Breaking Barriers: Workshop on Open Data Practices in AutoUI Research

Patrick Ebel
ebel@informatik.uni-leipzig.de
ScaDS.AI, Leipzig University
Leipzig, Germany

Pavlo Bazilinskyy
p.bazilinskyy@tue.nl
Eindhoven University of Technology
Eindhoven, Netherlands

Angel Hsing-Chi Hwang
hh695@cornell.edu
Virtual Embodiment Lab, Cornell University
New York City, USA

Wendy Ju
wendyju@cornell.edu
Information Science, Cornell Tech
New York City, USA

Hauke Sandhaus
hgs52@cornell.edu
Information Science, Cornell University
New York City, USA

Aravinda Ramakrishnan Srinivasan
a.r.srinivasan@leeds.ac.uk
Institute for Transport Studies, University of Leeds
Leeds, United Kingdom

Qian Yang
qy242@cornell.edu
Computing and Information Science, Cornell University
New York City, USA

Philipp Wintersberger
philippwintersberger@gmail.com
Department of Digital Media, University of Applied Sciences Upper Austria
Hagenberg, Austria
Institute of Visual Computing and Human-Centered Technology, TU Wien
Wien, Austria

ABSTRACT

While the benefits of open science and open data practices are well understood, experimental data sharing is still uncommon in the AutoUI community. The goal of this workshop is to address the current lack of data sharing practices and to promote a culture of openness. By discussing barriers to data sharing, defining best practices, and exploring open data formats, we aim to foster collaboration, improve data quality, and promote transparency. Special interest groups will be formed to identify parameter sets for recurring research topics, so that data collected in different individual studies can be used to generate insights beyond the results of the individual studies. Join us at this workshop to help democratize knowledge and advance research in the AutoUI community.

CCS CONCEPTS

• **Human-centered computing** → **Human computer interaction (HCI)**; • **Computing methodologies** → **Modeling and simulation**.

KEYWORDS

Automotive User Interfaces, Datasets, Open Science

ACM Reference Format:
Patrick Ebel, Pavlo Bazilinskyy, Angel Hsing-Chi Hwang, Wendy Ju, Hauke Sandhaus, Aravinda Ramakrishnan Srinivasan, Qian Yang, and Philipp Wintersberger. 2023. Breaking Barriers: Workshop on Open Data Practices in AutoUI Research. In *15th International Conference on Automotive User Interfaces and Interactive Vehicular Applications (AutomotiveUI '23 Adjunct), September 18–22, 2023, Ingolstadt, Germany.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3581961.3609835

AutomotiveUI '23 Adjunct, September 18–22, 2023, Ingolstadt, Germany

ACM ISBN 979-8-4007-0112-2/23/09.
https://doi.org/10.1145/3581961.3609835

1 INTRODUCTION

There are many benefits to an open science approach. Open science, and open data practices in particular, can improve rigor and reproducibility of research [15], thereby increasing trust in scientific results [1]. Making data and other resources, such as code, accessible so that experiments can be easily reproduced by other researchers can significantly improve scientific efficiency and encourage collaboration [13, 16]. In addition, open science approaches make scientific resources accessible, allow a broader community to participate in research, and make scientific findings more understandable to a wider population [2]. They thus contribute to the democratization of knowledge.

These benefits serve as the main incentives for many academics to share experimental data and associated materials. While these practices are standard in fields such as medicine [5, 12] or governance [11], sharing data and making code open source is the

exception, not the norm in the AutoUI community. While there are some positive examples of researchers making their data and code available to the research community [4, 17, 18], often, and especially when it comes to naturalistic driving studies, authors do not share their data [9, 14, 21]. One reason for this may be that these studies are often conducted in collaboration with automotive original equipment manufacturers (OEMs). However, even data collected in naturalistic driving studies supported by the government that claim that their data is openly available, such as the SHRP2 Naturalistic Driving Study [19], present researchers with various barriers (e.g. to obtain a *qualified researcher* status) when they attempt to access the data.

In recent years, many papers presented at the AutoUI conference have addressed similar topics, such as take-over requests (TORs) [6, 20], driver distraction [8, 14] or external human-machine interfaces [3, 7]. Thus, there is a huge amount of data, both within and between individual research labs, that could potentially be combined into large, structured datasets that would allow additional analysis beyond the purpose of the work for which the data was initially collected. Most researchers are aware of the benefits of open data and open source. For example, in a recent Dagstuhl seminar on computational models of human-automated vehicle interaction, "requests for open-access and agreed-upon datasets" were intensively discussed [10]. However, many labs still do not share their data. The reasons for this are not yet well understood. Some known issues of concerns are human participant and bystander privacy, possibilities for individual identification in the age of rapidly advancing computer vision technology, variance in international regulation on data collection and distribution, compatibility of data standards, concerns of generating obligations for future support, unauthorized redistribution of data, and unanticipated collection and monetization by uninvolved third-parties.

This workshop aims to change the current silo mentality within the automotive domain when it comes to data-sharing practices. Thereby, we want to discuss whether and how such data could be integrated into a larger open data platform. Such a platform could help to accelerate research, improve data quality, encourage collaboration, promote transparency, and support accessibility. In particular, we want to discover (1) the barriers that prevent researchers from publishing data openly to the wider research community, (2) what is needed to promote data sharing in our community, (3) define a set of best practices on how data and code should be shared, and (4) define open data formats for common research topics that already exist. We, therefore, plan to set up small 'special interest groups' around recurring themes (such as the ones listed above) that specify open data formats in terms of independent and dependent variables. In particular, we want to identify small but appropriate parameter sets and develop a small-scale proof of concept that can act as a seed for a growing open data platform.

2 SCHEDULE AND ACTIVITIES

To progress towards the idea of an open-data platform in the realm of the AutoUI community, the workshop will be surrounded by pre- and post-workshop activities. The workshop itself will consist of three sessions. The tentative schedule and activities are presented in Table 1, but may be adjusted according to the results of the pre-questionnaire.

2.1 Before the Workshop

Prior to the event, all participants will be asked to complete a questionnaire that explores why many labs do not share their data and what needs to change to encourage data sharing. In this questionnaire, we want to identify the expected advantages and disadvantages of sharing experimental data in the AutoUI domain, and the reasons why this is not yet the case. In addition, we will ask all participants to anonymize, document, and prepare a dataset on an appropriate topic from their research and bring it to the workshop.

2.2 Introduction and Dataset Presentations

After an introduction of the workshop organizers and the anticipated topics and outcomes, participants will briefly introduce themselves, the data they are working with, and their experience with open data practices. In addition, some of the organizers will share their own experiences (both positive and negative) with the topic.

2.3 Group Work 1: Data Formats and Standards

The participants are then divided into different thematic groups (e.g. driver distraction, TORs, eHMIs). Group members bring their own data (sub)sets with descriptions and discuss (a) the minimum viable data content, (b) open questions on anonymization, parameter standardization (e.g. how to define measurement points, which units to use, etc.) and come up with a proposal on how to make data collected in different studies comparable. The aim of this activity is to create a unified description for the relevant topics in order to synthesize the results of different studies.

2.4 Group Work 2: Actionable Activities

The different groups work on topics related to best practices in open data that are relevant to the AutoUI community. The major topics we want to address are:

(1) Documentation and Data Description: How to create a unified description and how to make datasets comparable?
(2) Data Standards: Set standards for anonymization, quality, and processing. How to handle small and large datasets?
(3) Community Engagement: How to motivate researchers and industry partners to share their data?
(4) Platform Discussions: How to share the data? Which platforms are best suited?

2.5 Wrap-Up and Post-Workshop Activities

The post-workshop activities will build upon the results of the workshop. In particular, we want the participants of the groups that worked together to merge their datasets according to the formats that they defined in the respective session. Based on the results, we will accompany the groups and motivate them to provide a small-scale evaluation emerging from the novel possibilities. For example, the group working on TORs could perform a regression model over the entire data set and publish the results.

Table 1: Tentative schedule of activities of the workshop

Tentative Schedule		Activities
Introduction	09:00 - 9:30	Introduction of workshop topics and pre-workshop-questionnaire Introduction of organizers Self-introductions
Session 1	09:30 - 10:00	Presentation of existing efforts towards open data sharing
Coffee Break	10:00 - 10:15	-
Session 2	10:15 - 11:30	Group Work 1: Data formats and standards for individual topics
Coffee Break	11:30 - 11:45	-
Session 3	11:45 - 12:30	Group Work 2: Actionable activities to develop best practices for data sharing
Closing	12:30 - 13:00	Presentations and wrap-up

3 ATTENDANCE

We are planning for a half-day workshop and expect about 20 participants (excluding the organizers). We will reach out to researchers and practitioners who may be interested in the workshop topic (in particular attendees from industry), in addition to the regular AutoUI participants. Also, the number of participants may vary depending on the format of the workshop.

4 EXPECTED OUTCOME

In addition to publishing the results of the event on the dedicated workshop website[1], we will provide a data upload (e.g., a GitHub repository) where examplary datasets can be uploaded in the data formats identified during the workshop. We hope that our efforts will lead to a set of best practices for promoting open data in our community. Accordingly we aim to publish a white paper on best practices for data description and data sharing in AutoUI research. In the long run, this event will encourage collaboration between research labs working on similar topics, allow novel comparisons and evaluations of the collected data, and generate publications for future ACM AutoUI conferences.

5 BIOGRAPHIES

Patrick Ebel is a Junior Research Group Leader at the Center for Scalable Data Analytics and Artificial Intelligence (ScaDS.AI) at Leipzig University. He received his Ph.D. in Computer Science from the University of Cologne and his M.Sc. in Automotive Systems from the TU Berlin. His research focuses on the analysis of large naturalistic driving data and computational models for human-vehicle interaction.

Pavlo Bazilinskyy is an assistant professor at TU Eindhoven focusing on AI-driven interaction between automated vehicles and other road users. He finished his Ph.D. at TU Delft in auditory feedback for automated driving as a Marie Curie Fellow, where he also worked as a postdoc. He was the head of data research at SD-Insights. Pavlo is a treasurer of the Marie Curie Alumni Association (MCAA).

Angel Hsing-Chi Hwang is a human-computer interaction (HCI) researcher and Ph.D. candidate at Cornell University. Her research explores the role of emerging technology in the future of (team)work, culture, and social interaction. All in all, she researches, designs, and build technology that can introduce greater creativity and inclusiveness to our future workplaces.

Wendy Ju is an Associate Professor at the Jacobs Technion-Cornell Institute at Cornell Tech. Her work in the areas of human-robot interaction and automated vehicle interfaces highlights the ways that interactive systems can be designed to be safer, more predictable, and more socially appropriate.

Hauke Sandhaus is a Ph.D. student in Information Science at Cornell University. Previously, he worked as a UX Design-Technologist at the Volkswagen Group Future Center, focusing on the user experience of fully autonomous vehicles. He holds a master's degree in Human-Computer Interaction from Bauhaus University. His research integrates technology and policy design to address the ethical implications of AI-driven autonomous systems.

Aravinda Ramakrishnan Srinivasan is a research fellow at the Institute for Transport Studies, University of Leeds, UK. He received the B.Tech. degree in electronics and communication engineering from the SASTRA University, Tirumalaisamudram, Tamil Nadu, India, and the M.S. and Ph.D. degrees in mechatronics and mechanical engineering from the University of Tennessee, Knoxville, TN, USA. He was a postdoctoral research fellow at the Lincoln Centre for Autonomous Systems, University of Lincoln, UK before his current tenure at Leeds. His research interests include machine-learning, artificial intelligence, autonomous vehicles, and robotics applications in everyday life.

Qian Yang is an Assistant Professor in Information Science at Cornell University who received her Ph.D. from Carnegie Mellon University. Her main research goal is to better understand how to bring human-centered thinking to bear on evermore complex AI systems and create interactive tools that make such AI more accessible for human-centered design innovation.

Philipp Wintersberger is a Professor of Interactive Systems at the University of Applied Sciences Upper Austria (Hagenberg) and lecturer at TU Wien. He received his Ph.D. in Human-Computer Interaction from Johannes Kepler University Linz. His research focuses on the effects of safety-critical AI-controlled automated systems on user behavior.

[1] https://autouimodelingws.jimdosite.com/

REFERENCES

[1] Christopher Allen and David M. A. Mehler. 2019. Open science challenges, benefits and tips in early career and beyond. *PLOS Biology* 17, 5 (May 2019), e3000246. https://doi.org/10.1371/journal.pbio.3000246

[2] Valeria Arza and Mariano Fressoli. 2018. Systematizing benefits of open science practices. *Information Services & Use* 37, 4 (Jan. 2018), 463–474. https://doi.org/

10.3233/isu-170861

[3] Pavlo Bazilinskyy, Lars Kooijman, Dimitra Dodou, Kirsten Mallant, Victor Roosens, Marloes Middelweerd, Lucas Overbeek, and Joost de Winter. 2022. Get Out of The Way! Examining eHMIs in Critical Driver-Pedestrian Encounters in a Coupled Simulator. In *Proceedings of the 14th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*. ACM. https://doi.org/10.1145/3543174.3546849

[4] P. Bazilinskyy, S.M. Petermeijer, V. Petrovych, D. Dodou, and J.C.F. de Winter. 2018. Take-over requests in highly automated driving: A crowdsourcing survey on auditory, vibrotactile, and visual displays. *Transportation Research Part F: Traffic Psychology and Behaviour* 56 (July 2018), 82–98. https://doi.org/10.1016/j.trf.2018.04.001

[5] Helen M. Berman, Gerard J. Kleywegt, Haruki Nakamura, and John L. Markley. 2014. The Protein Data Bank archive as an open data resource. *Journal of Computer-Aided Molecular Design* 28, 10 (July 2014), 1009–1014. https://doi.org/10.1007/s10822-014-9770-y

[6] Henrik Detjen, Sarah Faltaous, Jonas Keppel, Marvin Prochazka, Uwe Gruenefeld, Shadan Sadeghian, and Stefan Schneegass. 2022. Investigating the Influence of Gaze- and Context-Adaptive Head-up Displays on Take-Over Requests. In *Proceedings of the 14th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*. ACM. https://doi.org/10.1145/3543174.3546089

[7] Debargha Dey, Azra Habibovic, Melanie Berger, Devanshi Bansal, Raymond H. Cuijpers, and Marieke Martens. 2022. Investigating the Need for Explicit Communication of Non-Yielding Intent through a Slow-Pulsing Light Band (SPLB) eHMI in AV-Pedestrian Interaction. In *Proceedings of the 14th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*. ACM. https://doi.org/10.1145/3543174.3546086

[8] Patrick Ebel, Moritz Berger, Christoph Lingenfelder, and Andreas Vogelsang. 2022. How Do Drivers Self-Regulate their Secondary Task Engagements? The Effect of Driving Automation on Touchscreen Interactions and Glance Behavior. In *Proceedings of the 14th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*. ACM. https://doi.org/10.1145/3543174.3545173

[9] Patrick Ebel, Christoph Lingenfelder, and Andreas Vogelsang. 2023. On the Forces of Driver Distraction: Explainable Predictions for the Visual Demand of in-Vehicle Touchscreen Interactions. *Accident Analysis & Prevention* 183 (April 2023), 106956. https://doi.org/10.1016/j.aap.2023.106956

[10] Christian P Janssen, Martin Baumann, Antti Oulasvirta, Shamsi Tamara Iqbal, and Luisa Heinrich. 2022. Computational Models of Human-Automated Vehicle Interaction (Dagstuhl Seminar 22102). In *Dagstuhl Reports*, Vol. 12. Schloss Dagstuhl-Leibniz-Zentrum für Informatik.

[11] Marijn Janssen, Yannis Charalabidis, and Anneke Zuiderwijk. 2012. Benefits, Adoption Barriers and Myths of Open Data and Open Government. *Information Systems Management* 29, 4 (Sept. 2012), 258–268. https://doi.org/10.1080/10580530.2012.716740

[12] Patty Kostkova, Helen Brewer, Simon de Lusignan, Edward Fottrell, Ben Goldacre, Graham Hart, Phil Koczan, Peter Knight, Corinne Marsolier, Rachel A. McKendry, Emma Ross, Angela Sasse, Ralph Sullivan, Sarah Chaytor, Olivia Stevenson, Raquel Velho, and John Tooke. 2016. Who Owns the Data? Open Data for Healthcare. *Frontiers in Public Health* 4 (Feb. 2016). https://doi.org/10.3389/fpubh.2016.00007

[13] Jennifer C. Molloy. 2011. The Open Knowledge Foundation: Open Data Means Better Science. *PLoS Biology* 9, 12 (Dec. 2011), e1001195. https://doi.org/10.1371/journal.pbio.1001195

[14] Alberto Morando, Pnina Gershon, Bruce Mehler, and Bryan Reimer. 2020. Driver-initiated Tesla Autopilot Disengagements in Naturalistic Driving. In *12th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*. ACM. https://doi.org/10.1145/3409120.3410644

[15] Marcus R. Munafò, Brian A. Nosek, Dorothy V. M. Bishop, Katherine S. Button, Christopher D. Chambers, Nathalie Percie du Sert, Uri Simonsohn, Eric-Jan Wagenmakers, Jennifer J. Ware, and John P. A. Ioannidis. 2017. A manifesto for reproducible science. *Nature Human Behaviour* 1, 1 (Jan. 2017). https://doi.org/10.1038/s41562-016-0021

[16] Michael Nielsen. 2020. *Reinventing discovery* (2 ed.). Princeton University Press, Princeton, NJ.

[17] V. Onkhar, P. Bazilinskyy, D. Dodou, and J.C.F. de Winter. 2022. The effect of drivers' eye contact on pedestrians' perceived safety. *Transportation Research Part F: Traffic Psychology and Behaviour* 84 (Jan. 2022), 194–210. https://doi.org/10.1016/j.trf.2021.10.017

[18] Aravinda Ramakrishnan Srinivasan, Julian Schumann, Yueyang Wang, Yi-Shin Lin, Michael Daly, Albert Solernou, Arkady Zgonnikov, Matteo Leonetti, Jac Billington, and Gustav Markkula. 2023. The COMMOTIONS Urban Interactions Driving Simulator Study Dataset. https://doi.org/10.48550/ARXIV.2305.11909

[19] Trent Victor, Marco Dozza, Jonas Bärgman, Christian-Nils Boda, Johan Engström, and Gustav Markkula. 2014. Analysis of Naturalistic Driving Study Data: Safer Glances, Driver Inattention, and Crash Risk.

[20] Philipp Wintersberger, Andreas Riener, Clemens Schartmüller, Anna-Katharina Frison, and Klemens Weigl. 2018. Let Me Finish before I Take Over. In *Proceedings of the 10th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*. ACM. https://doi.org/10.1145/3239060.3239085

[21] Jannik Wolf, Marco Wiedner, Mohamed Kari, and David Bethge. 2021. HMInference: Inferring Multimodal HMI Interactions in Automotive Screens. In *13th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*. ACM. https://doi.org/10.1145/3409118.3475145