BEFORE THE PUBLIC UTILITIES COMMISSION
OF THE STATE OF CALIFORNIA

Order Instituting Rulemaking on Regulations Relating to Passenger Carriers, Ridesharing, and New Online-Enabled Transportation Services	R.12-12-011 (Filed December 20, 2012)

REPLY COMMENTS OF THE FUTURE OF PRIVACY FORUM ON THE COMMISSION'S DECISION AUTHORIZING AUTONOMOUS VEHICLE PASSENGER SERVICE WITH DRIVERS AND ADDRESSING IN PART ISSUES RAISED IN THE PETITIONS FOR MODIFICATION OF GENERAL MOTORS, LLC/GM CRUISE, LLC, LYFT, INC., AND RASIER-CA, LLC/UATC, LLC FOR PURPOSES OF A PILOT TEST PROGRAM FOR DRIVERLESS AUTONOMOUS VEHICLE PASSENGER SERVICE

I. INTRODUCTION

Autonomous vehicle technologies ("AVs") are enabled by the collection of new types of data, making consumers' right to privacy and data protection a critical issue to consider when regulating these vehicles. Given the large amount of data autonomous vehicles will collect, as well as consumers' expectations of privacy, the Commission's actions are of significant interest to the Future of Privacy Forum ("FPF") as well as to many other organizations – the majority of opening comments raise data and privacy concerns. Of the 13 parties that filed opening comments: six parties recommended the Commission limit the data collected in pilots;[1] nine advised that if service data is collected, it must remain confidential;[2] and nine also advocated that communications between passengers and remote operators must remain private to protect consumer interests.[3] Privacy is enshrined as a personal right in California's Constitution."[4]

[1] *See* 60 Plus Comments at 5, SAFE Comments at 14, Zoox Comments at 4, Auto Alliance Comments at 4, GM Comments at 9, Waymo Comments at 10 and Lyft Comments at 12-13.
[2] *See* 60 Plus Comments at 5, Zoox Comments at 4-5, Auto Alliance Comments at 4, GM Comments at 9, Lyft Comments at 13, SVLG TechNet SDC CMTA Joint Comments at 10-11.
[3] *See* 60 Plus Comments at 5, Zoox Comments at 6, GM Comments at 12, Lyft Comments at 14-15, SVLG TechNet SDC CMTA Joint Comments 5-6 and Waymo Comments at 12.
[4] "All people are by nature free and independent and have inalienable rights. Among these are enjoying and defending life and liberty, acquiring, possessing and protecting property, and pursuing and obtaining safety, happiness and privacy." CAL. CONST. art. I, § I.

FPF recognizes that data can be used to improve public services and serve valuable social, economic, and democratic functions. However, sharing municipal datasets about individuals' use of transportation services and citizens' activities carries inherent risks to individual privacy. We have significant concerns that without robust technical, operational, and legal controls and safeguards,[5] the Commission's request for all communications between a passenger and a remote operator, as well as its intent to publicly share that data, could put citizens' privacy at risk while undermining consumers' trust in TCPs and the government. In addition to the Proposed Decision's request for data, the San Francisco International Airport, the San Francisco Municipal Transportation Agency, and the San Francisco County Transportation Authority (the "City") requested in its opening comments that additional, highly sensitive information be provided from AV permitees, including detailed data such as the GPS locations and to-the-minute times of vehicle pick-ups and drop-offs. We join the numerous organizations advocating for enhanced consumer privacy protections regarding communications between passengers and remote operators, believe that the City's additional data request would create substantial privacy risks, and urge the Commission to carefully consider its approach to data collection and sharing.

II. DISCUSSION

Inadequate privacy protections for personal data can lead to significant financial, physical, reputational, organizational, and societal harms. Governments must be vigilant and resourceful to ensure that the data they collect and use does not create privacy risks for citizens. Core privacy risks when personal information is shared with government agencies can include: re-identification of individuals; inaccurate, unfair, or inefficient decisions based on biased or faulty data; and loss of public trust in all involved entities, including the Commission, the City, and TCPs.

The Commission's Proposed Decision and the additional data collection requested by the City would create substantial privacy risks for Californians. The Commission should take steps to mitigate those risks before moving forward. The risks include: 1) potential exposure of personal data disclosed to the Commission as a result of data breaches, inappropriate access by insiders, public records requests, court orders, or administrative data sharing arrangements between California agencies; 2) potential

[5] For examples of the types of safeguards that could protect individual privacy in government data sharing programs, see Micah Altman et al., *Towards a Modern Approach to Privacy-Aware Government Data Releases*, 30 Berkeley Tech L.J. 1968 (2015), https://cyber.harvard.edu/publications/2016/Privacy_Aware_Government_Data_Releases; BEN GREEN ET AL., OPEN DATA PRIVACY PLAYBOOK (2017), https://dash.harvard.edu/bitstream/handle/1/30340010/OpenDataPrivacy.pdf?sequence=5; and Kelsey Finch & Omer Tene, *Smart Cities: Privacy, Transparency, Community*, CAMBRIDGE HANDBOOK OF CONSUMER PRIVACY (2018), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3156014.

identification or re-identification of individuals based on the contents of their communications data; and 3) potential identification or re-identification of individuals based on location data or other identifiers. The Commission should mitigate these risks when fashioning the data collection mandates in this proceeding, because: the Commission cannot ensure that data will not be made public; and the data at issue – communications data and location information – is particularly sensitive and difficult to de-identify.

1. Because the Commission cannot ensure that data will not be made public, it should minimize data collection and incorporate privacy safeguards.

While we recognize the benefit of TCP data, we are concerned that the Commission cannot ensure that data it collects will not become available to the public, including data that is intended to remain confidential. We are particularly concerned in light of: 1) the potential vulnerability of such information to data breaches; 2) the risk of employees or other insiders improperly accessing confidential data; 3) the state's disclosure obligations under the California Public Records Act; and 4) the absence of clear rules to address potential access to data by other agencies, including law enforcement.

Our concern is heightened by the Proposed Decision's lack of specificity regarding how the Commission will ensure that communications records are appropriately and consistently "anonymized" and "disaggregated" before publication.[6] Disclosure control experts consider free-form communications data and location data to be particularly challenging to de-identify.[7] Before collecting sensitive or personal data, whether intended to be public or confidential, the Commission should establish clear, robust safeguards addressing potential privacy and security risks.[8]

2. Communication between passengers and remote operators could contain sensitive information, and the disclosure contemplated by the Commission could create serious privacy risks.

We share the concern of other commenters that the reporting requirement for communication between passengers and operators poses a significant privacy risk. We are particularly concerned by the lack of detail regarding how such information will be collected, used, and shared. This data will most likely be unstructured or free-format data, which experts find unpredictable and challenging to de-

---

[6] Proposed Decision, p. 23, 33, 35-36, 39, 42. The Proposed Decision uses the term "anonymous," but appears to refer to data that is more accurately described as "de-identified." For the sake of clarity, we use the term "de-identified" throughout.

[7] *See* DE-IDENTIFYING PERSONAL INFORMATION NISTIR 8053 (NIST Oct. 2015), http://nvlpubs.nist.gov/nistpubs/ir/2015/NIST.IR.8053.pdf; SP 800-188: DE-IDENTIFYING GOVERNMENT DATASETS (NIST draft. Aug. 2016), http://csrc.nist.gov/publications/drafts/800-188/sp800_188_draft2.pdf.

[8] For guidance, we note that FPF has worked with public entities in the past to set reasonable safeguards for both datasets that are intended for public release and those that, due to the sensitivity or identifiability of the data they contain, should not be widely accessible. *See, e.g.*, FUTURE OF PRIVACY FORUM, CITY OF SEATTLE OPEN DATA RISK ASSESSMENT (2018), https://fpf.org/wp-content/uploads/2018/01/FPF-Open-Data-Risk-Assessment-for-City-of-Seattle.pdf and Finch & Tene, *supra* note 5.

identify.[9] Further, requiring AV passenger services to provide versions of all communications data suitable for public release within 24 hours may impede the technical and operational efforts required to adequately de-identify such data. In large datasets of unstructured or text data, the risk of personal data being overlooked is high and the consequences can be serious.[10] Finally, passengers who are not certain how and for what purposes their communications will be collected may bes reluctant to use AV passenger services or to candidly communicate with remote operators about critical issues.

3. The City's requests for additional service data raise additional, significant privacy and security concerns.

The City's opening comments recommend that the scope of the Public Data Reporting Requirement be expanded to include additional, highly sensitive consumer data elements. Among the additional elements, the City requests that the Commission require AV permittees to add GPS location with timestamps of every passenger pick-up and drop-off, linked to a specific trip and identifier; and geographical information regarding trip location within census blocks or zip codes for each day and month.[11] Mandated reporting of vast quantities of sensitive location data creates heightened privacy and security risks for individuals, and the precise nature of GPS data renders it uniquely sensitive. There is substantial evidence that historic logs of location data can be analyzed to identify individuals' travel – even when names are not linked to the underlying data.[12] Leaders within the City's own Open Data team have acknowledged the special risks and considerations necessary for location data in other projects, and should incorporate those lessons here.[13] If the CPUC is inclined to adopt any of the City's

---

[9] *See* NIST, *supra* note 7; Green et al., *supra* note 5; FPF, *supra* note 8. For example, directly identifying information (such as names) and indirectly identifying information (such as locations, times, or events) may not be clearly marked or even readily discernible in text narratives. For example, in some cases attempting to redact names can lead to individuals with uncommon names being mistakenly exposed in public data sets. In other cases, important information could be mistaken for personal information and be erroneously removed (such as not distinguishing a communication about a passenger named Allison from one about Allison Street in San Francisco).

[10] *See, e.g.*, Lauren Fitzpatrick, *CPS privacy breach bared confidential student information*, CHI. SUN-TIMES (Feb. 2, 2017), http://chicago.suntimes.com/news/cps-privacy-breach-bared-confidential-student-information/; Vince Lattanzio, *Philly paying $1.4 million after posting confidential gun permit information online*, NBC PHILADELPHIA, July 22, 2014, http://www.nbcphiladelphia.com/news/local/Philly-Paying-14M-After-PostingConfidential-Gun-Permit-Information-Online-268147322.html.

[11] *See* San Francisco International Airport, the San Francisco Municipal Transportation Agency, and the San Francisco County Transportation Authority Comments at 5.

[12] Several studies have demonstrated that even de-identified data can be "reverse engineered" to reveal passenger names and trip pick-up and drop-off location information. *See, e.g.,* Neustar Research, Riding with the Stars: Passenger Privacy in the NYC Taxicab Dataset (Sept. 15, 2014), https://research.neustar.biz/2014/09/15/riding-with-the-stars-passenger-privacy-in-the-nyc-taxicab-dataset/; Chris Whong, *FOILing NYC's Taxi Trip Data* (Mar. 18, 2014), https://chriswhong.com/open-%20data/foil_nyc_taxi/.

[13] *See* DATASF, OPEN DATA RELEASE TOOLKIT: PRIVACY EDITION https://datasf.org/resources/open-data-release-toolkit/.

recommendations, the Commission should 1) specify less precise, neighborhood-level data for all trips; and 2) enact concrete policies and procedures providing privacy and security safeguards for such data.

III. CONCLUSION

To achieve its stated mission to protect consumers, the Commission must take a thoughtful approach to consumers' data privacy and security. We commend the Commission for its desire to strike a balance between AV passenger safety and protecting individual privacy, such as by limiting its reporting requirements primarily to non-sensitive vehicle data and by requesting both confidential and public versions of sensitive communications data. However, as outlined by FPF and echoed by other stakeholders, it is crucial that the Commission minimize data collected to that essential to providing for the safety and consumer protection of AV passengers and commit to clear and robust privacy safeguards for any data that it collects. AVs have the potential to provide many benefits to Californians, and the Commission has an important role in ensuring that those benefits do not come at the expense of individual privacy.

Respectfully submitted,

LAUREN SMITH
FUTURE OF PRIVACY FORUM
1400 I STREET NW, SUITE 450
WASHINGTON DC 20005
(410) 775-6527
LSMITH@FPF.ORG
POLICY COUNSEL

Dated: May 4, 2018 in San Francisco, California