Ninth Workshop on Human-In-the-Loop Data Analytics (HILDA)

Remco Chang
Tufts University
Medford, Massachusetts, USA
remco@cs.tufts.edu

Kexin Rong
Georgia Institute of Technology
Atlanta, Georgia, USA
krong@gatech.edu

Roee Shraga
Worcester Polytechnic Institute
Worcester, Massachusetts, USA
rshraga@wpi.edu

Abstract

HILDA brings together researchers and practitioners to exchange ideas and results on human-data interaction. It explores how data management and analysis can be made more effective when taking into account the people who design and build these processes as well as those who are impacted by their results. Following the past few years, we plan to continue to focus on this year's workshop on early-stage research that is promising and exciting, which includes pairing each accepted paper with a mentor. The theme for this edition of the workshop is HILDA and Large Language Models. However, the workshop is not limited to this theme and other topics are also of interest. In this summary, we describe the workshop, its main focus areas and our review and mentorship plan.

CCS Concepts

• **Information systems** → **Data management systems**; • **Human-centered computing**;

Keywords

Human-in-the-loop, Data Preparation, Human-centered Computing, Data Analytics, Information Systems, Data Visualization, Data Exploration, Data Management, Labeling, Machine Learning, Large Language Models

ACM Reference Format:
Remco Chang, Kexin Rong, and Roee Shraga. 2025. Ninth Workshop on Human-In-the-Loop Data Analytics (HILDA). In *Companion of the 2025 International Conference on Management of Data (SIGMOD-Companion '25), June 22–27, 2025, Berlin, Germany*. ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3722212.3724485

1 Introduction

Human In the Loop Data Analytics (HILDA) has been hosted at SIGMOD for the past 8 years with a hiatus in 2021. The key focus of HILDA is to evaluate, understand, and formally reason about the participation of humans in data management, with the eventual goal of building optimized data management systems and techniques that treat humans as first-class citizens, alongside data. This is an understudied area of data management, which has traditionally focused on optimizing the computational aspects of database systems, while overlooking the central role of humans. Today, the critical bottleneck in data analysis is not the lack of data, or our ability to analyze it at scale, but the lack of human cognition and time to make sense of the findings.

HILDA was created as a venue for papers that traditionally would not find a home at SIGMOD, KDD, or the more HCI-oriented conferences like CHI and VIS. Over the last years, HILDA and its sister workshops DSIA at VIS and IDEA at KDD have been instrumental in raising a new generation of researchers who are now faculty or researchers. Today, we can submit papers with performance benchmarks to HCI and visualization venues and papers with user studies to SIGMOD and find capable reviewers. As these topics gain acceptance in other venues, HILDA will emphasize mentoring and early-stage research, especially for work that does not fit easily into traditional computation-heavy SIGMOD categories.

To fulfill this vision, HILDA'25 will be a hands-on workshop with *mentoring* at the center. We introduced the mentoring program at HILDA 2022 by borrowing ideas from the successful PLATEAU workshop. Instead of submitting fleshed-out work, we value early-stage contributions and promising ideas. While submissions are still short papers, every accepted paper will get a mentor from the PC who improves the paper together with the authors. The authors are expected to incorporate mentor feedback into the camera-ready version for publication. During the workshop, we will facilitate discussions and feedback, which will include an explicit in-person role for the mentor for each paper. This hands-on format is very different from other workshops at SIGMOD, but we believe it will be more engaging and will especially support early-stage researchers.

The theme for this edition is *HILDA and Large Language Models (LLMs)*. The recent success of LLMs like OpenAI's ChatGPT underscores the critical role of human interaction in the effective use of LLMs, where users actively engage with models through prompts, validate their responses, and provide interactive feedback and examples. We encourage research on guidelines and best practices for effective human-LLM collaboration. We also encourage research that focuses on human-driven evaluation of LLMs. Human evaluation remains crucial for assessing LLM outputs, particularly for scenarios lacking clear ground truth or involving nuanced safety considerations, where automated metrics alone may be insufficient.

HILDA'25 will pursue its core goal of bringing together database researchers interested in the distinctive ways that people impact data management tasks and attracting like-minded researchers in other communities such as Human-Computer Interaction, Information Visualization, Data Mining, and Machine Learning. We hope that, through HILDA 2025, we will foster interdisciplinary efforts that tackle important challenges in human-data interaction.

2 Topics of Interest

HILDA encourages both standard research papers and more unusual works—for instance papers that describe in-progress work,

SIGMOD-Companion '25, Berlin, Germany

ACM ISBN 979-8-4007-1564-8/2025/06
https://doi.org/10.1145/3722212.3724485

reports on experiences, question accepted wisdom, raise open problems, or propose speculative new approaches. We welcome work that proposes innovations in design to improve the way people can work with data management systems, as well as work that studies empirically how humans interact with existing systems. We welcome research that comes from the traditions of the database systems community, and also reports on industry activities, and research on data topics from communities that study people and organizations. A sample of topics that are in the spirit of this workshop include, but are not limited to:

- novel query interfaces
- interactive query refinement
- data exploration and analysis
- data visualization
- human-assisted data integration and cleaning
- perception-aware data processing
- database systems designed for highly interactive use cases
- empirical studies of database use
- evaluating and ensuring fairness in data-driven decision making processes
- understanding the outcomes of processes through provenance and explanations
- interactive debugging of complex data systems
- crowd-powered data infrastructure

Submissions can also examine any of the above topics from an application or domain perspective.

3 Review Process and Mentorship

HILDA reviews are single blind. All submitted papers will be reviewed by at least three reviewers who will determine the fit of the work for HILDA's unique mentorship process this year, the quality of the work, and its potential for future research. Every accepted paper will be assigned a mentor who will engage with the authors providing constructive feedback through one-on-one discussion. We expect that the authors will work closely with their mentors to improve the substance and direction of their work. We will encourage the authors to integrate feedback from discussion with the mentor in the camera-ready submission. Authors and mentors can withdraw without repercussions due to unforeseen conflicts. In such situations, the program chairs will try to find another suitable mentor. All accepted papers will have the opportunity to present the work during the workshop.

The mentors of HILDA'25 are:

- Aamod Khatiwada (Northeastern University)
- Andra-Denis Somech (Delft University of Technology)
- Avigdor Gal (Technion - Israel Institute of Technology)
- Bar Genossar (Technion - Israel Institute of Technology)
- Brit Youngmann (Technion - Israel Institute of Technology)
- Danyel Fisher (Independent Consultant)
- Fatemeh Nargesian (University of Rochester)
- Gerardo Vitagliano (MIT CSAIL)
- Grace Fan (Northeastern)
- Jean-Daniel Fekete (Inria)
- Kanit Wongsuphasawat (Databricks)
- Kaustav Bhattacharjee (New Jersey Institute of Technology)
- Madelon Hulsebos (Centrum Wiskunde & Informatica)
- Sainyam Galhotra (Cornell University)
- Slava Novgorodov (Tel Aviv University)
- Stavros Sintos (University of Illinois Chicago)
- Tiziana Catarci (University of Rome "La Sapienza")
- Yash Govind (Informatica)
- Zhengjie Miao (Simon Fraser University)

4 Keynotes

Our workshop will feature two invited keynotes (30 min + Q/A time) from leaders in the field, who will share insights on the challenges of human-data interaction in data analysis:

- Arnab Nandi (The Ohio State University)
- Eugene Wu (Columbia University)

Eugene served as chair of the workshop in 2018 and his research at the intersection of data management and HCI remains extremely relevant for the HILDA community. Arnab chaired the workshop in 2019 and currently serves on its steering committee. Arnab will open the event with a special keynote celebrating 10 Years of HILDA, reflecting on the workshop's legacy and future directions.

5 Organizers and Steering Committee

The organizers (chairs) of HILDA'25 are:

- Remco Chang (Tufts University)
- Kexin Rong (Georgia Institute of Technology)
- Roee Shraga (Worcester Polytechnic Institute)

The steering committee of the HILDA workshop are:

- Carsten Binnig (TU Darmstadt)
- Juliana Freire (New York University)
- Arnab Nandi (The Ohio State University)
- Joseph M. Hellerstein (University of California, Berkeley)
- Aditya Parameswaran (University of California, Berkeley)