



# Steering Innovation for Autonomous Vehicles Towards Societally Beneficial Outcomes

Thomas Krendl Gilbert

Cathy Wu

Michael Dennis

June 2021

The Day One Project offers a platform for ideas that represent a broad range of perspectives across S&T disciplines. The views and opinions expressed in this proposal are those of the author(s) and do not reflect the views and opinions of the Day One Project or its S&T Leadership Council.

## Summary

Vehicle automation, coupled with simultaneous mobility revolutions of vehicle electrification and ridesharing, is set to have major impacts on society—perhaps the biggest impacts of any development in transportation since the introduction of cars over 100 years ago. But whether those impacts will be positive or not is still unknown. For example, widespread deployment of AVs could slash U.S. energy consumption by as much as 40% due to improved driving efficiency; alternatively, it could double U.S. energy consumption due to increased availability of cheap transport options.<sup>1</sup> Similar uncertainty surrounds the potential impacts of AVs on physical safety, transportation access for disabled communities, overall traffic efficiency, and long-term greenhouse-gas emissions. Guiding the evolution of AVs towards the future we want requires evaluating AVs using metrics that prioritize societally beneficial outcomes. The Biden-Harris administration should create an Evaluation Innovation Engine at the Department of Transportation (DOT) to propose, refine, and standardize public-interest metrics for AVs.

The Evaluation Innovation Engine (EIE) would do for AV metrics what the Defense Advanced Research Projects Agency (DARPA) Grand Challenge did for AV development: ignite productive competition among companies to achieve state-of-the-art performance. The EIE should have two main tasks (1) convening stakeholders to discuss potential metrics and providing opportunities for public comment on how proposed metrics should be prioritized, and (2) administering annual funding rounds of ~\$72 million each for private firms and other entities to create, test, and optimize algorithms for publicly beneficial AV outcomes. The EIE should be overseen by the Secretary of Transportation and staffed by representatives from pertinent DOT offices (Office of Civil Rights, Office of Small and Disadvantaged Business Utilization, Office of Public Affairs) and administrations (National Highway Traffic Safety Administration (NHTSA), Federal Highway Administration (FHWA), Federal Motor Carrier Safety Administration (FMCSA), Federal Transit Administration (FTA)), as well as a broad coalition of civil-society advocates.

## Challenge and Opportunity

The Biden administration has recently faced pressure to speed up AV deployment and relax safety standards to counter Chinese tech progress and accelerate job creation.<sup>2</sup> A major policy impediment to responsible widescale deployment is the absence of suitable and sufficient metrics for the performance of AV fleets. At present, federal performance guidelines for AVs are oriented almost exclusively around the physical safety of individual AVs. The NHTSA has adopted and widely publicized SAE International's six "Levels of Driving Automation"<sup>3</sup> as a roadmap for evaluating the successful operation of AVs, but performance in each of these levels is defined

---

<sup>1</sup> Wadud, Z.; MacKenzie, D. Leiby, P. (2016). Help or hindrance? The travel, energy and carbon impacts of highly automated vehicles. *Transportation Research Part A: Policy and Practice*, 86: 1–18.

<sup>2</sup> Shepardson, D. (2021). Cruise urges Biden to back autonomous vehicle deployment boost. *Reuters*, June 1.

<sup>3</sup> SAE International. (2018). SAE International Releases Updated Visual Chart for Its "Levels of Driving Automation" Standard for Self-Driving Vehicles. December 11.

and evaluated only according to accident avoidance. The FMCSA and FHWA have similarly prioritized crash reduction above all else.

The single-minded federal focus on AV safety has contributed to a lack of metrics focused on outcomes such as AV equity, fuel efficiency, or economic impacts. The limited metrics that do exist are neither precise nor standardized for industry use. Part of the problem is that it is not clear whose responsibility it is to develop metrics for aspects of AVs that are not directly related to on-road outcomes. The National Institute on Standards and Technology (NIST), for instance, has begun to evaluate performance and measurement of industrial AVs.<sup>4</sup> But this work is being pursued without considering the effects of new fuel-efficiency standards already proposed by the Department of Energy (DOE) and the Environmental Protection Agency (EPA), and without coordinating among essential stakeholder-led work such as SAE International's Standards Development effort.<sup>5</sup> Similarly, traffic throughput is an easily simulated roadway feature that is already well-recognized and well-studied by technical communities. But throughput generally falls under the purview of different state and local governments who may not be interested in measuring the effects of AVs effects on regional traffic (or may lack the means and authority needed to do so).

## Plan of Action

Fragmentation of the transportation-stakeholder ecosystem means that there is currently no obvious pathway for developing and implementing a comprehensive suite of AV metrics.<sup>6,7</sup> The federal government should act to correct this problem by establishing an Evaluation Innovation Engine (EIE) at the Department of Transportation (DOT). Tasking a single entity with overseeing and coordinating efforts to examine, prioritize, and enact AV performance metrics will do much to guide AVs towards a future that is beneficial for all.

The EIE should undertake two complementary efforts to track and standardize robust AV-performance metrics while simultaneously strengthening DOT leadership on standard-setting for AVs. First, the EIE should establish a **Metrics Commission** to discuss needs and proposals for various performance metrics. Second, EIE should create a **Grand Challenge Ecosystem** that administers approximately \$72 million in funds every year for private companies and other entities to create, test, and optimize algorithms for publicly beneficial AV outcomes. More detail on each of these efforts is provided below.

---

<sup>4</sup> National Institute of Standards and Technology. (2021). Mobility Performance of Robotic Systems.

<sup>5</sup> Werner, A. (2019). New colours for autonomous driving: an evaluation of chromaticities for the external lighting equipment of autonomous vehicles. Colour Turn, 1.

<sup>6</sup> Wu, C.; et al. (2021). Flow: A Modular Learning Framework for Mixed Autonomy Traffic. IEEE Transactions on Robotics

<sup>7</sup> Vinitsky, E.; et al. (2018). Benchmarks for reinforcement learning in mixed-autonomy traffic. In: Conference on Robot Learning, Proceedings on Machine Learning Research, 87: 399–409.

## Effort 1: Metrics Commission

Following the successful examples of the Federal Aviation Administration (FAA) and the NHTSA<sup>8</sup>, the Metrics Commission will deliberate on unproven and needed metrics for assessing how AV fleets, at multiple scales of deployment, impact different aspects of society. The Commission will solicit public feedback to inform which metrics should be prioritized for development and implementation and assign priority metrics to appropriate agencies. For example, the Commission could ask DOE to pursue development of metrics for how AV deployment affects traffic congestion. The Commission would comprise representatives from relevant federal agencies outside DOT (e.g., NIST, DOE, EPA), civil-society groups (e.g., the National Association of Counties<sup>9</sup>, National League of Cities<sup>10</sup>, Self-Driving Coalition for Safer Streets,<sup>11</sup> the American Association of State Highway and Transportation Officials (AASHTO),<sup>12</sup> and the American Association of Motor Vehicle Administrators (AAMVA)<sup>13</sup>), and industry leaders (e.g., Waymo, Tesla, and the “Big Three” vehicle manufacturers in the United States—General Motors Company, Fiat Chrysler Automobiles, and Ford Motor Company).

Mission Scope of Metrics Commission	Timeline of Deliverables	Prospective Partners	Example Metrics
-Catalyze high-potential metrics -Prioritize metrics based on period of public comment, input from civil society groups	~4 months: curate list of possible metrics ~2 months: period of public comment to prioritize metrics ~6 months: competition to achieve specified performance on curated metrics	-Federal agencies (NHTSA, FHWA, FMCSA, FTA) -Major automakers (GM, Fiat Chrysler, Ford) -Large tech firms (Tesla, Waymo, Cruise)	-Near term: physical safety / accessibility -Medium term: energy consumption / commute times -Long term: food deserts / reshaping the urban landscape

We expect that the Metrics Commission would deliberate on a particular “metrics round” for a 6–12 month period. 2–4 months would be dedicated to curating a list of possible metrics. 1–2 months would be dedicated to a public-comment period to help revise and prioritize those metrics, aiming to boil the final list down to about five. The remaining 3–6 months would be dedicated to an open-funding competition for companies to develop and demonstrate schemes

<sup>8</sup> Pruitt, C. (1979). People doing what they do best: The professional engineers and NHTSA. *Public Administration Review*, 39(4): 363–371.

<sup>9</sup> Istrate, E.; Harris, J. (2017). The future of work: The rise of the gig economy. National Association of Counties, November.

<sup>10</sup> Rouse, D.C.; et al. (2018). *Preparing communities for autonomous vehicles*. American Planning Association.

<sup>11</sup> Kang, C. (2016). Self-driving cars gain powerful ally: The government. *The New York Times*, September 19.

<sup>12</sup> Dakota, North. "American Association of State Highway and Transportation Officials." (2010).

<sup>13</sup> Hallmark, S.; Veneziano, D.; Litteral, T. (2019). *Preparing local agencies for the future of connected and autonomous vehicles*. Minnesota Department of Transportation, Research Services & Library, No. MN/RC 2019-18.

for achieving outstanding performance on those metrics in practice. The goal of the open-funding competition would be to financially support companies in coming up with smart and innovative ways to ensure that deployment of AVs at multiple scales achieves societally beneficial outcomes including and beyond improved physical safety. We expect that many companies will experience growing pains as they struggle to meet performance benchmarks for a wide suite of metrics. Lessons learned from the open-funding competition will make it easier for all companies to meet performance benchmarks for a wide suite of metrics if and when those metrics become standard components of AV evaluations. Lessons learned from the open-funding competition will also help federal agencies decide whether proposed metrics are ready for general use. The open-funding competition would steer grants towards companies with the most robust proposals, as evaluated by factors including (i) innovative use of tools such as artificial intelligence and simulation techniques, demonstrated partnerships, and clear deployment pathways.

## *Effort 2: Grand Challenge Ecosystem*

The Grand Challenge Ecosystem will provide funding for labs and companies to create, test, and optimize algorithms for AV outcomes aligned with priority metrics laid out by the Metrics Commission. Funded efforts could include planning, simulations, field tests, operations, data management, or control. We expect that by matching just 0.5% of the most recent annual DOT budget of the Department of Transportation for each of five annual open funding rounds—an estimated total of \$362 million, or \$72 million per round—the Grand Challenge Ecosystem will accelerate development and deployment of AV software, hardware, data, and partnerships oriented towards societally beneficial outcomes. This proposed funding level represents less than a quarter of the DOT-led budget for the comparable Smart City Challenge.<sup>14</sup> The Grand Challenge Ecosystem will also foster emergence of clear market incentives for companies to meet or exceed performance thresholds determined to be in the public interest.

## *Synergies*

The Metrics Commission and the Grand Challenge Ecosystem are designed to complement each other under the broader EIE, balancing competitive seed funding for metrics development and testing with deeper investment in technologies and strategies for achieving outcomes measured by those metrics. We expect the experiences of Grand Challenge Ecosystem participants to inform specifications of metrics ultimately recommended by the Metrics Commission for wide use. The existence of the Ecosystem would also help give the Metrics Commission power and purpose by ensuring that there is a pool of companies ready and willing to road-test new metrics as they are proposed. In turn, the Commission's recommendations will signal to current and future companies which metrics and technologies are most urgently needed and would be most likely to receive federal endorsement and funding. In this way, the federal government's capacity

---

<sup>14</sup> Shaheen, S.; Cohen, A.; Martin, E. (2017). The US Department of Transportation's Smart City Challenge and the Federal Transit Administration's Mobility on Demand Sandbox: Advancing Multimodal Mobility and Best Practices Workshop. Transportation Research Circular, E-C219.

to set broad agendas and national priorities will be made to align with the capacity of private interests to innovate in the market.

## Conclusion

The appeal of the proposal outlined in this memo can be framed in terms of financial and regulatory costs. Uncoordinated, unfocused federal investments into AV performance standards may generate unnecessary cost overruns and delays due to public officials repeatedly changing and updating requirements while AV fleets are being developed and rolled out. Such a scenario recently played out after the 2009 stimulus bill, when billions of dollars intended for California high-speed rail was redirected to unnecessary consultant work that led to years of delays and no final product.<sup>15</sup> By contrast, targeting federal funds for private-sector research and development (R&D) activities definitively aligned with federal priorities will yield actionable performance outcomes faster and at a fraction of the price. The Evaluation Innovation Engine comprises a pre-emptive, common-sense approach for optimizing AV deployment by (1) identifying and prioritizing a suite of metrics oriented at societally beneficial outcomes (i.e., the work of the Metrics Commission), and (2) funding companies to develop technologies and strategies for realizing those outcomes in practice (i.e., the activities funded under the Grand Challenge Ecosystem). Lessons learned and capacities developed as part of the Grand Challenge Ecosystem will enable the Metrics Commission to further refine and standardize metrics for broad use, while the existence of the Grand Challenge Ecosystem will ensure that metrics proposed by the Metrics Commission help shape the evolution of the AV market for the public good.

---

<sup>15</sup> Vartabedian, R. (2019). How California's faltering high-speed rail project was 'captured' by costly consultants. Los Angeles Times, April 26.

## Frequently Asked Questions

### 1. Have there been any promising precedents for the proposed Evaluation Innovation Engine?

This proposal draws inspiration from several past federal and private-sector initiatives. One is the Federal Aviation Agency (FAA), which was established as an independent body before being folded into DOT once DOT was founded. Just as the FAA helped set clear performance standards for—and accelerate domestic progress—in aviation, so too will the EIE help ensure strong, consistent performance and progress in the AV sector. The more recent success of the Advanced Research Projects Agency-Energy (ARPA-E) at DOE<sup>16,17</sup> exemplifies how federal funding can advance high-impact technologies that lack targeted investment from the private sector. MIT’s “Innovation Orchard”, which aims to foster promising high-risk tech development from industry, inspired the Grand Challenge Ecosystem component of the EIE.<sup>18</sup>

### 2. What challenges might the Biden-Harris administration encounter from industry in launching the EIE?

The administration may receive pushback if the EIE was overly prescriptive, as AV manufacturers and stakeholders are wary of onerous requirements that could slow their innovation efforts. However, the authors of this memo have personally found that there is a strong desire for a central leadership and standardization of the still-nascent AV ecosystem. A prominent government effort to offer clear guidelines for AV developers would likely be met with a warm reception as such guidelines would provide regulatory certainty and direction.

### 3. Why is it important to include non-governmental stakeholders in the Metrics Commission?

Befitting the diverse makeup of municipal and regional advocacy groups, the proposed commission will have process-, performance-, and outcome-oriented goals. Organizationally, the commission will establish critical buy-in from major stakeholders and align public-private incentives through a combination of voice, federal standard-setting, and reduction of market uncertainty, being of mutual benefit to all participants. Regarding performance, the commission will ensure that public-private consensus is reached on standards in a manner that is equitable, transparent, and politically accountable, as opposed to an industry-led consortium that would

---

<sup>16</sup> Bonvillian, W.B.; Van Atta, R. (2011). ARPA-E and DARPA: Applying the DARPA model to energy innovation. *The Journal of Technology Transfer*, 36(5): 469.

<sup>17</sup> Goldstein, A.; et al. (2020). Startups supported by ARPA-E were more innovative than others but an investment gap may remain. *Nature Energy*, 5(10): 741–742.

<sup>18</sup> Learly, K. (2017). MIT’s ‘Innovation Orchard’ is Investing \$200 Million in Tough Tech Startups. *Futurism*, September 22.

be ill-prepared to unilaterally evaluate or anticipate the emergent effects of AVs. In terms of outcome, the commission will establish a sense of urgency regarding the immediate need for firms, agencies, and researchers to prepare appropriate metrics for the anticipated societal effects of AVs.

#### **4. How would tradeoffs and multi-stakeholder disagreements be handled by the Metrics Commission?**

We expect that in the first several years of its existence, the Commission will focus on performance outcomes for which there is near-unanimous agreement that metrics need to be created or improved. However, we also anticipate that the Commission will encounter disagreements about how to weigh different societal considerations when developing and prioritizing metrics (e.g., environmental vs. economic impacts) and long-term stakeholder interests. The Commission would strive to raise these issues proactively and to solicit feedback during public-comment periods to help resolve such disagreements before they become intractable. If the Commission is unable to resolve a disagreement itself, we expect that its input on both sides of the debate would help leadership in the administration and in Congress address the disagreement through executive direction or legislation, respectively.

#### **5. How can we be sure that the EIE will not stifle innovation?**

The EIE is not intended to issue prescriptive technical standards that could slow innovation. Rather, the EIE is designed to complement existing efforts focused on maximizing safety and reducing costs by guiding and funding companies to pursue a broad suite of societally beneficial outcomes that currently deliver little or no direct financial returns. By incentivizing development and adoption of socially beneficial metrics, the EIE will spur—not stifle—innovation in a way that benefits society writ large.

#### **6. How does this proposal fit within existing reporting standards or regulations?**

At present, major AV firms are required to report “vehicle disengagements” (i.e., situations in which a human driver retook control of the wheel of a car) to NHTSA. However, these requirements are notoriously vague and subject to the discretion of individual firms, which get to decide what constitutes a disengagement (vs. system anomaly vs. other situations). Beyond this early and poorly specified standard, few other metrics exist even for voluntary monitoring in-house at major firms. Our proposal thus fills a major gap in motivating the creation and continued maturation of additional metrics for AV performance in a variety of settings.



## About the Authors



**Thomas Krendl Gilbert** is an interdisciplinary Ph.D. candidate in Machine Ethics and Epistemology at UC Berkeley. He researches the predicaments that emerge when artificial intelligence reshapes the context of organizational decision-making. His work investigates how specific algorithmic learning procedures reframe classical ethical questions and recall the foundations of democratic political philosophy, namely the significance of popular sovereignty for resolving ambiguities in norms. This work has concrete implications for the design of automated vehicle systems that are fair for distinct subpopulations, safe when enmeshed with municipal practices, and accountable to public concerns.



**Cathy Wu** is an Assistant Professor at MIT in the Laboratory for Information and Decision Systems. She holds a Ph.D. from UC Berkeley, as well as a B.S. and M.Eng in Electrical Engineering and Computer Science from MIT. She completed a postdoc at Microsoft Research. Cathy's interests are broadly in machine learning and mobility. She studies the technical challenges surrounding the integration of autonomy into societal systems. Her work has been acknowledged with several awards, including the 2019 Institute of Electrical and Electronics Engineers (IEEE) Intelligent Transportation Systems Society (ITSS) Best Ph.D. Dissertation Award, the 2016 IEEE ITSC Best Paper Award, and fellowships from the National Science Foundation (NSF), the Berkeley Chancellor, the National Defense Science and Engineering Graduate (NDSEG) Graduate Program, and the Dwight David Eisenhower Transportation Fellowship Program. Her work has appeared in the press, including *Wired* and *Science* magazines.



**Michael Dennis** is a Ph.D. candidate at the Center for Human-Compatible AI at UC Berkeley. He studies how AI ought to make decisions, both in isolation and when interacting with other AI systems or human agents. His work has implications for AI safety, robustness in reinforcement learning, and multi-agent systems. He believes that strong normative theories of decision making are critical for understanding how AI ought to be designed, as well as for making interaction between AI systems and society at large more robustly beneficial.



## About the Day One Project

The Day One Project is dedicated to democratizing the policymaking process by working with new and expert voices across the science and technology community, helping to develop actionable policies that can improve the lives of all Americans, and readying them for Day One of the next presidential term. For more about the Day One Project, visit [dayoneproject.org](https://dayoneproject.org).