Safety through Agility: Using Mixed Reality to tune shared autonomy systems

Data Collection

What data will you collect or create?

We will collect driving data of human subjects as they are evaluated within the driving simulator. We have acquired an IRB for this at Penn.

How will the data be collected or created?

The data will be recorded during the use of the driving simulator

Documentation and Metadata

What documentation and metadata will accompany the data?

The driving data will be recorded with timestamps and an anonymous driver ID.

Ethics and Legal Compliance

How will you manage any ethical issues?

We have acquired an IRB for this at Penn.

How will you manage copyright and Intellectual Property Rights (IP/IPR) issues?

All anonymized data will be made available in an open-source dataset

Storage and Backup

How will the data be stored and backed up during the research?

Only anonymized data will be stored in public github repos.

No data related to the human subject name or ID will be stored.

How will you manage access and security?

Only anonymized data will be stored in public github repos.

No data related to the human subject name or ID will be stored.

Selection and Preservation

Which data are of long-term value and should be retained, shared, and/or preserved?

Only anonymized data will be retianed in public github repos.

What is the long-term preservation plan for the dataset?

The data will be preserved for a duration of 6 years.

Data Sharing

How will you share the data?

Only anonymized data will be stored in public github repos.

Are any restrictions on data sharing required?

No restrictions on anonymized data

Responsibilities and Resources

Who will be responsible for data management?

Co-I Helen Loeb will responsible for data management of the anonymized public data.

What resources will you require to deliver your plan?

We only require github

Planned Research Outputs

Dataset - "Guardian Angel"

Anonymized driving simulator data

Planned research output details

Title	Type	Anticipated release date	Initial access level	Intended repository(ies)	Anticipated file size	License	Metadata standard(s)	May contain sensitive data?	May contain PII?
Guardian Angel dataset	Dataset	2025-09-29	Open	None specified	10 MB	Creative Commons Attribution Non Commercial Share Alike 4.0 International	None specified	No	No

Data Management Plan

This supplementary document describes our data management plan, discussing types of data, standards, policies, and plans for data archiving and dissemination.

Types of Data
We expect this project to create numerous types of data which may include, but are not limited to: source code for implemented protocols; simulation, testing, and other software; measurements from testbed implementation; publications; technical reports; presentation and demonstration materials (*e.g.,* slides, videos);and curricular materials. Most of these data types will be generated in obvious ways, with the possible exception of measurements, which will be acquired through logging or data capture mechanisms built in to implementation packages.

Data and Metadata Standards
Standard plain text data formats (*e.g.,* ASCII, CSV, PDF) will be used for data and metadata format and content. Measurement data and simulation/emulation output will be initially stored in proprietary formats (*e.g.,* MATLAB .mat files) but later extracted to readable plain text formats as above. Data and metadata will also be written in researchers' individual notebooks and later converted to electronic formats for documentation or publication.

Policies for Access and Sharing and Provisions for Appropriate Protection/Privacy
Throughout the proposed project, we expect numerous publications and technical reports to be created. In accordance with DoT policy, these documents will be made available publicly, through appropriate conference and journal publication management systems (*e.g.,* IEEE Xplore, ACM Digital Library) which may require subscription or fee and/or archived on cps-vo.org and/or a public-facing research website to be created and maintained jointly by the Lead PIs. Other types of data will be made available as appropriate, either by contacting the PIs or by accessing a public repository on the above website. We do not expect to charge for any data access, but data sharing agreements may be necessary to allow access to certain "sensitive" data sets. In case of peer-reviewed publication, we may delay public release of documentation and data until the review process has completed. No embargo periods are currently required, so data will be made available at the time of publication. We retain data distribution rights, excepting company proprietary technology and data used or produced.

With the exception of spectrum data collected, and to the best of our capabilities, subject to IP policies of collaborators' institutions or employers, we will make all such resources publicly available. When necessary to comply with policies enforced on collaborators, we will maintain a private archive on the above-referenced domain to which access can be granted on a user-by-user or group-by-group basis, using the underlying UNIX file permission and access control systems. In extreme cases, when deemed absolutely necessary by policy, we will refrain from storing certain sensitive information on any Internet-facing server in order to protect privacy, confidentiality, or intellectual property. In such cases, the sensitive information, data set, or file will be stored and backed-up on an external drive and only accessed by approved devices.

With respect to any wireless data collected, we will ensure that stored radio data are in full compliance with applicable laws, including requirements for data privacy. We will place mechanisms including hardware and software filters to restrict wireless recording to specific bands that are open (e.g. ISM bands) or we have supporting FCC licenses. We will not permit any third party to access the data

unless they accept the same responsibilities pertaining to data management and privacy.

Policies and Provisions for Re-use, Re-distribution

We expect researchers in industry and at other universities to take advantage of software and data that we make available. We retain the right to place permission restrictions on reuse and redistribution of data, but none are currently in place. The exception is any data considered proprietary or intellectual property of an industry collaborator, in which the company has control over data release and re-distribution rights. In this case, decisions for reuse and redistribution of data will be made on a case-by-case basis.

Plans for Archiving and Preservation of Access

Research products and data are planned to be archived and preserved access to according to standard University of Pennsylvania (Penn) practices and policies. There are a number of data servers at Penn campuses, running a variety of UNIX based operating systems and serving most of the necessary services to our community, including the server hosting the research website mentioned above. All data on the Penn servers has data backups including backups with offsite storage, to mitigate data loss due to hardware failure, and vital data is backed up daily and can be recovered from these backups if necessary.

Data security from outside (*e.g.,* Internet) is provided by a state-of-the-art firewalls. None of the data servers have their services exposed to the outside world directly, although many services are accessible through VPN or proxy machines. Incoming communication from outside is constantly monitored and suspicious activity is reported to administrators. Internal security is provided with a number of measures. Communication with servers is only possible for those possessing valid credentials. Physical access to data centers is restricted to trained and duly authorized staff members.