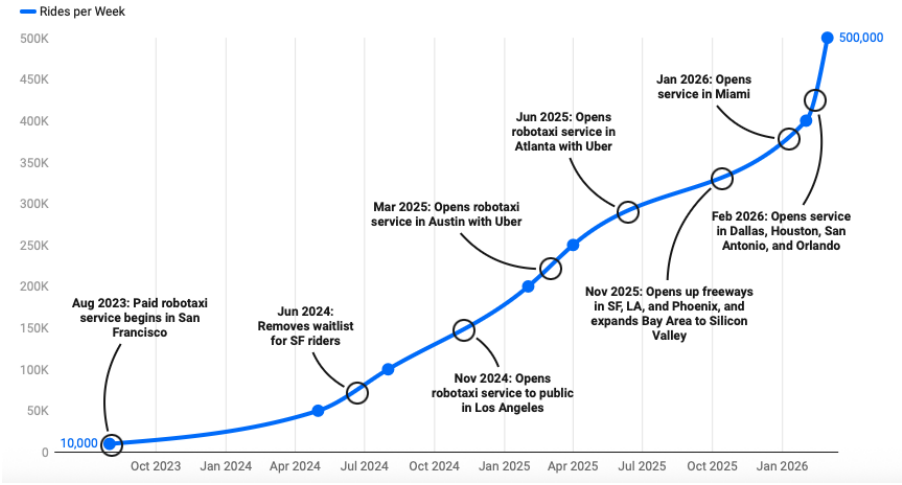
20 May, 2026

Outline

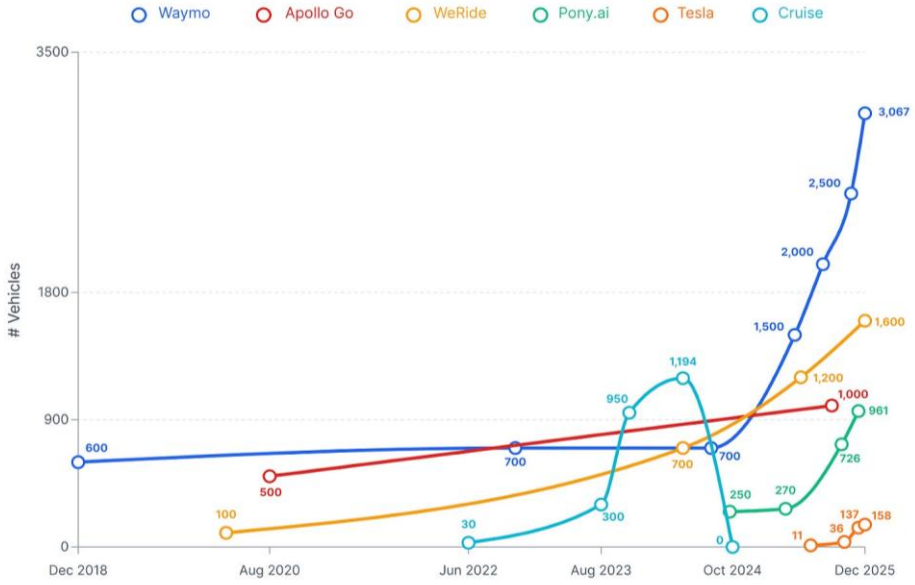
- State and trends of the technology
- Challenges ahead
- Autopia's part in tackling AD Challenges

State of the technology

Are robotaxis ready?

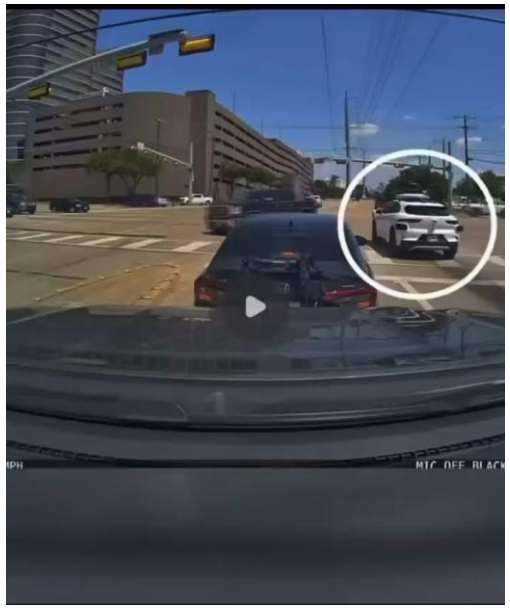


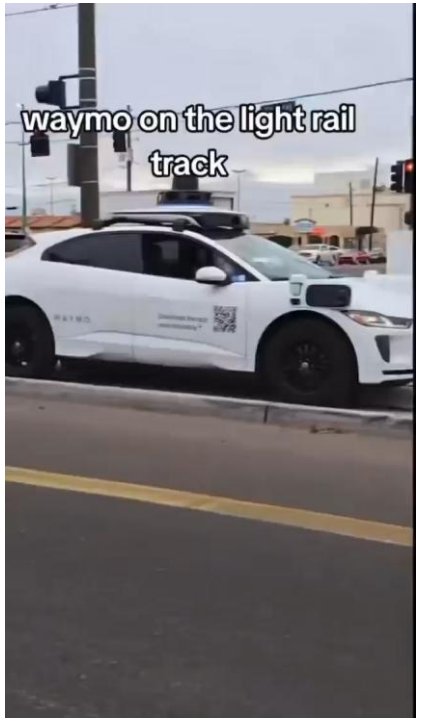
Source: TechCrunch, “Waymo’s Skyrocketing Ridership in One Chart”, 2026.

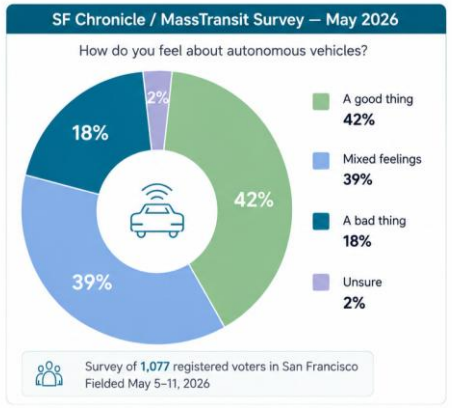


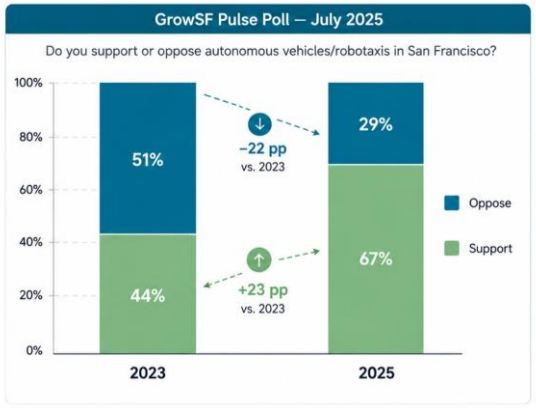
Source: Boston Consulting Group , “Here, at Last: The Evolution of the Robotaxi”, 2026

Are robotaxis ready?









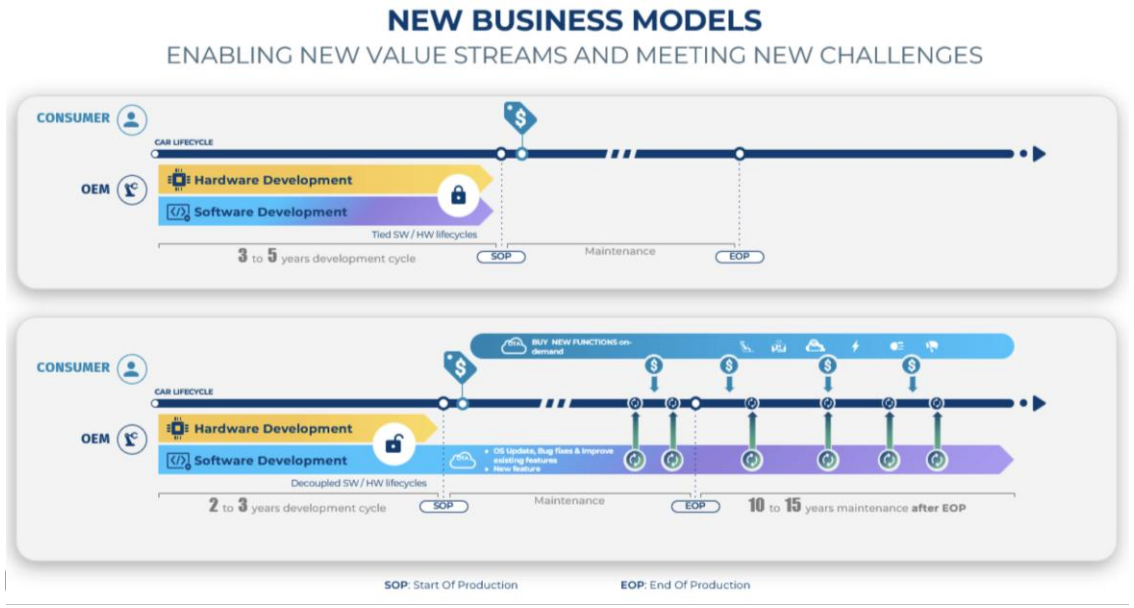
Sources: SF Chronicle / MassTransit (2026), GrowSF Pulse Poll (2025).

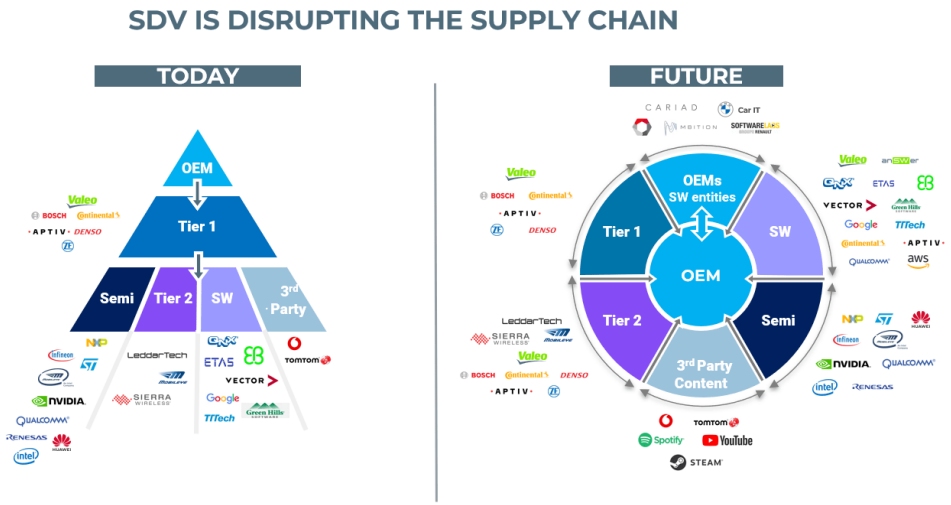
Software Defined Vehicles

- Vehicle where core functions are governed by **software**, not fixed hardware
- Enables **continuous evolution** through remote updates & upgrades
- Decouples applications from hardware, making vehicles more **flexible and adaptable**
- Acts as a **digital mobility platform**, integrating cloud and services

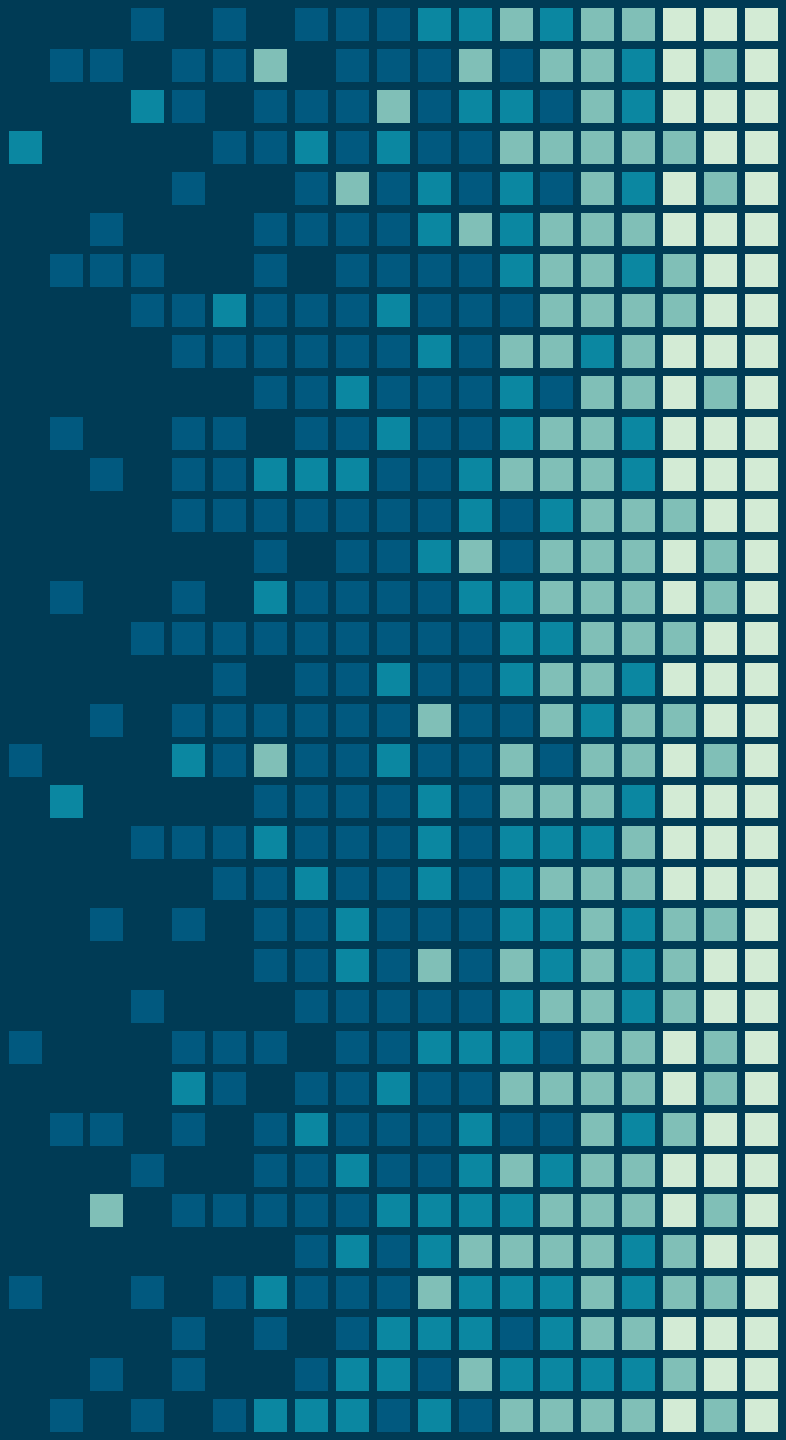
👇

- Vehicle software projected to **grow** from 100M → 1B lines of code this decade
- Duplicated efforts, fragmented platforms, and talent shortage slow down **innovation**
- Non OEMs & Big Tech entering the market with software-first approaches, creating **risks of vendor lock-in**

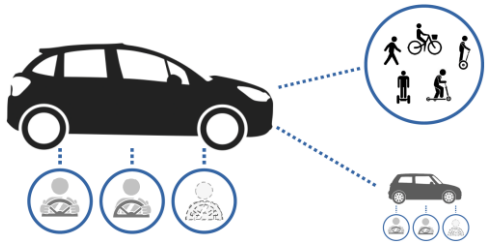




Challenges ahead



Current challenges for a massive deployement of AV

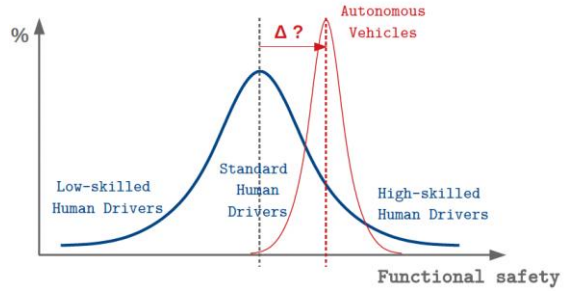


User acceptance

Human-in-the-loop decision-making
AV as threat for (work, privacy, efficiency, urban resilience…)

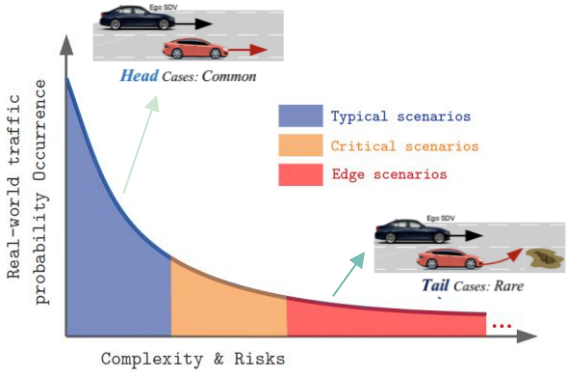
Trustworthy AI

Transparency, robustness and accountability



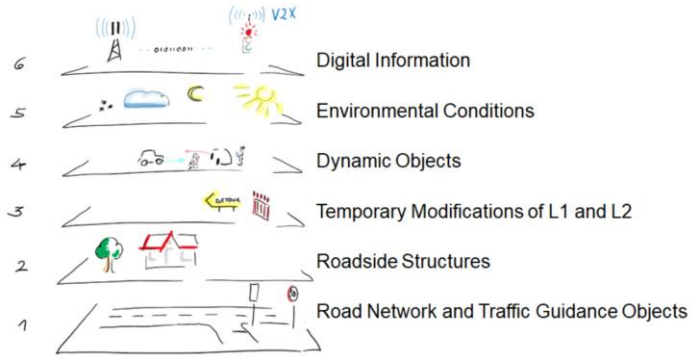
AI Generalization

Edge cases and the long-tail problem



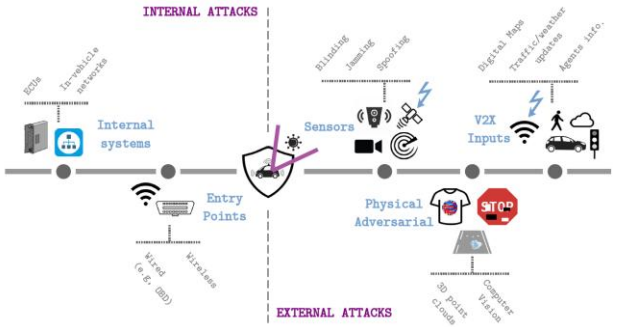
Scalability

Dynamic maps with geometry, topology and semantics on-the-fly



Assurance Co-Design

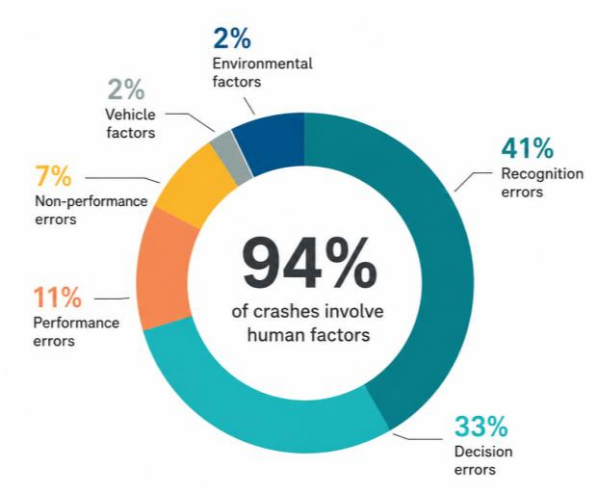
Safety-security interaction in the cloud-Edge-Vehicle continuum



Complexity management: AI vs human

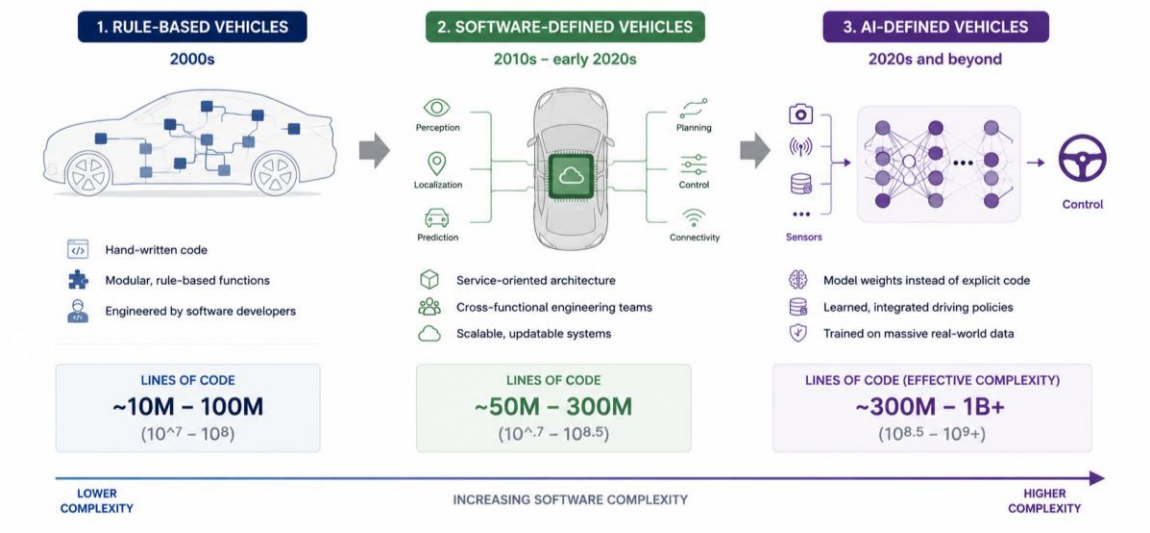
Place de l'Étoile, Paris (France)

Tahrir square, El Cairo (Egypt)



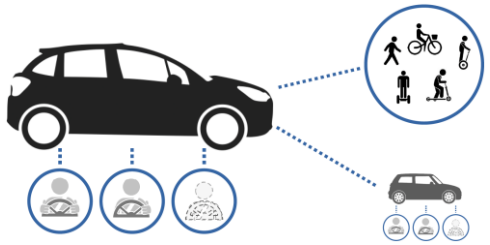
Adapted from NHTSA National Motor Vehicle Crash Causation Survey (NMVCCS), 2008.

vs



Adaptation based on industry reports and public AV software architecture estimates (2023–2025).

Current challenges for a massive deployment of AV

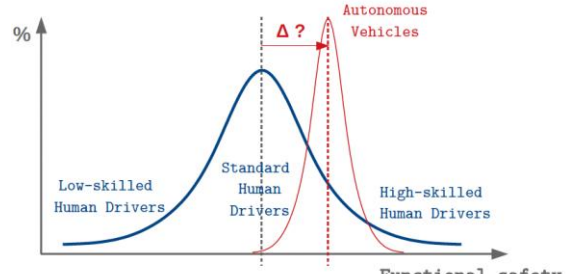


User acceptance

Human-in-the-loop decision-making
AV as threat for (work, privacy, efficiency, urban resilience…)

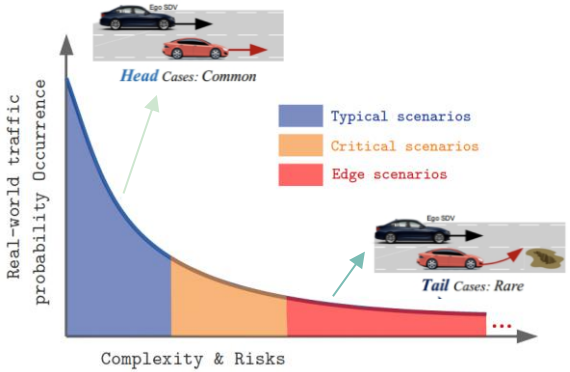
Trustworthy AI

Transparency, robustness and accountability



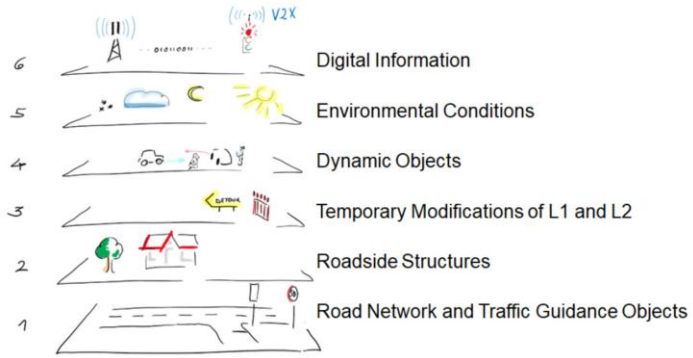
AI Generalization

Edge cases and the long-tail problem



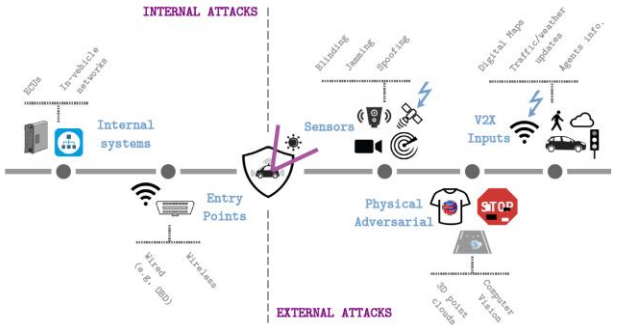
Scalability

Dynamic maps with geometry, topology and semantics on-the-fly

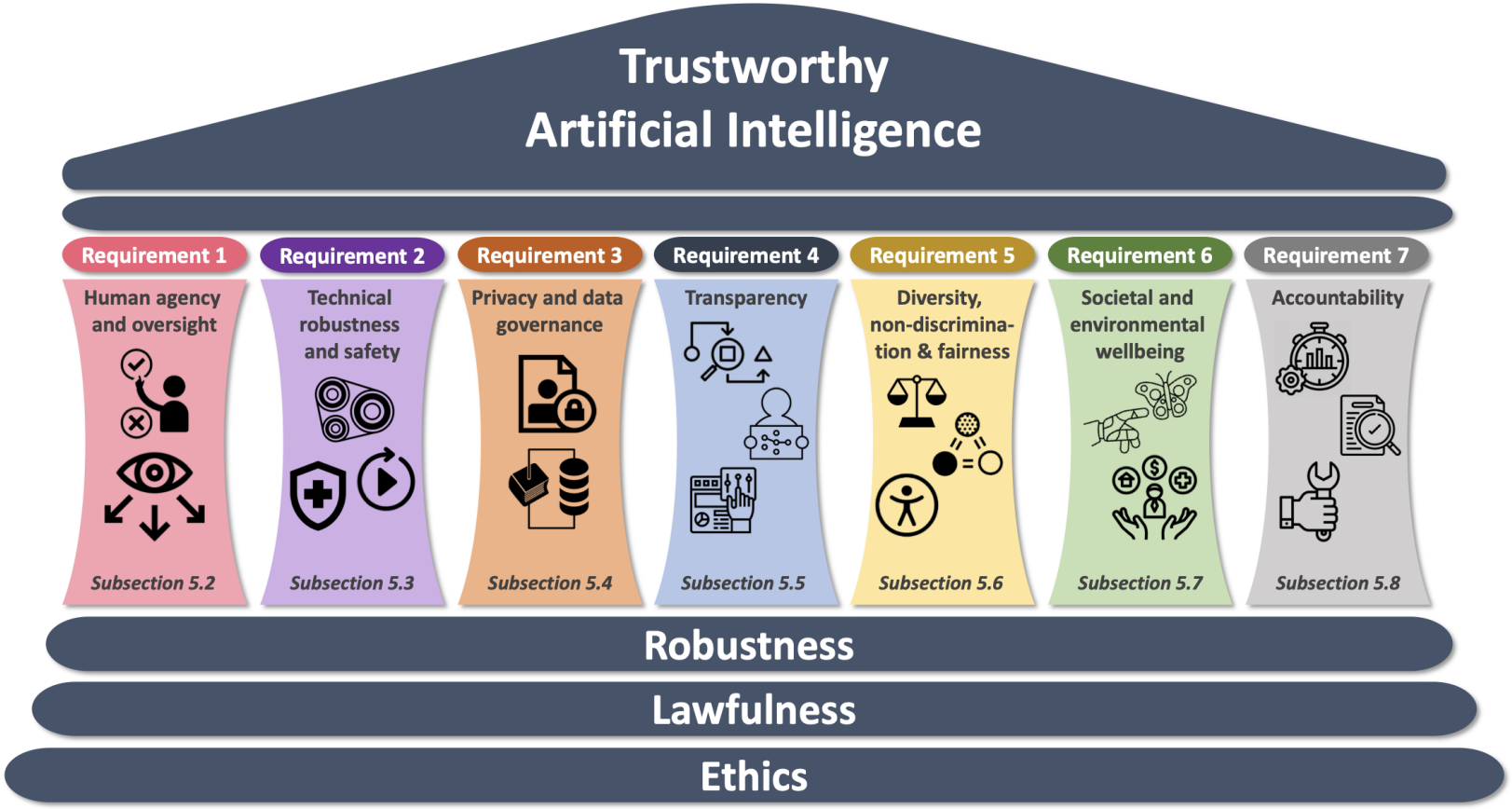


Assurance Co-Design

Safety-security interaction in the cloud-Edge-Vehicle continuum

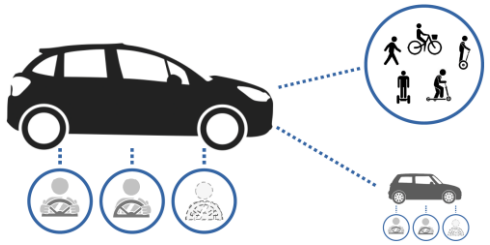


Which are the pillars for a trustworty AI?



Díaz-Rodríguez, N., Del Ser, J., Coeckelbergh, M., de Prado, M. L., Herrera-Viedma, E., & Herrera, F. Connecting the dots in trustworthy Artificial Intelligence: From AI principles, ethics, and key requirements to responsible AI systems and regulation. Information Fusion, 101896. 2023

Current challenges for a massive deployement of AV

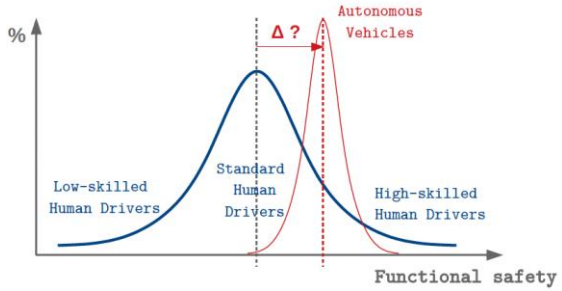


User acceptance

Human-in-the-loop decision-making
AV as threat for (work, privacy, efficiency, urban resilience…)

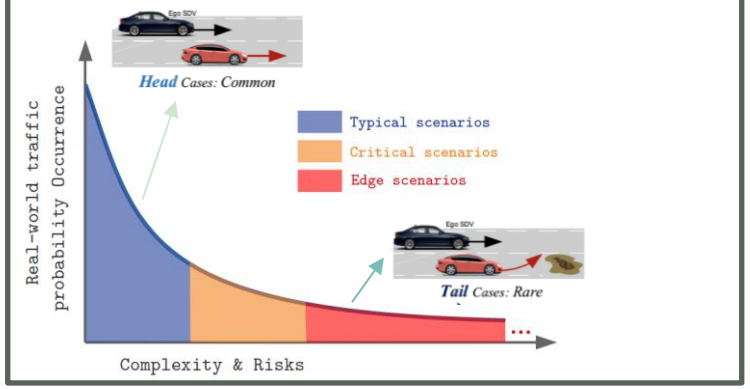
Trustworthy AI

Transparency, robustness and accountability



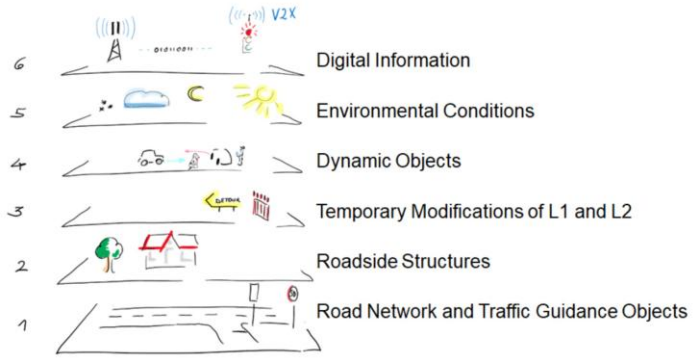
AI Generalization

Edge cases and the long-tail problem



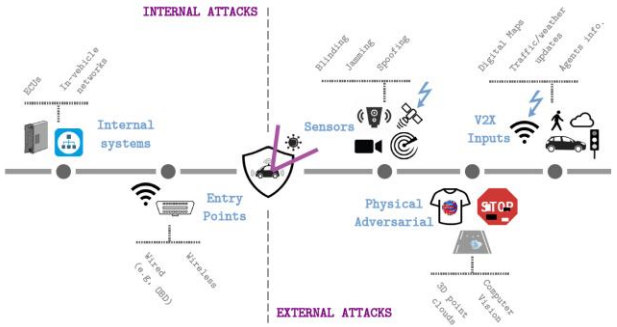
Scalability

Dynamic maps with geometry, topology and semantics on-the-fly

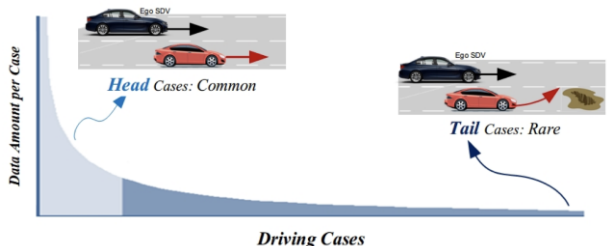


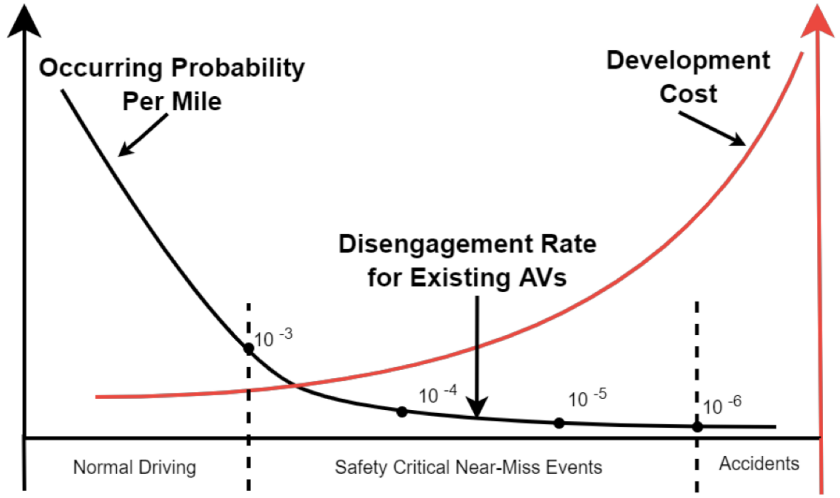
Assurance Co-Design

Safety-security interaction in the cloud-Edge-Vehicle continuum

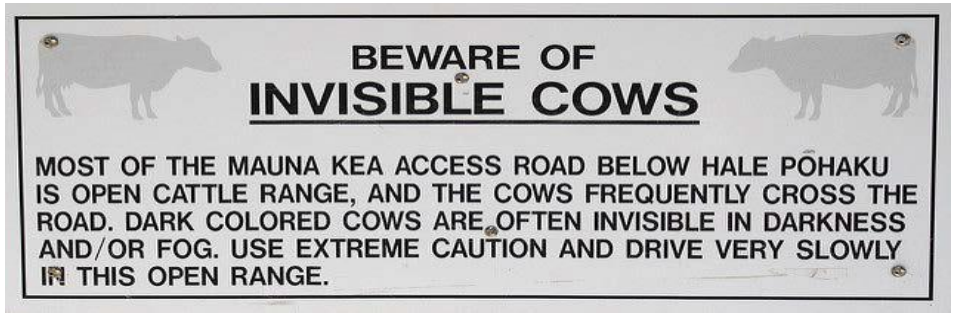


Old challenges still ahead: the long-tail distribution

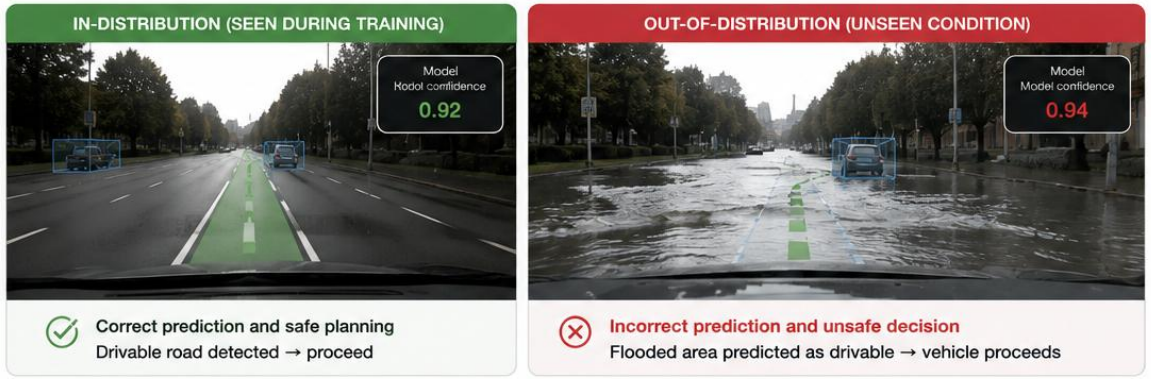




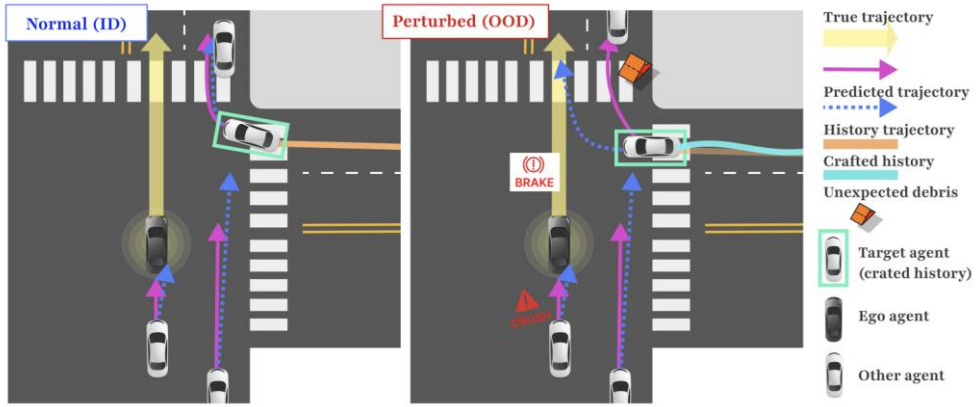
Zhou, W., Cao, Z., Deng, N., Liu, X., Jiang, K., & Yang, D. Long-Tail Prediction Uncertainty Aware Trajectory Planning for Self-driving Vehicles. arXiv e-prints, arXiv-2207, 2022



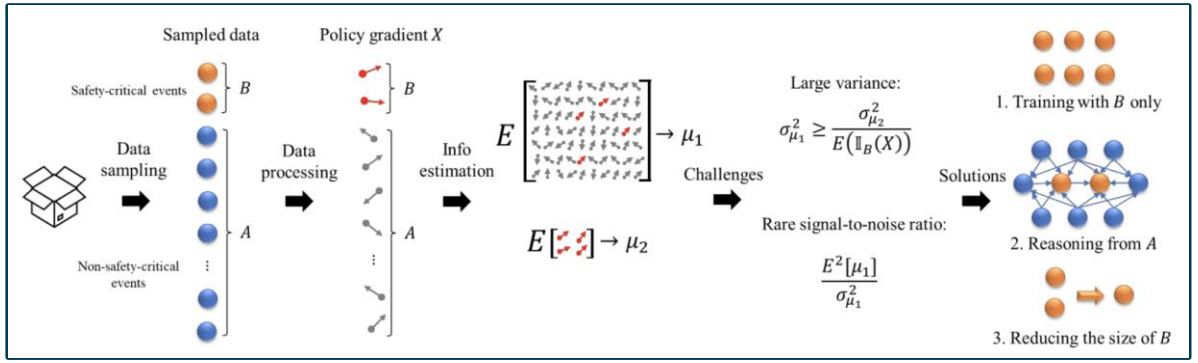
Generalizable and explainable AI: general thoughts



Adaptation inspired by Reuters coverage of the Waymo flooded-road recall (2026)

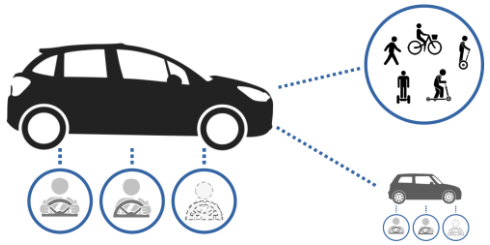


T. Guo, T. Banerjee, R. Liu, and L. Su, "Building Real-time Awareness of Out-of-distribution in Trajectory Prediction for Autonomous Vehicles," arXiv:2409.17277, 2024.



Liu, H. X., & Feng, S. Curse of rarity for autonomous vehicles, Nature communications, 2022

Current challenges for a massive deployment of AV

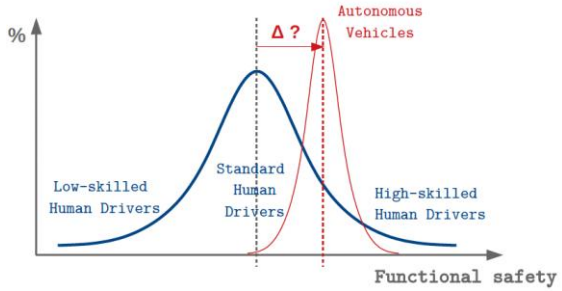


User acceptance

Human-in-the-loop decision-making
AV as threat for (work, privacy, efficiency, urban resilience…)

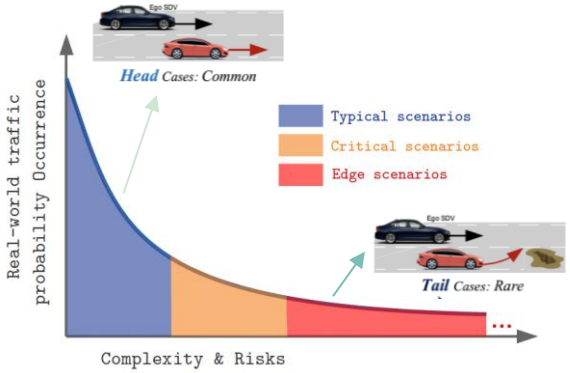
Trustworthy AI

Transparency, robustness and accountability



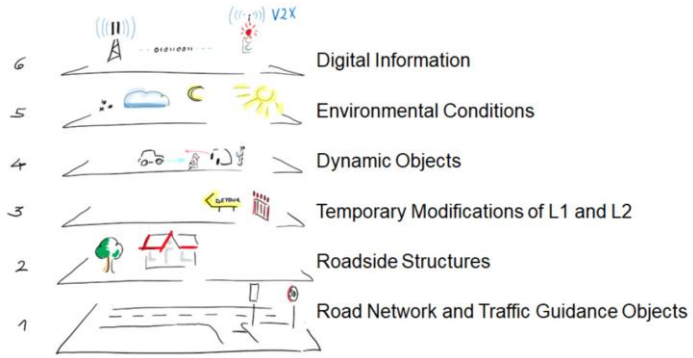
AI Generalization

Edge cases and the long-tail problem



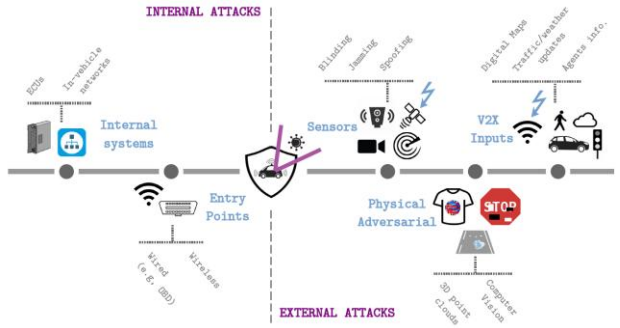
Scalability

Dynamic maps with geometry, topology and semantics on-the-fly

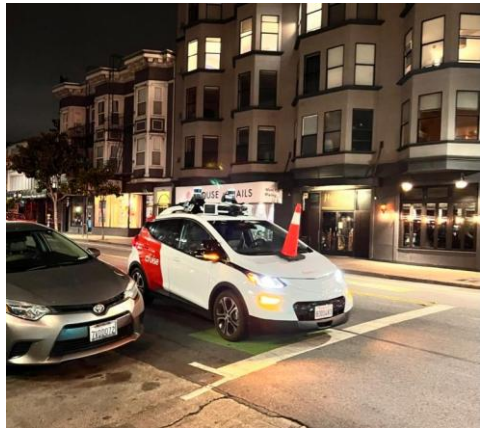


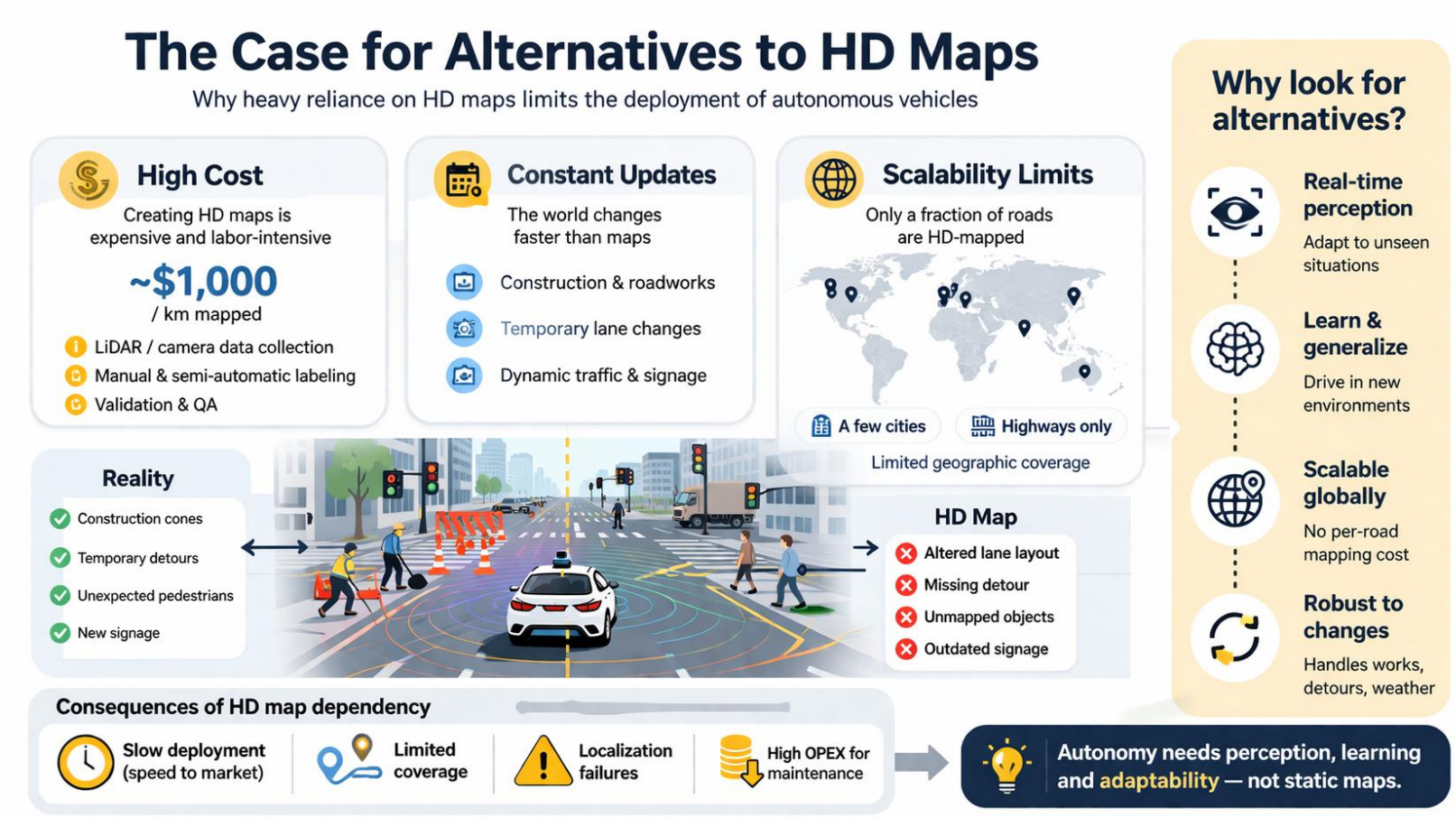
Assurance Co-Design

Safety-security interaction in the cloud-Edge-Vehicle continuum



Is it manageable the dependency of HD maps? Can standalone AVs “survive”?

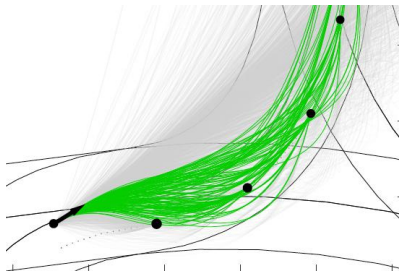




AI–based summary of HD maps challenges

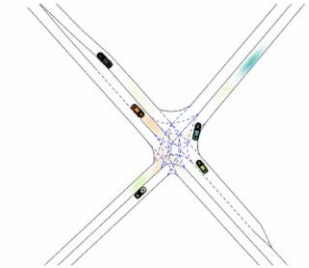
Tackling trustworthy decision-making challenges

How do we contribute to AV challenges?



Human-like planning

Safe-by-design context-aware amotion planning relying on primitives emulating human-cognition principles

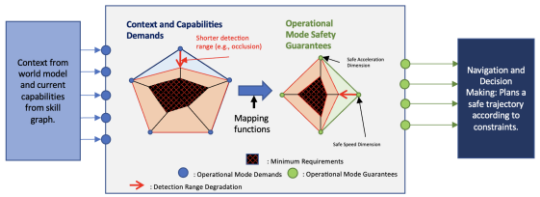


Interaction-aware (causal) prediction

Cause-effect relationships discovery between agents, enabling multi-modal footprint-based prediction

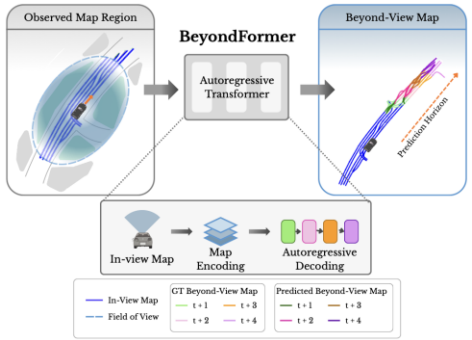
Fail-degraded navigation & control

Skills-graphs, contract-based safety, scalable/generalized risk assessment and decision-making methods



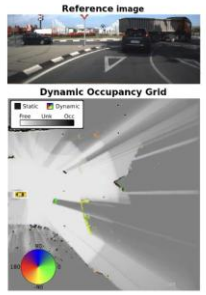
Dynamic maps

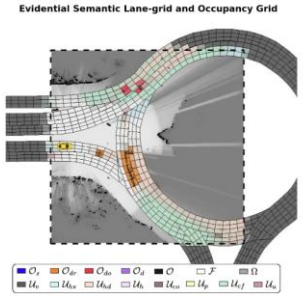
On-the-fly generation of in-view and out-of-view maps with geometry, topology, and semantics



Collective trustworthy perception

Semantic occupancy grids shared via V2X among vehicles





Dynamic maps

World modelling using probabilistic grids

LiDAR-based Perception Framework

Sensors:
- 1 Ibeo Lux: 4 layers, 100°
- 2 VLP16: 16 layers, 360°

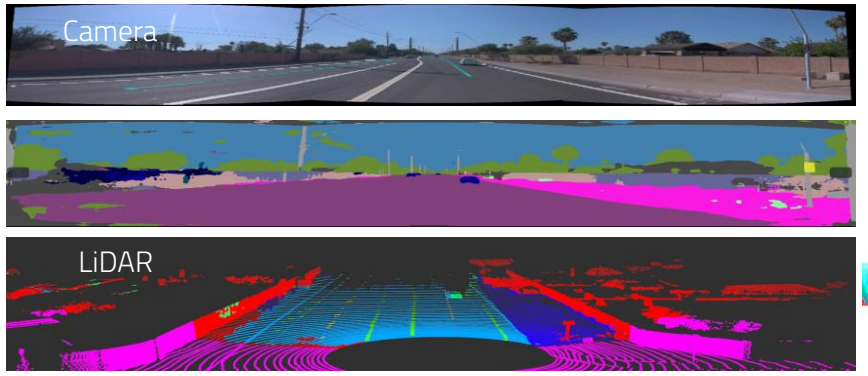
Tasks involved:
- Obstacle-ground point cloud classification
- Dynamic Occupancy Grid (DOG)
- Classified Occupancy Grid (COG)
- Road users object-level tracking

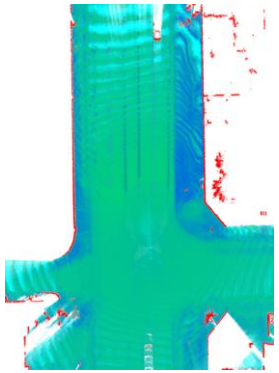
Environment semantic representation

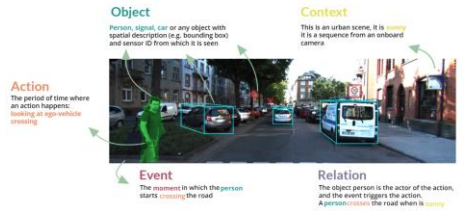
Evidential Semantic Lane-grid for Unknown Space Analysis and High-level Representation of Dynamic Occupancy Grids

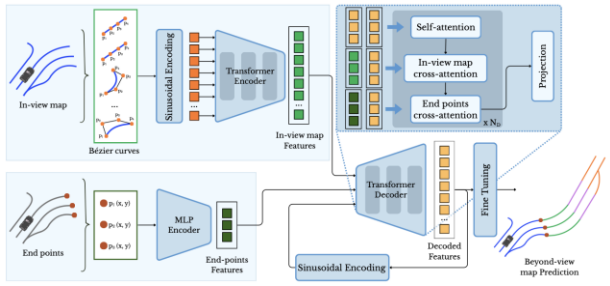
V. Jiménez-Bermejo, V. Trentin, A. Artuñedo and J. Villagra

LOGIC+: drivable area estimation using LiDAR and camera

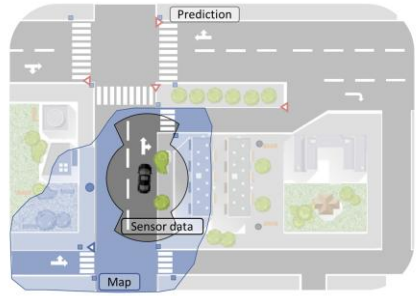




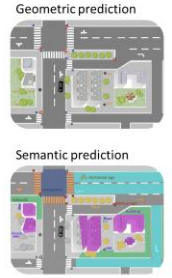


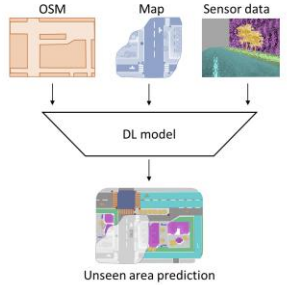


BeyondFormer



(a)





(b)

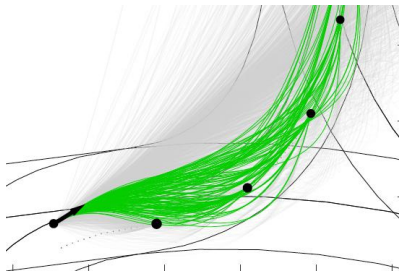
Out-of-view Map Prediction

Jiménez, V., Godoy, J., Artuñedo, A., & Villagra, J. Object-level semantic and velocity feedback for dynamic occupancy grids. IEEE Transactions on Intelligent Vehicles, 2023.

Jiménez, Trentin, Artuñedo & Villagra. Evidential Semantic Lane-Grid for Unknown Space Analysis and High-Level Representation of Dynamic Occupancy Grids. IEEE Transactions on Intelligent Vehicles, 2024

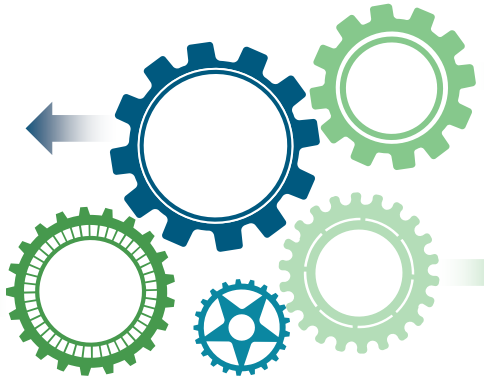
Hortelano, Jiménez-Bermejo & Villagra. LOGIC+: LiDAR-Only Geometric-Intensity Confidence Grids for Drivable Area Estimation. IEEE Open Journal of Intelligent Transportation Systems, 2026

How do we contribute to AV challenges?



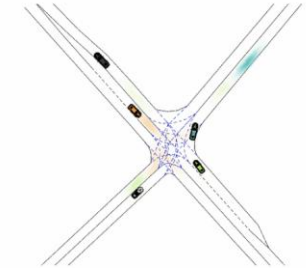
Human-like planning

Safe-by-design context-aware amotion planning relying on primitives emulating human-cognition principles



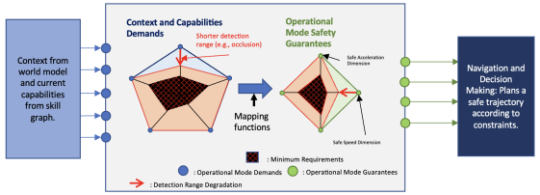
Interaction-aware (causal) prediction

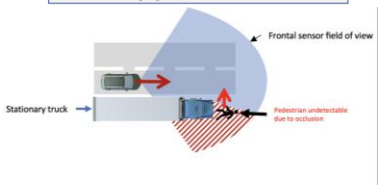
Cause-effect relationships discovery between agents, enabling multi-modal footprint-based prediction



Fail-degraded navigation & control

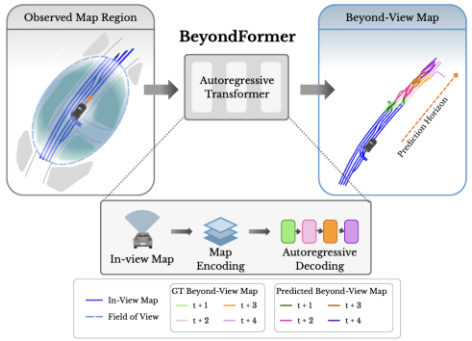
Skills-graphs, contract-based safety, scalable/generalized risk assessment and decision-making methods





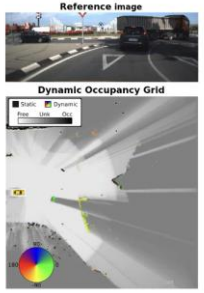
Dynamic maps

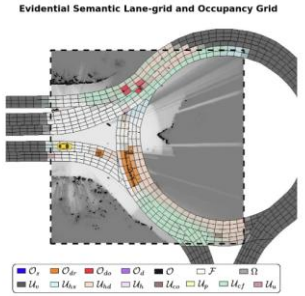
On-the-fly generation of in-view and out-of-view maps with geometry, topology, and semantics



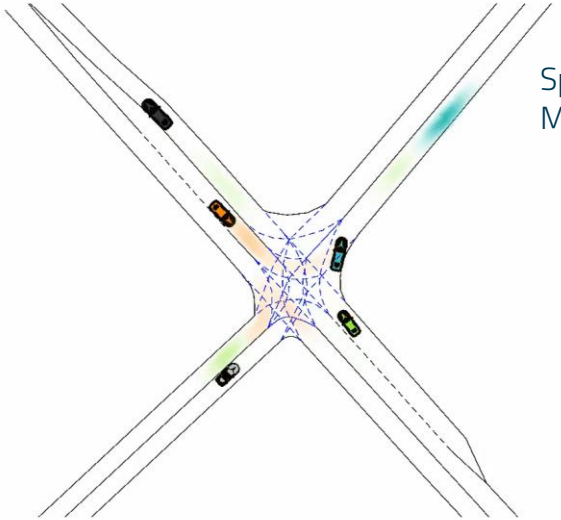
Collective trustworthy perception

Semantic occupancy grids shared via V2X among vehicles

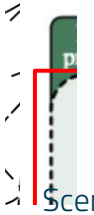




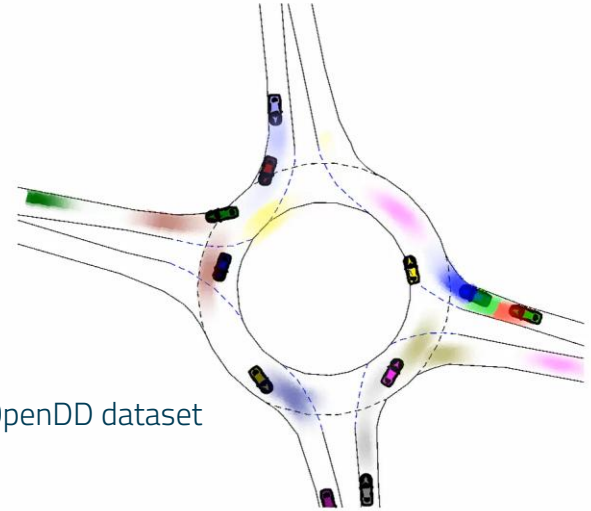
Interaction-aware motion prediction



Spatio-Temporal and
Multimodal Motion Grid

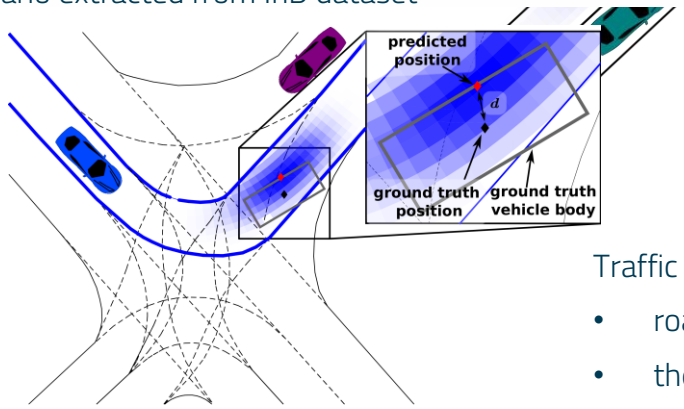


Scenario extracted from OpenDD dataset



)etection

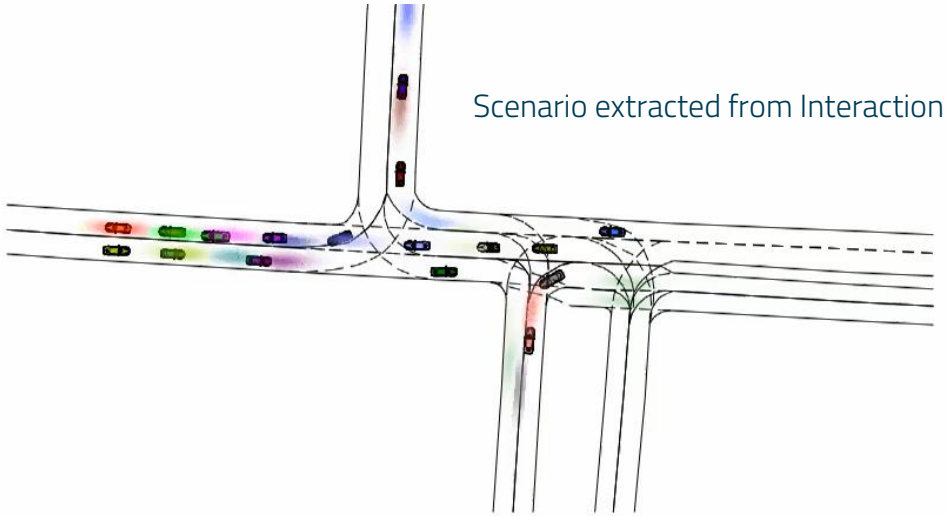
Scenario extracted from InD dataset



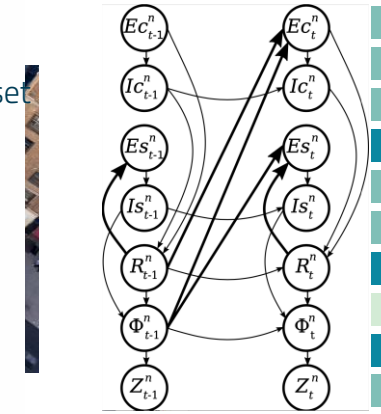
Traffic situ

- road
- there
- scene

ML-Guided Markov Chains for Motion Prediction



Scenario extracted from Interaction dataset



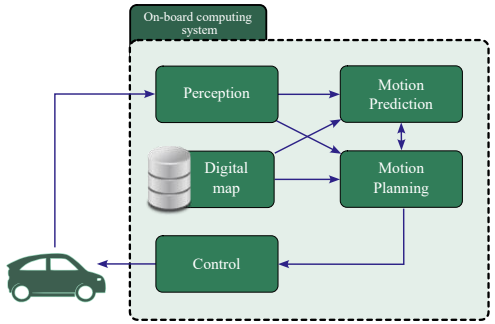
ware Intentions

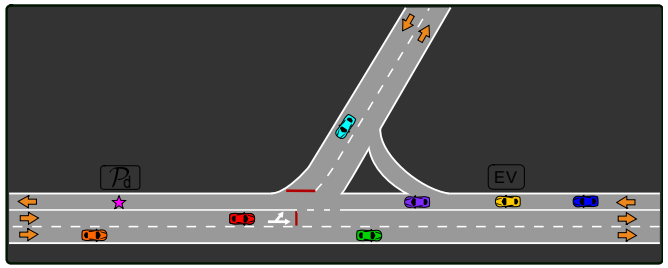
Trentin, Ma, Villagra & Al-Ars. Learning-enabled multi-modal motion prediction in urban environments, IEEE IV, 2023

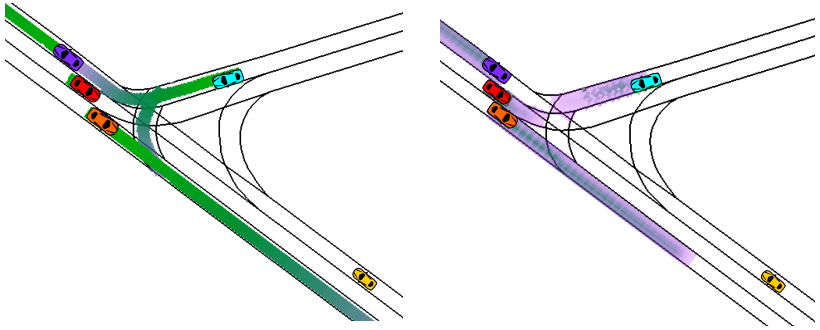
Trentin, Artuñedo, Godoy & Villagra, Multi-Modal Interaction-Aware Motion Prediction At Unsignalized Intersections. IEEE Transactions on Intelligent Vehicles, 2023

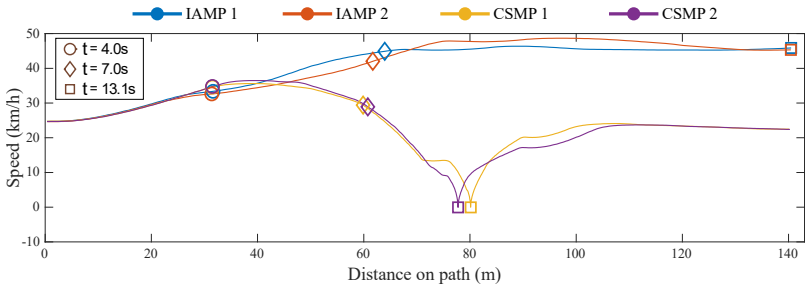
Trentin, Jiménez-Bermejo, Medina-Lee, Artuñedo & Villagra. Handling the Hidden: Occlusion-Aware Motion Prediction for Autonomous Vehicles. IEEE Open Journal of Intelligent Transportation Systems, 2026.

IAMP²: Interaction aware motion planning and prediction



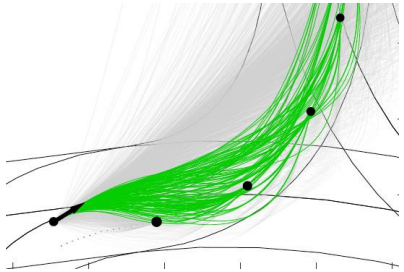






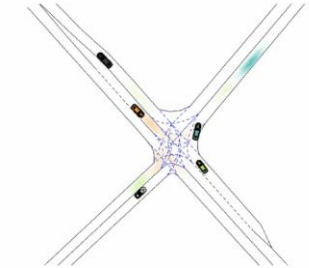
Medina-Lee, Godoy, Artuñedo & Villagra. Speed profile generation strategy for efficient merging of automated vehicles on roundabouts with realistic traffic. IEEE Transactions on Intelligent Vehicles, 2022
Medina-Lee, Trentin, Hortelano, Artuñedo, Godoy & Villagra. $IA\ (MP)^2$: Framework for Online Motion Planning Using Interaction-Aware Motion Predictions in Complex Driving Situations. IEEE Trans. on Intelligent Vehicles, 2024.

How do we contribute to AV challenges?



Human-like planning

Safe-by-design context-aware amotion planning relying on primitives emulating human-cognition principles

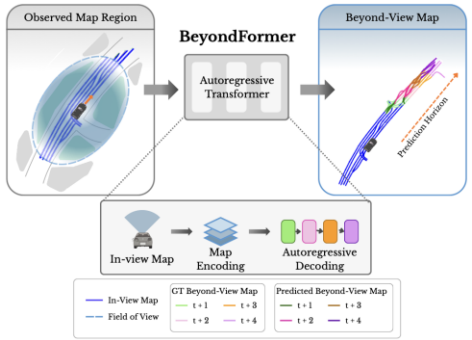


Interaction-aware (causal) prediction

Cause-effect relationships discovery between agents, enabling multi-modal footprint-based prediction

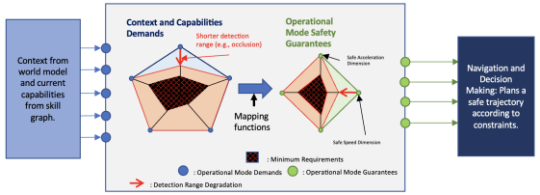
Dynamic maps

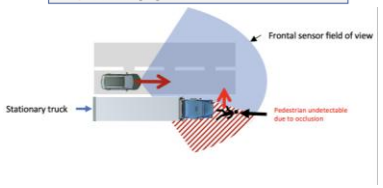
On-the-fly generation of in-view and out-of-view maps with geometry, topology, and semantics



Fail-degraded navigation & control

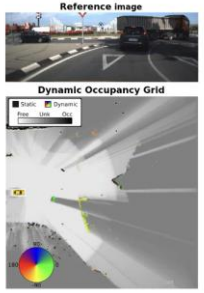
Skills-graphs, contract-based safety, scalable/generalized risk assessment and decision-making methods

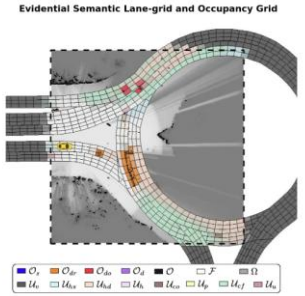




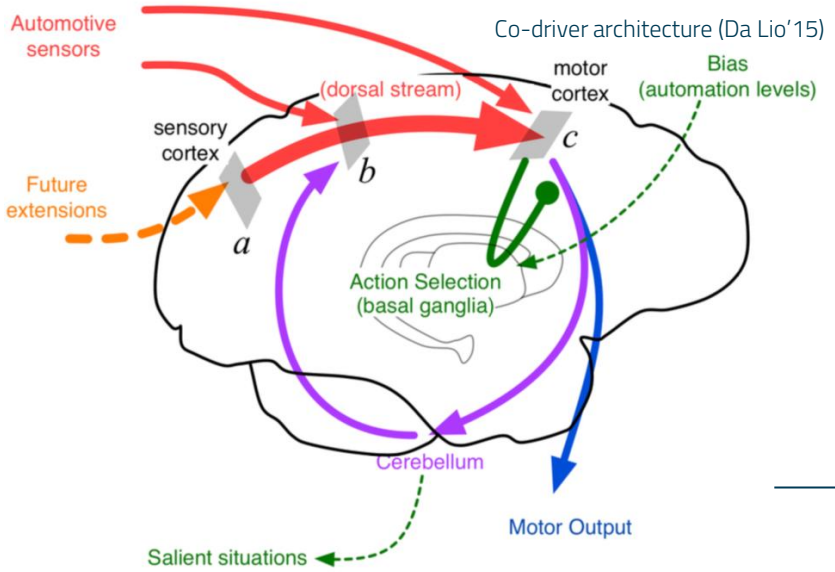
Collective trustworthy perception

Semantic occupancy grids shared via V2X among vehicles

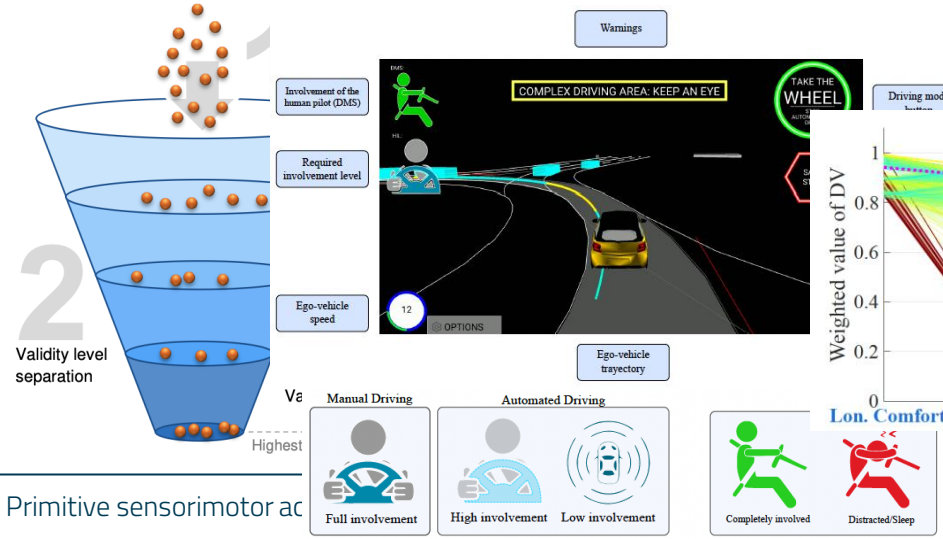




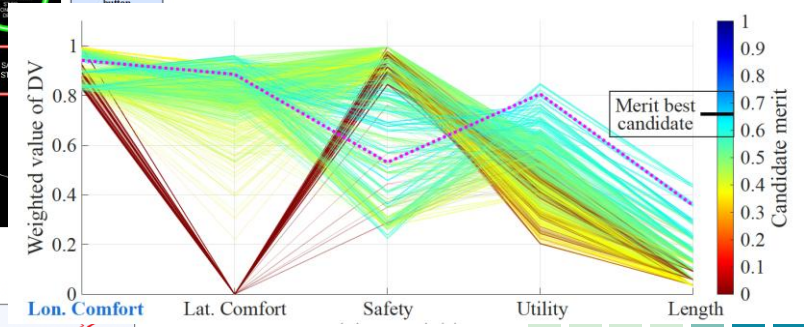
Human-in-the-loop decision-making

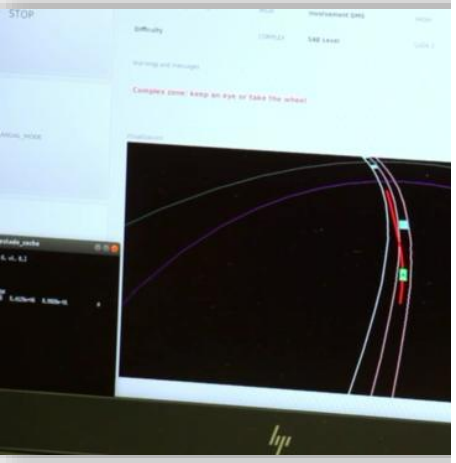


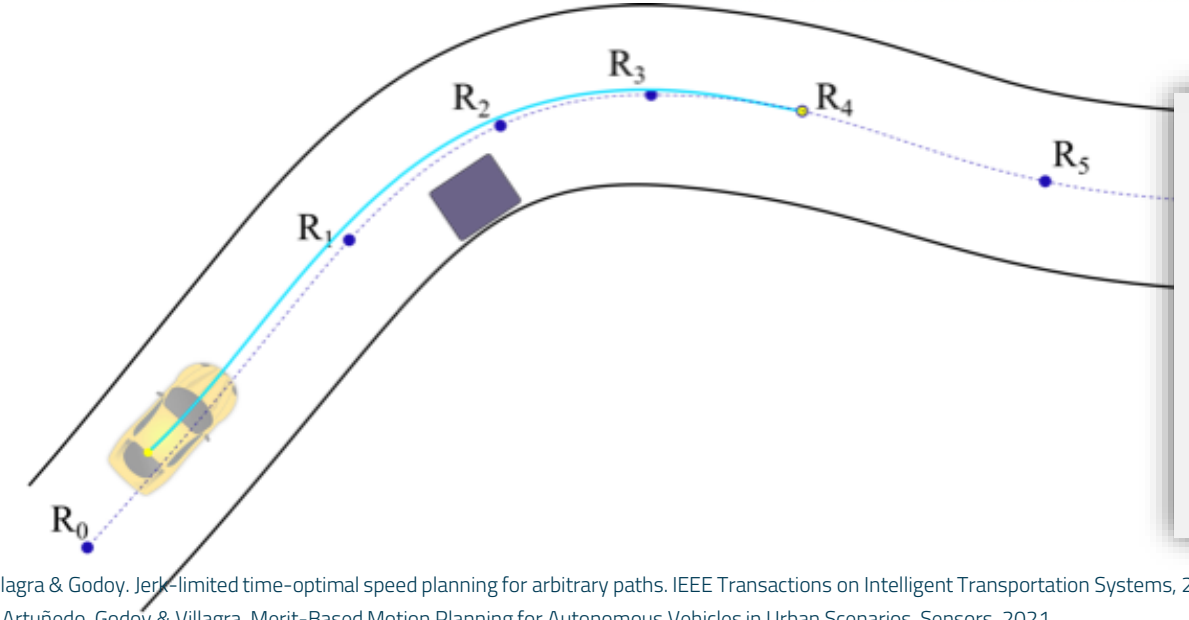
Co-driver architecture (Da Lio'15)



Primitive sensorimotor ac





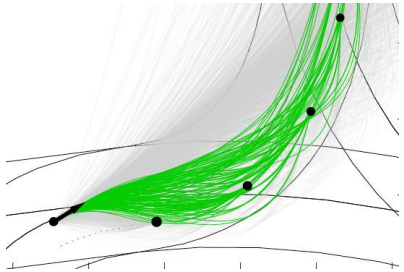


Artuñedo, Villagra & Godoy. Jerk-limited time-optimal speed planning for arbitrary paths. IEEE Transactions on Intelligent Transportation Systems, 2021

Medina-Lee, Artuñedo, Godoy & Villagra. Merit-Based Motion Planning for Autonomous Vehicles in Urban Scenarios. Sensors, 2021.

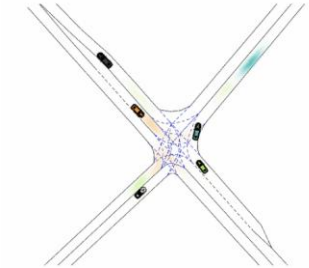
Medina-Lee, Artuñedo, Godoy, Trentin & Villagra. Self-configuring Motion Planner for Automated Vehicles Based on Human Driving Styles. IEEE Intelligent Vehicles Symposium (IV), 2024.

How do we contribute to AV challenges?



Human-like planning

Safe-by-design context-aware amotion planning relying on primitives emulating human-cognition principles

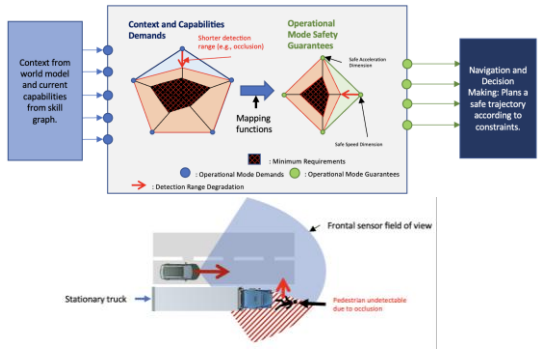


Interaction-aware (causal) prediction

Cause-effect relationships discovery between agents, enabling multi-modal footprint-based prediction

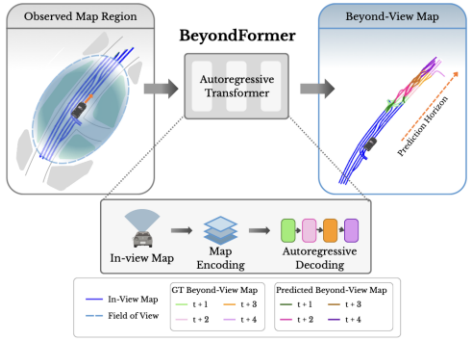
Fail-degraded navigation & control

Skills-graphs, contract-based safety, scalable/generalized risk assessment and decision-making methods



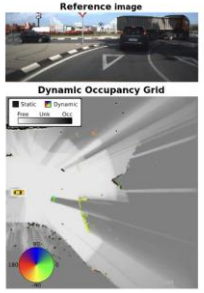
Dynamic maps

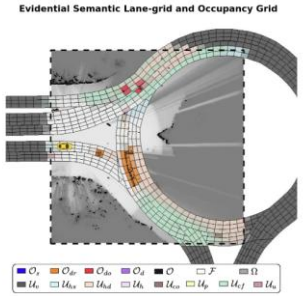
On-the-fly generation of in-view and out-of-view maps with geometry, topology, and semantics



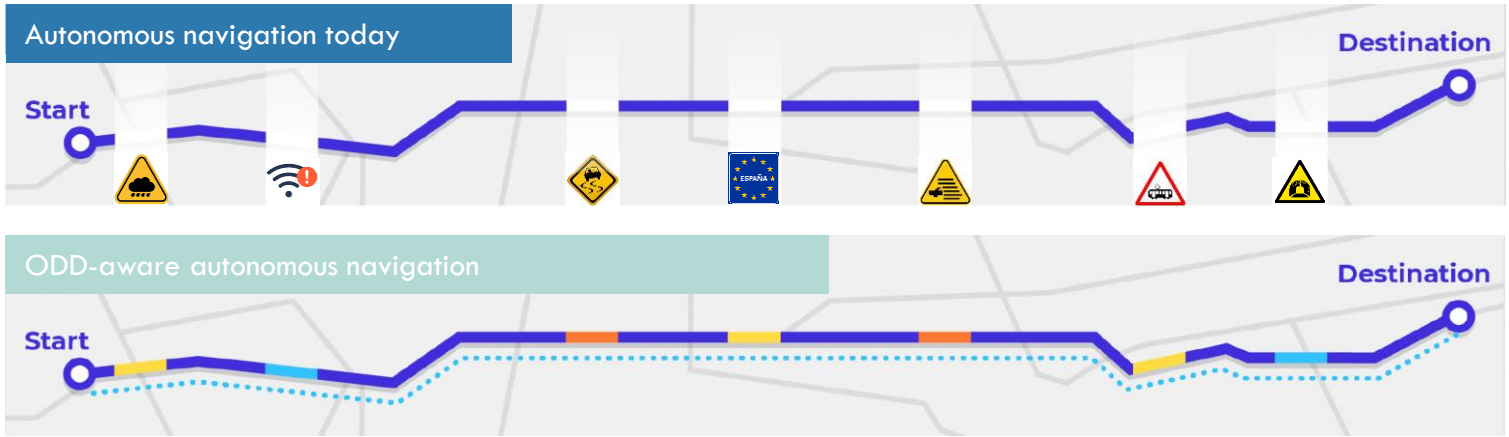
Collective trustworthy perception

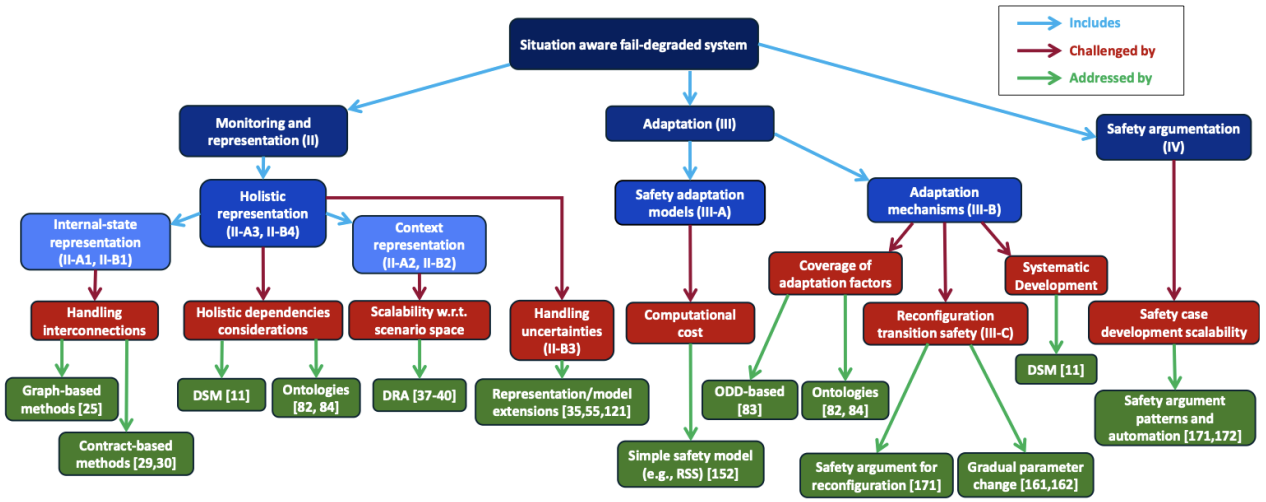
Semantic occupancy grids shared via V2X among vehicles

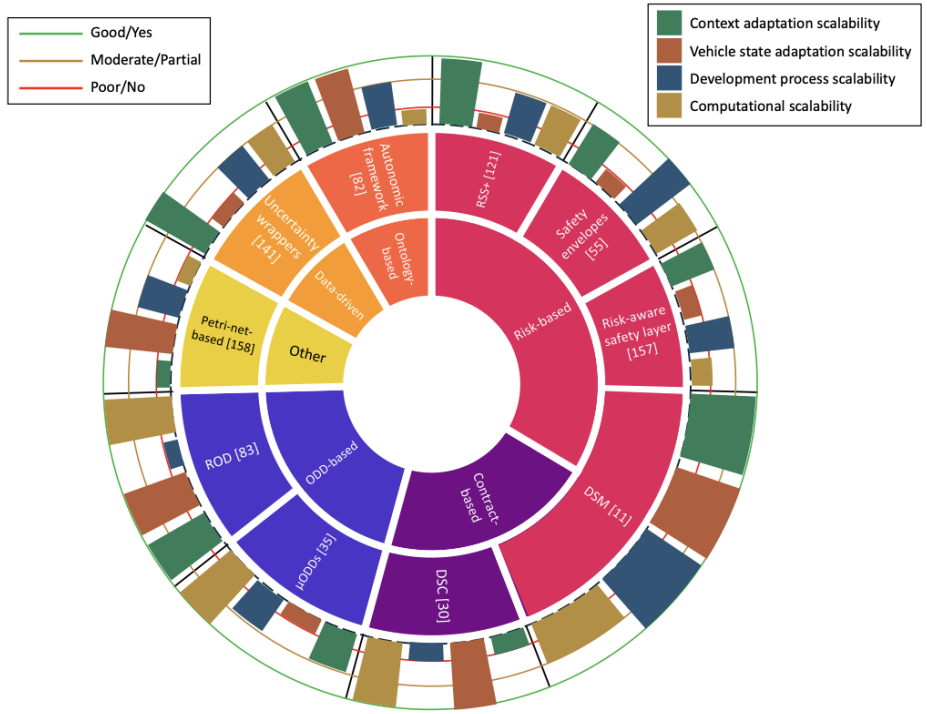




Scalable fail-degraded navigation

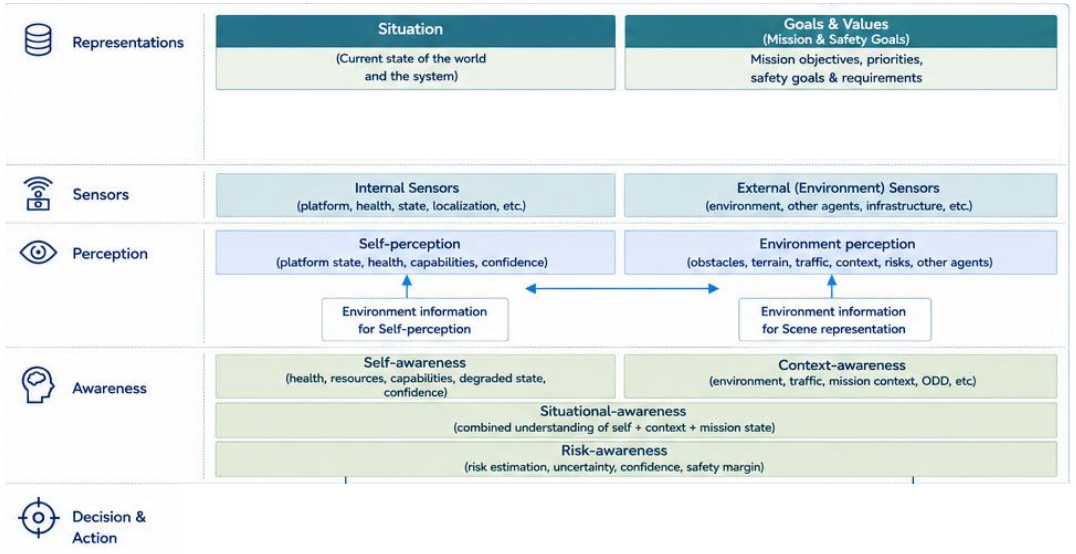




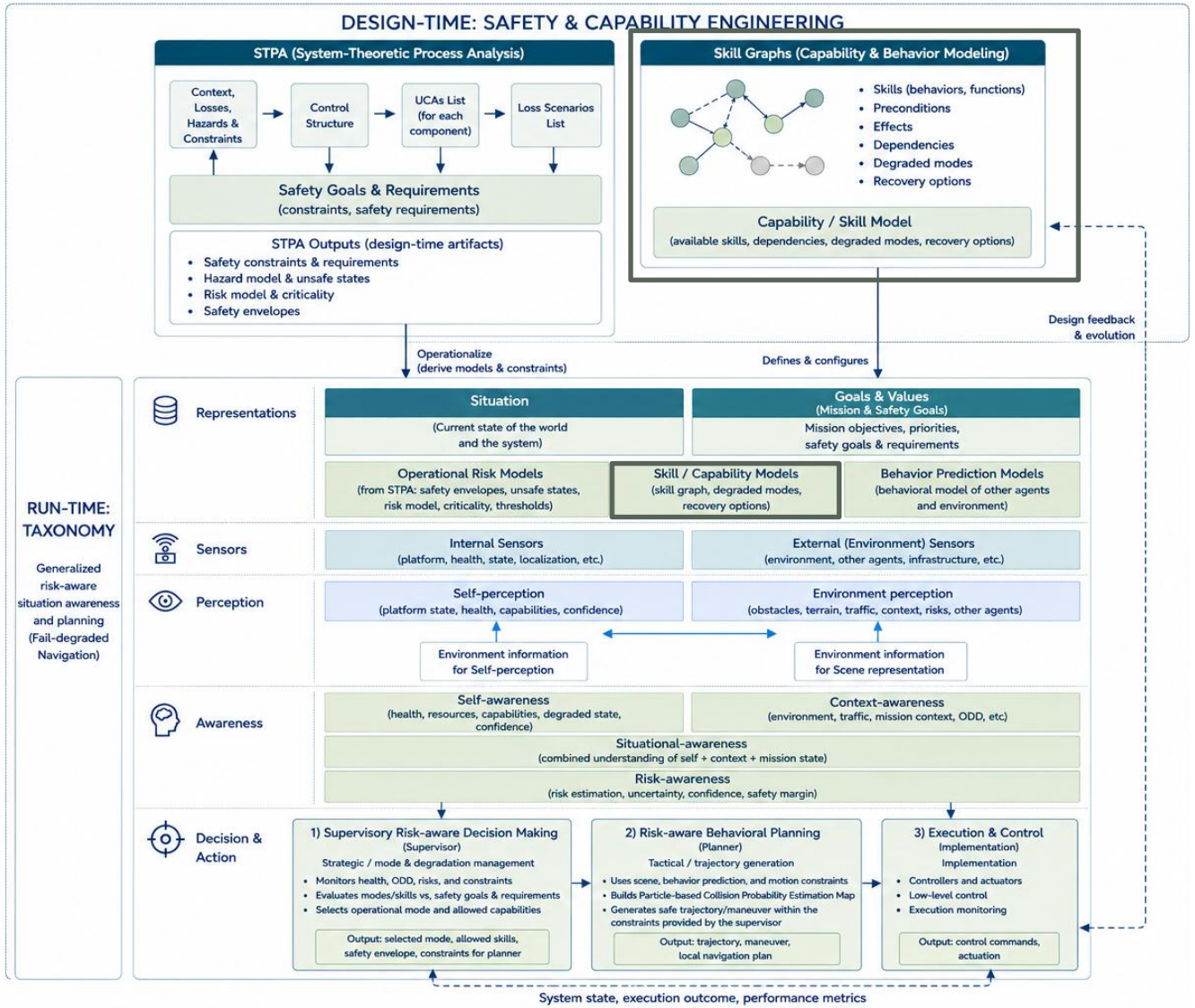


Hossam, Villagra, Navas & Milanés. Scalable Fail-Degraded Systems for Autonomous Vehicles: A Survey. IEEE Transactions on Intelligent Transportation Systems, 2025

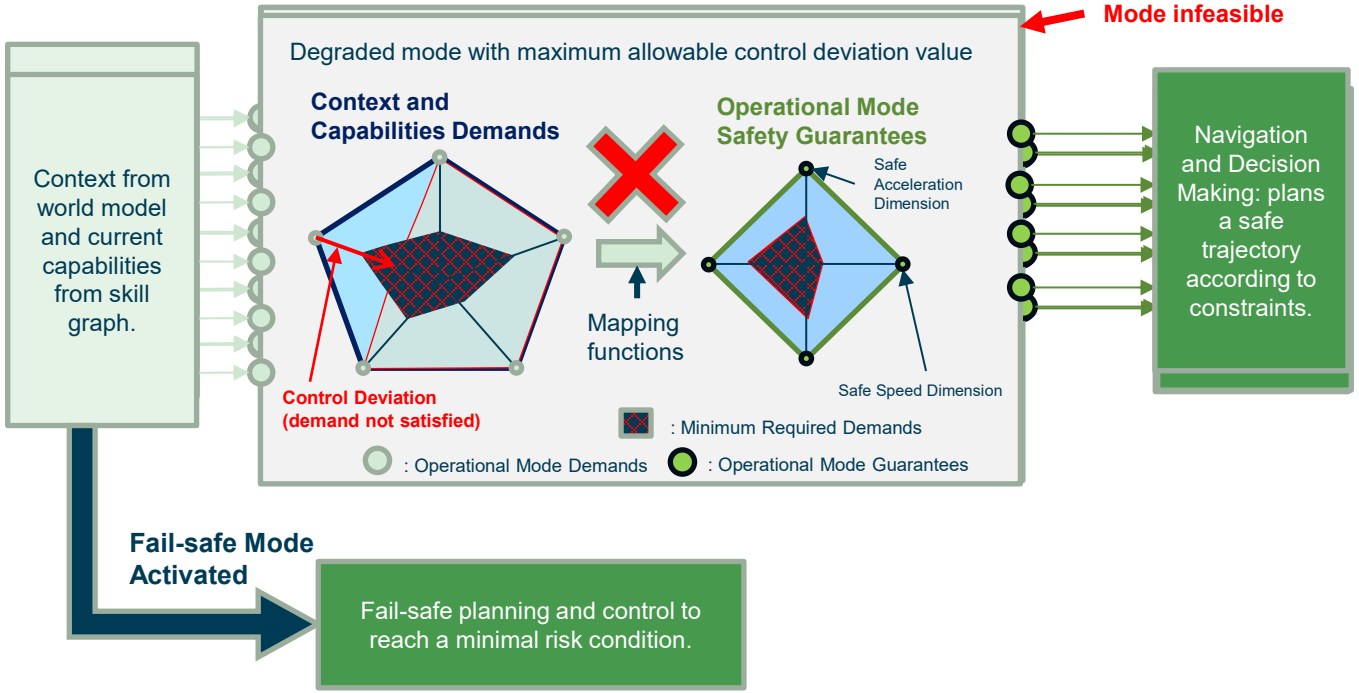
Taxonomy for monitoring and awareness

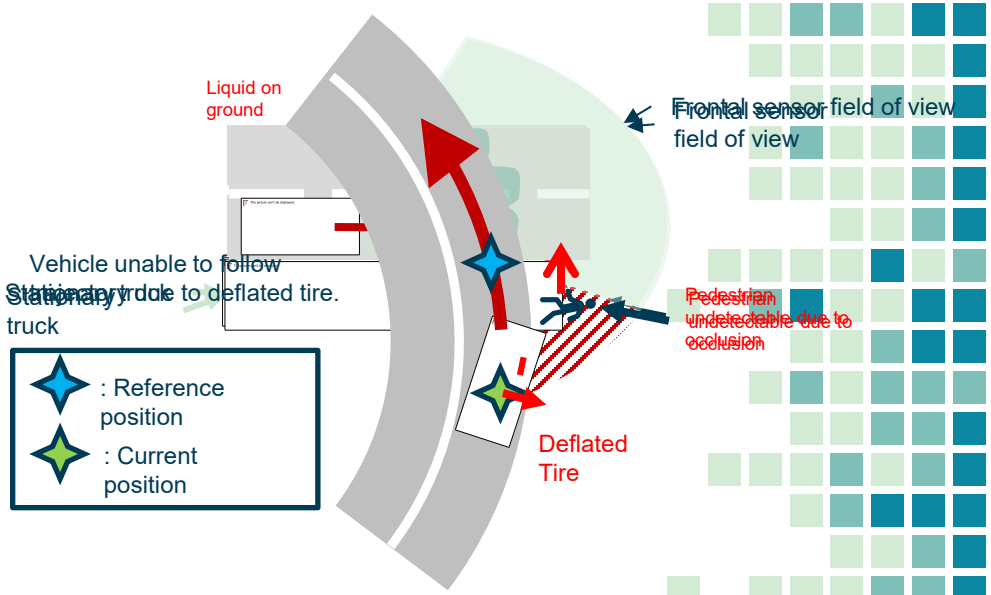


Autopia approach for monitoring, awareness and planning

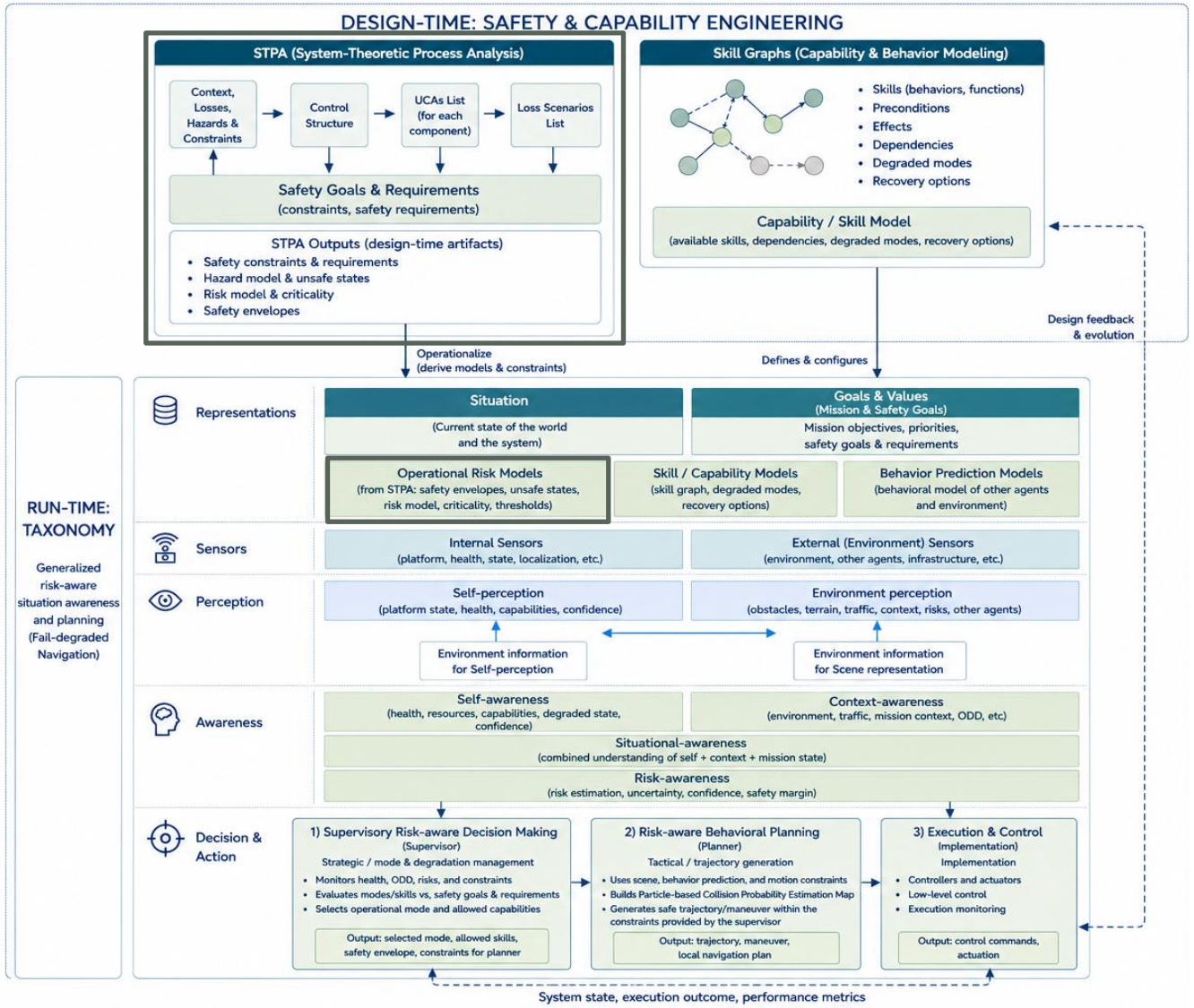


Skill-graphs for fail-degraded navigation

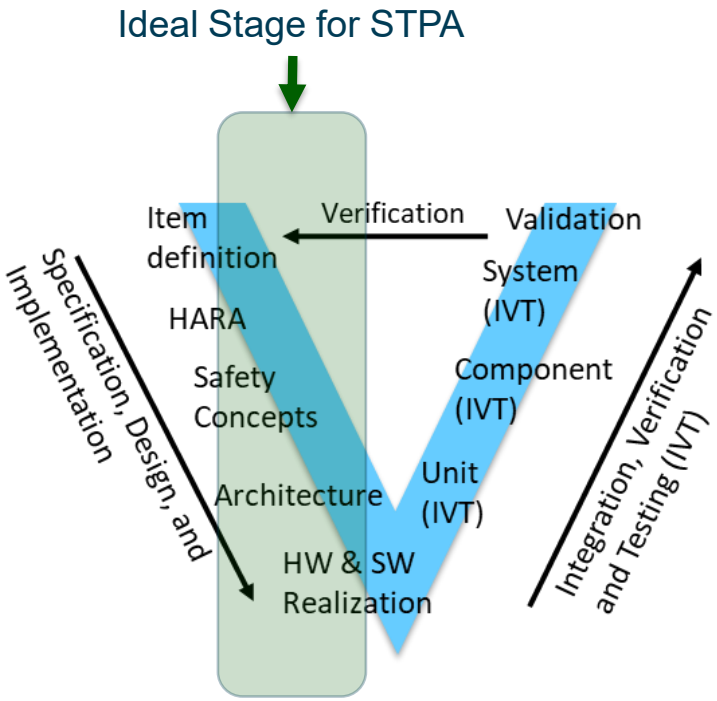


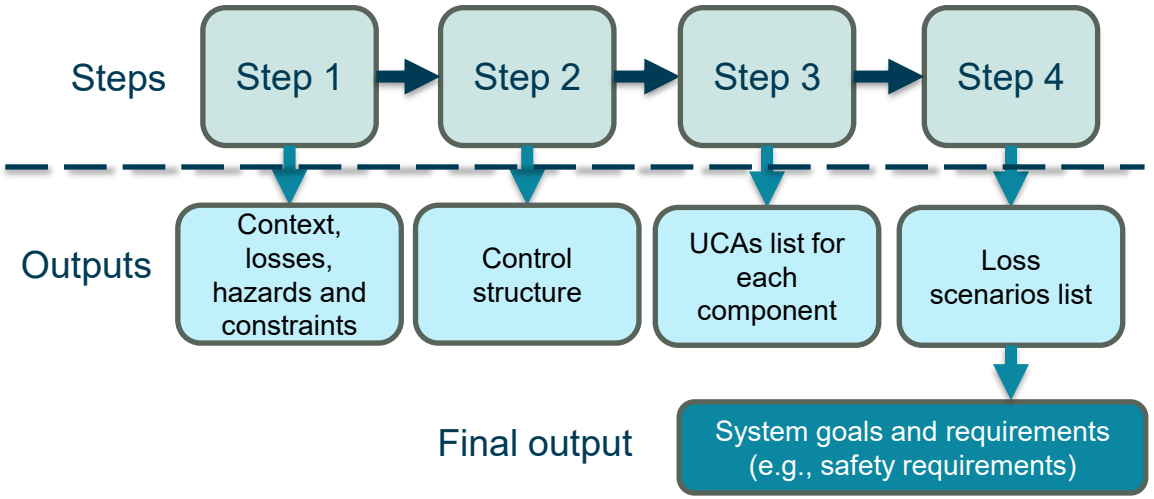


Autopia approach for monitoring, awareness and planning



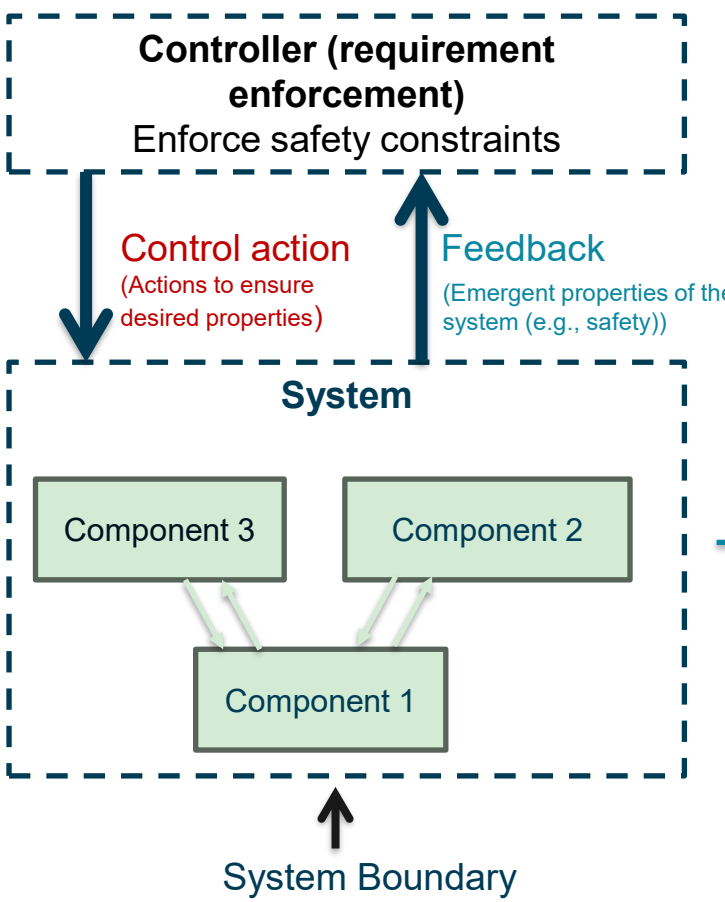
STPA: guiding principles

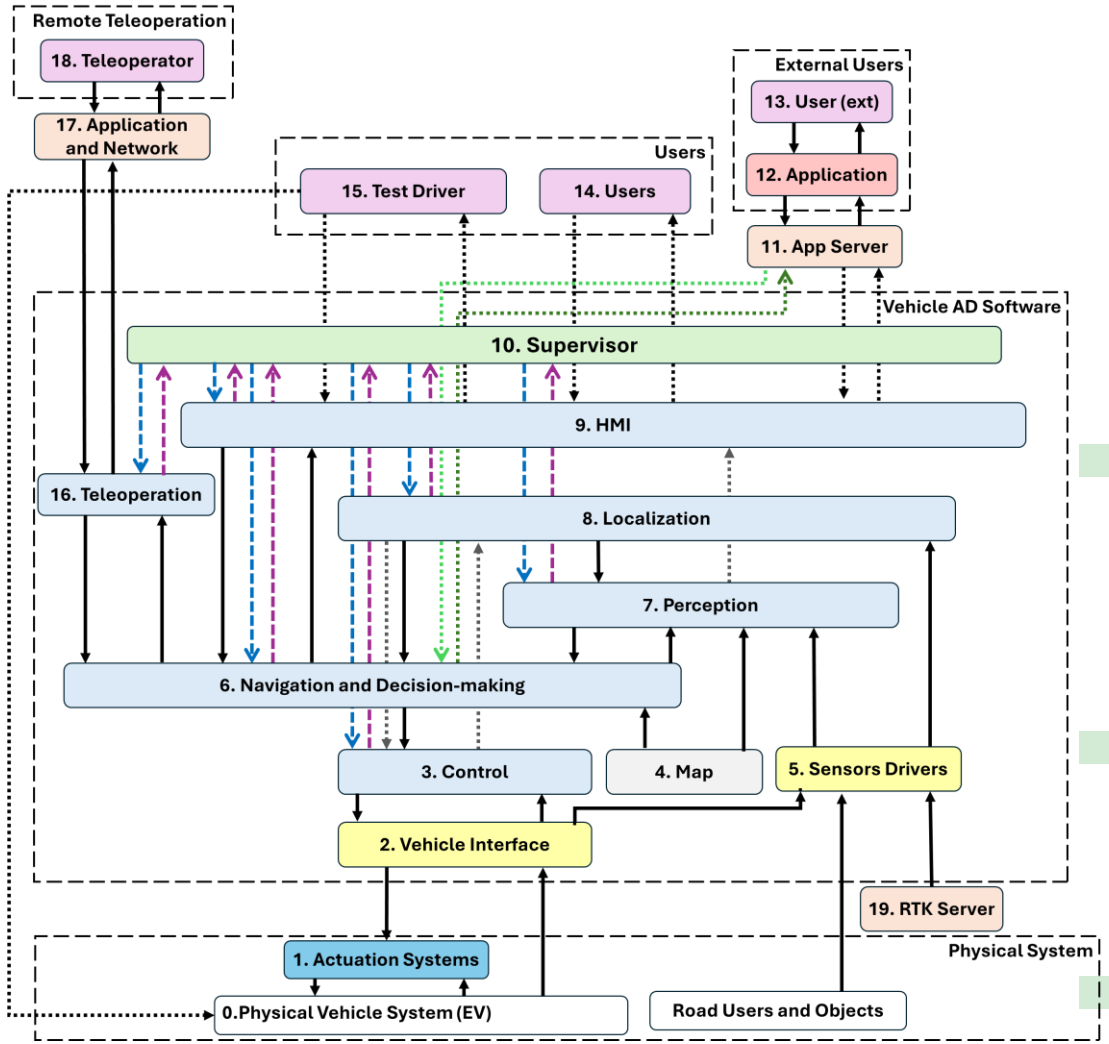




- Proactive hazard analysis method
  - Most effective during design and development stages, reducing late redesign and mitigation costs
- System-theoretic safety approach
  - Models safety as a control problem and analyzes the system as a whole to prevent losses
- Structured derivation of safety requirements:
  - Identifies hazards, unsafe control actions (UCAs), loss scenarios, and safety constraints.

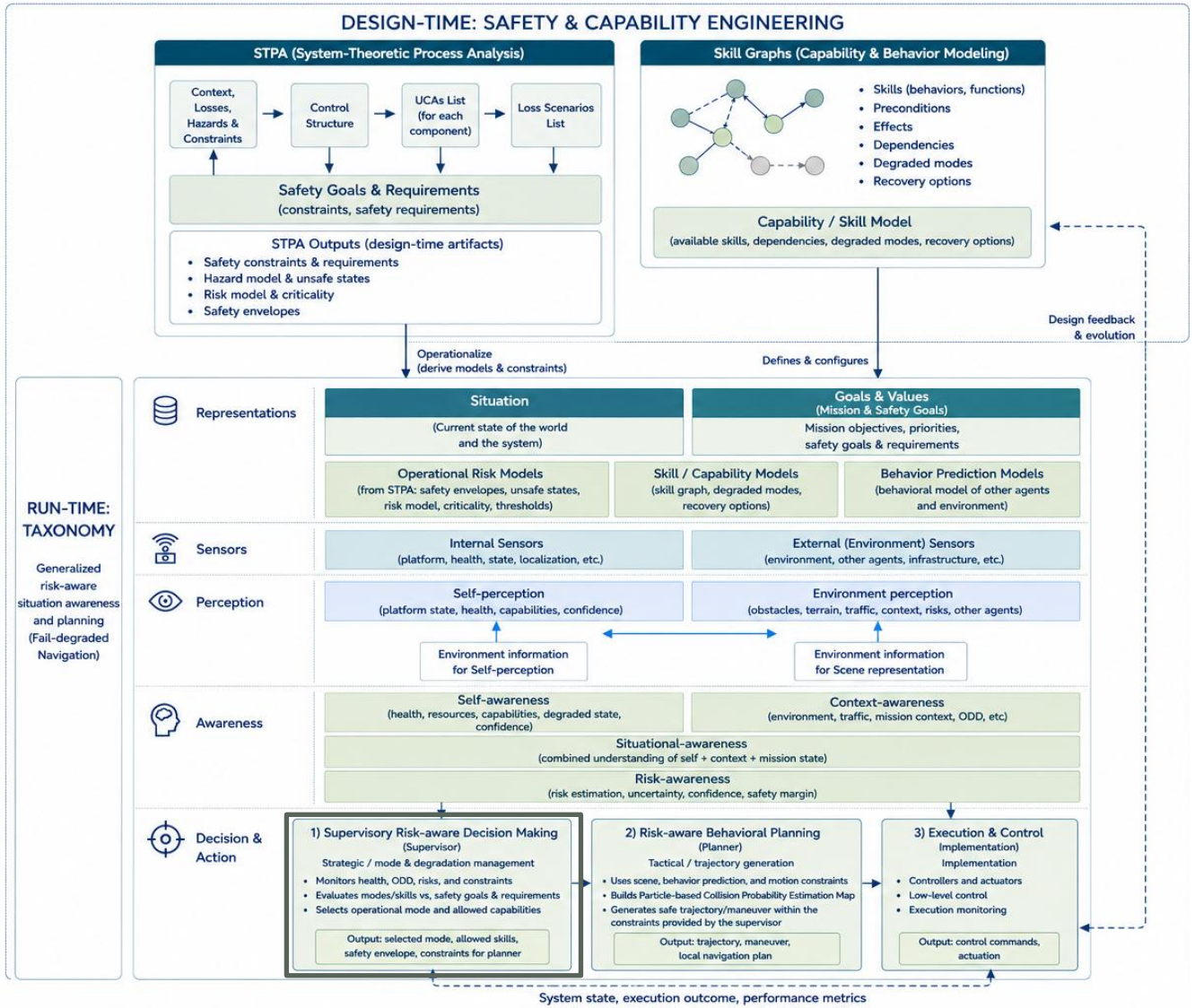
STPA: application to AUTOPIA stack





N. Leveson and J. Thomas, STPA Handbook. 2018
Berger, J. STPA Guide. VTT Technical Research Centre of Finland. VTT Research Report No. VTT-R-00848-23, 2024

Autopia approach for monitoring, awareness and planning



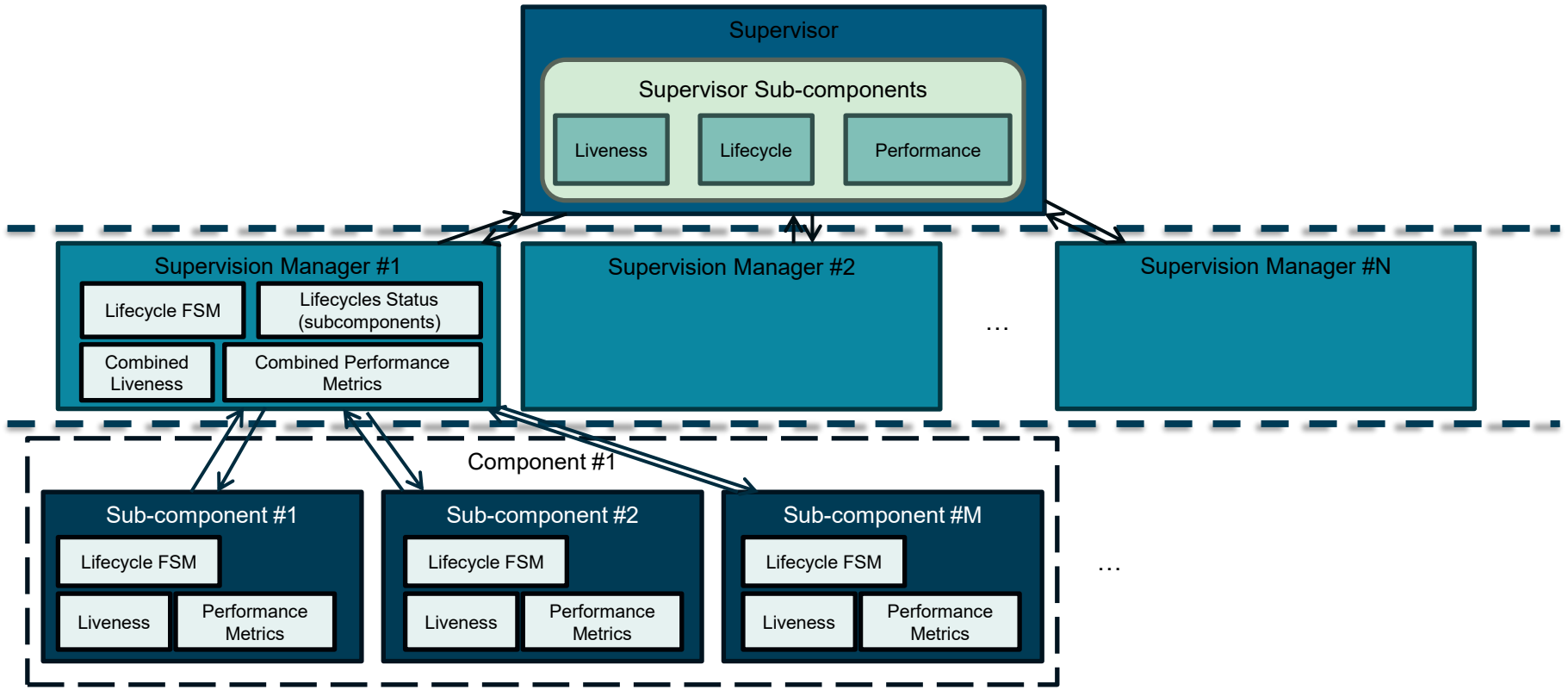
Supervisory overall scheme

Global ID	Module-local ID	Module	Signal	Loss Scenario Description	Context	Event ID	Cause	Cause type keywords	RPN (Loss)
364	5	Localization	Pose estimate	Localization does not provide pose estimate while <>.	<taking a curve, avoiding an obstacle/pedestrian, replanning trajectory>	UCA-8.a.1.	5- Flawed operation mode state implementation (state not set to operating). (1)	mode_management	720
365	6	Localization	Pose estimate	Localization does not provide pose estimate while <>.	<taking a curve, avoiding an obstacle/pedestrian, replanning trajectory>	UCA-8.a.1.	6- Unsafe state command by supervision (e.g., to ready state). (1)	mode_management	720
372	13	Localization	Pose estimate	UCA-8.a.3. Localization provides pose estimate with large offset error from true states while <> in <>.	<in a junction O1.3, in a curve, in a merge maneuver, avoiding a pedestrian>, <rain, clouds, GNSS interference zone>	UCA-8.a.3.	2- Not updating the localization estimate due to poor operation mode maintenance. (2)	mode_management	720
389	30	Localization	Pose estimate	Localization stops providing pose estimate while <>	<in a junction O1.3, in a merge maneuver, avoiding obstacles, reaching trajectory end point>	UCA-8.a.7.	5- Flawed operation mode state implementation (state not set to operating). (1)	mode_management	576
390	31	Localization	Pose estimate	Localization stops providing pose estimate while <>	<in a junction O1.3, in a merge maneuver, avoiding obstacles, reaching trajectory end point>	UCA-8.a.7.	6- Unsafe state command by supervision (e.g., to ready state). (1)	mode_management	576
376	17	Localization	Pose estimate	UCA-8.a.4. Localization provides pose estimate too late (missing updates) while <>.	<overtaking another vehicle, in a roundabout with traffic, replanning trajectory>	UCA-8.a.4.	2- Localization operation mode instability (unsafe interaction with supervisor). (3)	mode_management	432
380	21	Localization	Pose estimate	UCA-8.a.5. Localization provides pose estimate with very low frequency while <>.	<overtaking another vehicle, in a roundabout with traffic, replanning trajectory>	UCA-8.a.5.	3- Localization operation mode instability (unsafe interaction with supervisor). (3)	mode_management	324
384	25	Localization	Pose estimate	UCA-8.a.6. Localization provides pose estimate intermittently while <>.	<overtaking another vehicle, in a roundabout with traffic, replanning trajectory>	UCA-8.a.6.	3- Localization operation mode instability (unsafe interaction with supervisor). (3)	mode_management	324

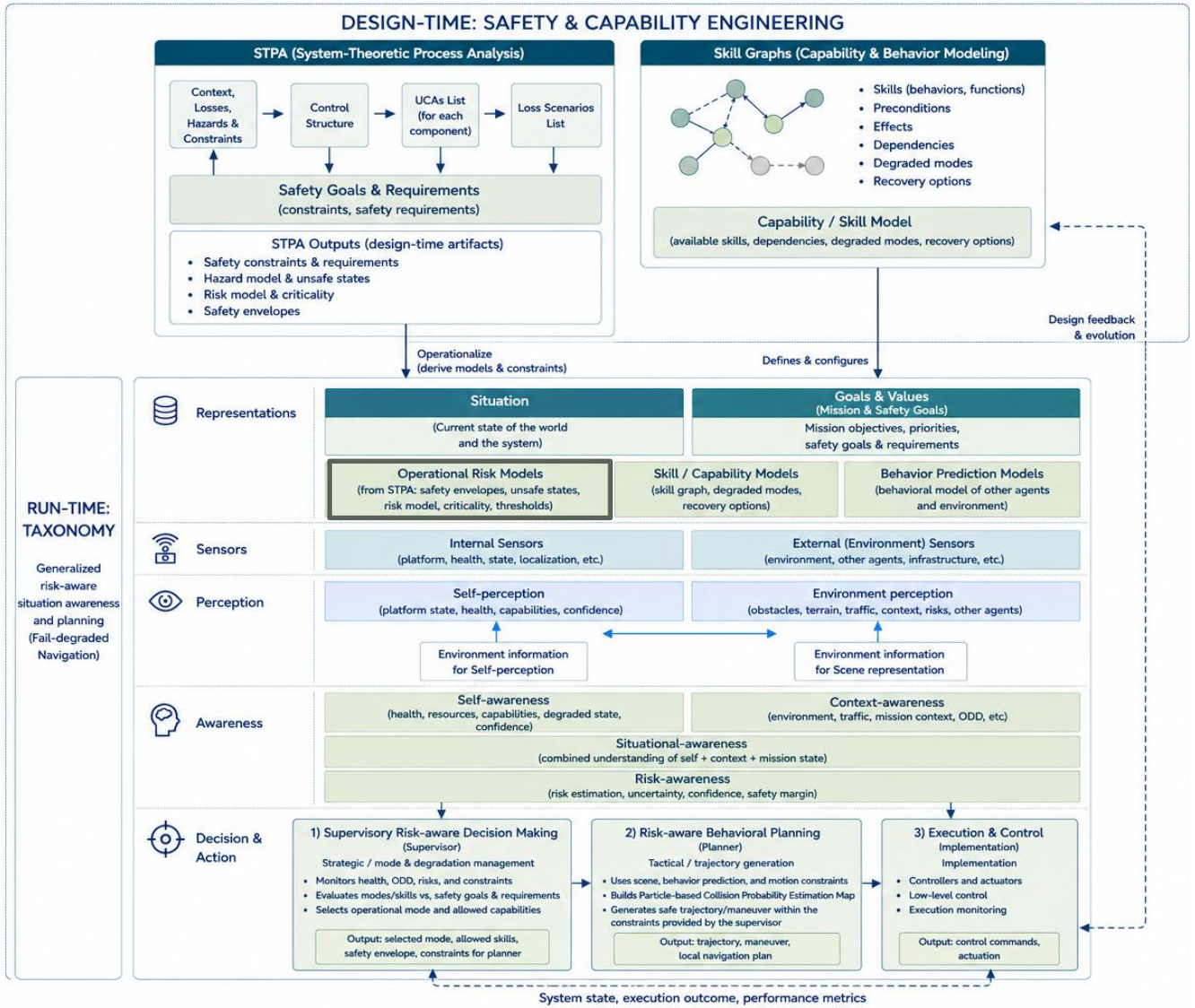
...

Risk prioritization based on:
$R \propto S \cdot L \cdot 1/T \cdot CoK$, with

- S: severity
- L: likelihood
- T: available response time
- CoK: confidence/strength of knowledge



Autopia approach for monitoring, awareness and planning

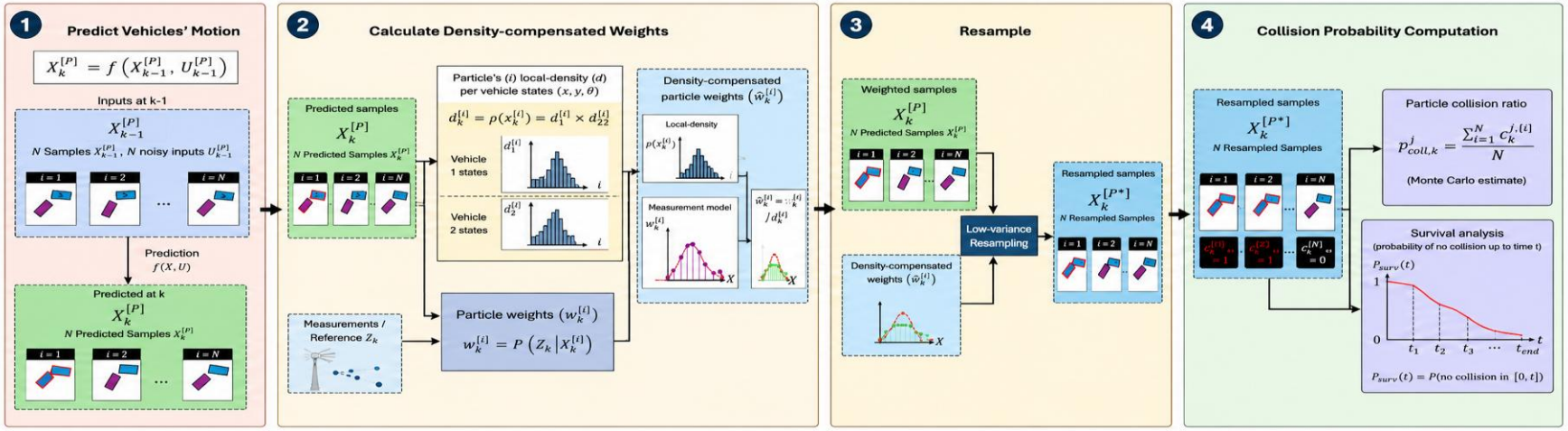


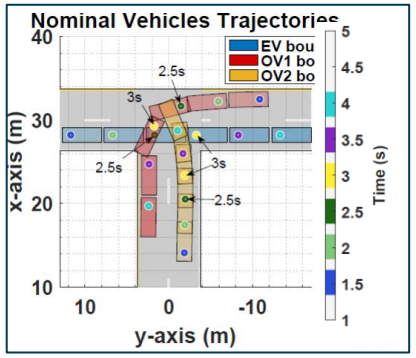
Particle-based Collision Probability Estimation

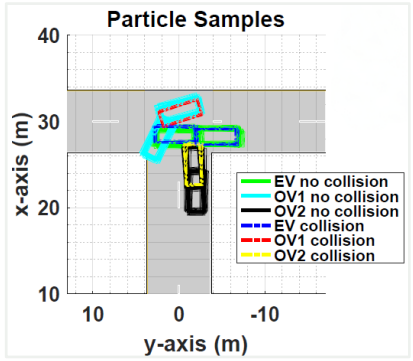
Existing approaches struggle to capture **low-probability/high-impact** collision scenarios caused by **uncertainty propagation** and complex multi-agent **interactions**

👇

- Effective probabilistic representation of traffic participants uncertainties
- Accurate simulation of uncertainty distribution in the presence of collisions

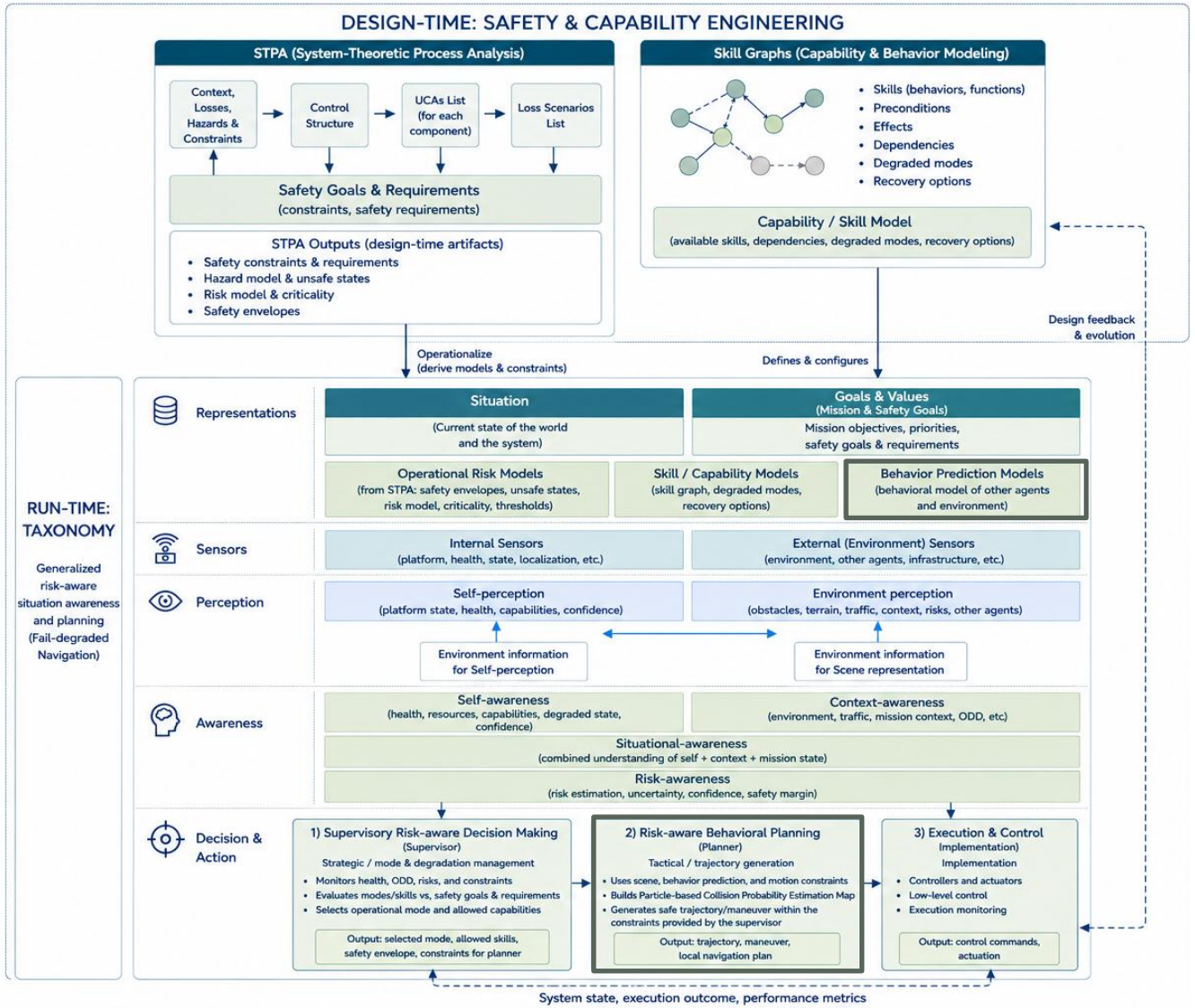




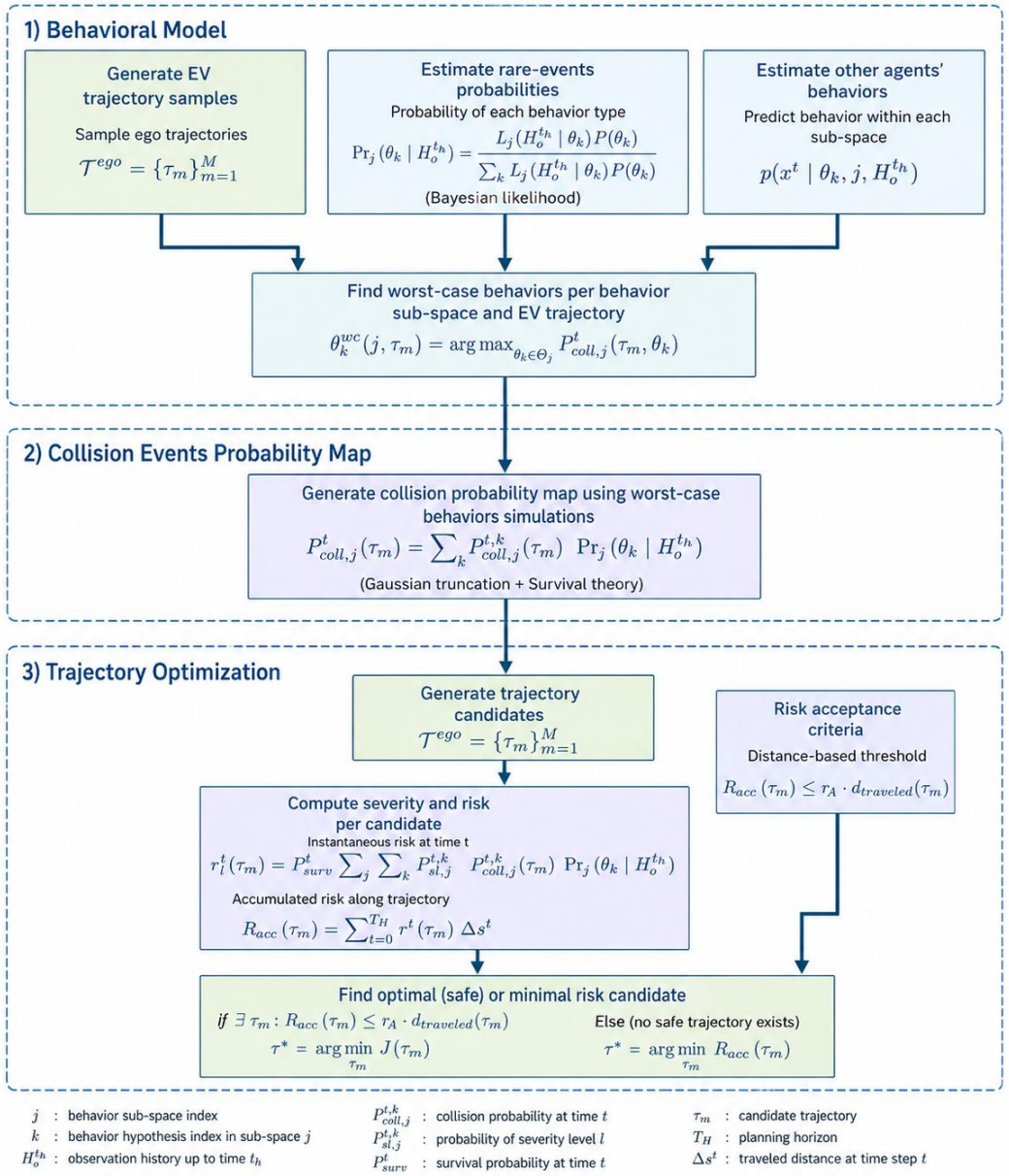


Hossam, Trentin, Jiménez-Bermejo, Villagra, Navas & Milanés. A Particle-Based Collision Probability Estimation Framework for Uncertainty-Aware Risk Evaluation in Autonomous Vehicles. IEEE Access, 2026

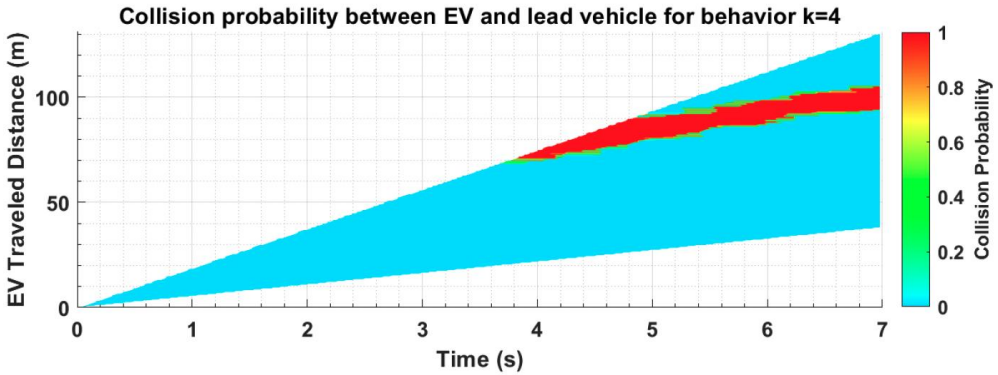
Autopia approach for monitoring, awareness and planning



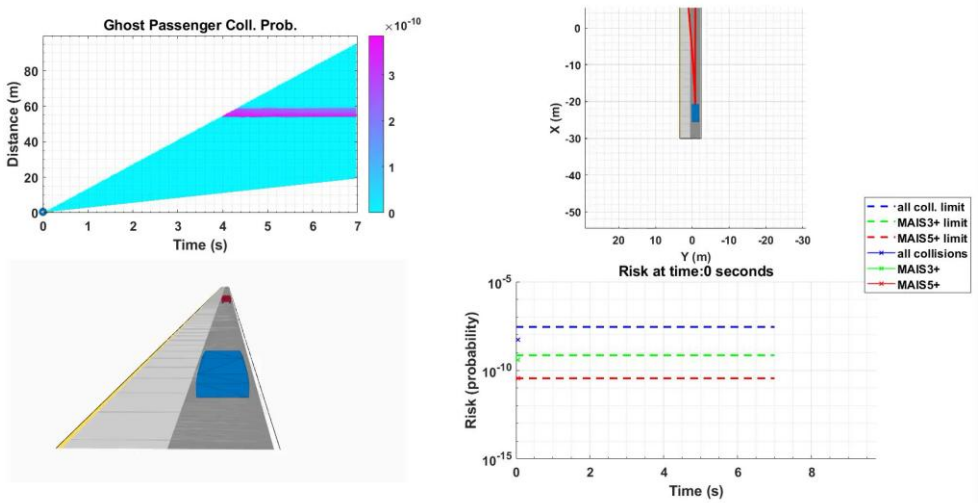
Risk-aware monitoring and decision-making: main principles



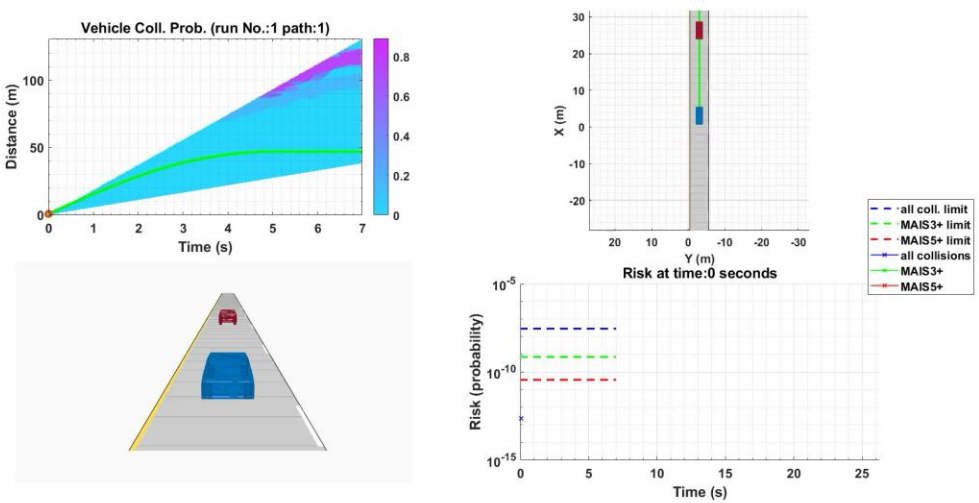
- Probabilistic framework that enables **interpretable** run-time reasoning about risk over the full planning horizon
- **Unified representation** of **collision and injury risk**, allowing consistent handling of **nominal** interactions, **rare** critical events, and **inevitable** collision situations
- Risk is represented across multiple severity levels aligned with **automotive safety** definitions → risk-constrained **planning + minimum-risk** maneuver selection



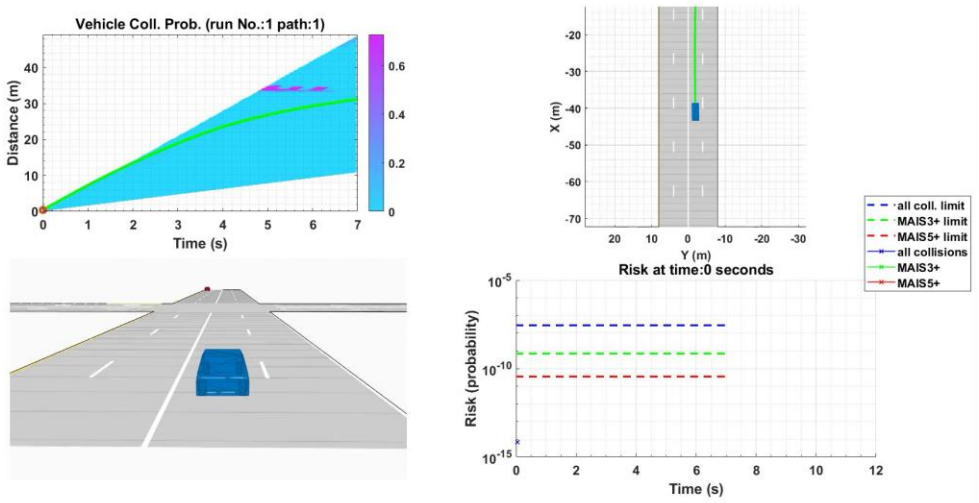
Risk-aware monitoring and decision-making: first results



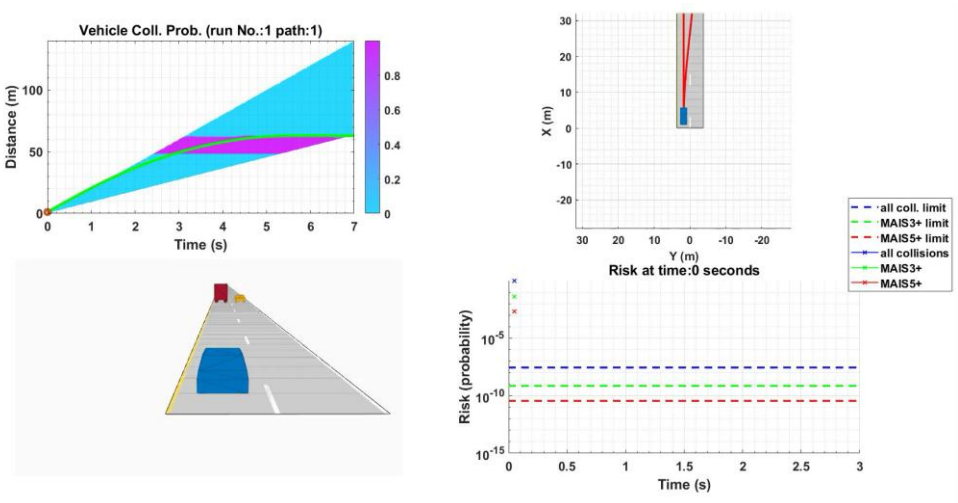
Door opening



Emergency stop



Intersection



Inevitable collision

Hossam, Artuñedo, Villagra, Risk-aware Monitoring and Decision-making framework for Autonomous Vehicles, to be submitted to IEEE Transactions on Intelligent Vehicles, 2026

Autopia in numbers

Since 2020:

- **Funding**: 4.5M€+ funding secured — 6 national/regional, 8 EU projects + 5 industry contracts
- **Scientific activity**: 82 scientific publications — 32 journals, 31 conferences, 3 books, 16 book chapters (64 co-authors / 12 countries)
- **Innovation**: TRL7 validated prototypes — road vehicles (2), bus (1), military platoon (1), semi-autonomous excavator (1)
- **Knowledge transfer**: 1.5M€ industrial contracts, 2 patents (1 with Renault)
- **Education**: 9 PhDs (4 completed / 5 ongoing), 25+ MSc/BSc theses
- **Team growth**: 6 → 18 members

Jorge Villagra

Jorge Godoy

Antonio Artuñedo

Permanent staff

Juan L. Hortelano

Marcos Moreno

Abdallah Hossam

Chema Barberá

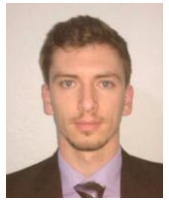
Ignacio López

PhD students

Davide Miceli

Management

Vinicius Trentin

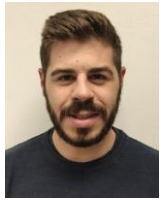


Victor Jimenez

Clara Gómez

Post-doctoral researchers

Miguel Beteta



Gabriel Delgado

David Redondo

Diego Aceituno

Arantxa García

Bárbara Villalba

R&D engineers


Thanks for your attention
https://autopia.car.upm-csic.es
9420 HNW