Researching Responsible Natural Language Processing
Week 4 – 5th October 2025

What is a ~~design~~ HCI paper?

John Vines
Professor of Design Informatics
john.vines@ed.ac.uk

THE UNIVERSITY *of* EDINBURGH

design informatics

Designing Responsible Natural Language Processing

Outline of session

- Brief introduction to HCI as a design and research field
- Hear from you all about what papers you found
- Different types of research contributions in HCI

What is a design paper?

~~What is a design paper?~~

What is a HCI paper?

HCI as a “design” field

“Human–computer interaction is a discipline concerned with the **design**, evaluation and implementation of interactive computing systems for human use and with the study of major phenomena surrounding them.”

Hewett, et al. 1992. *ACM SIGCHI curricula for human-computer interaction*.

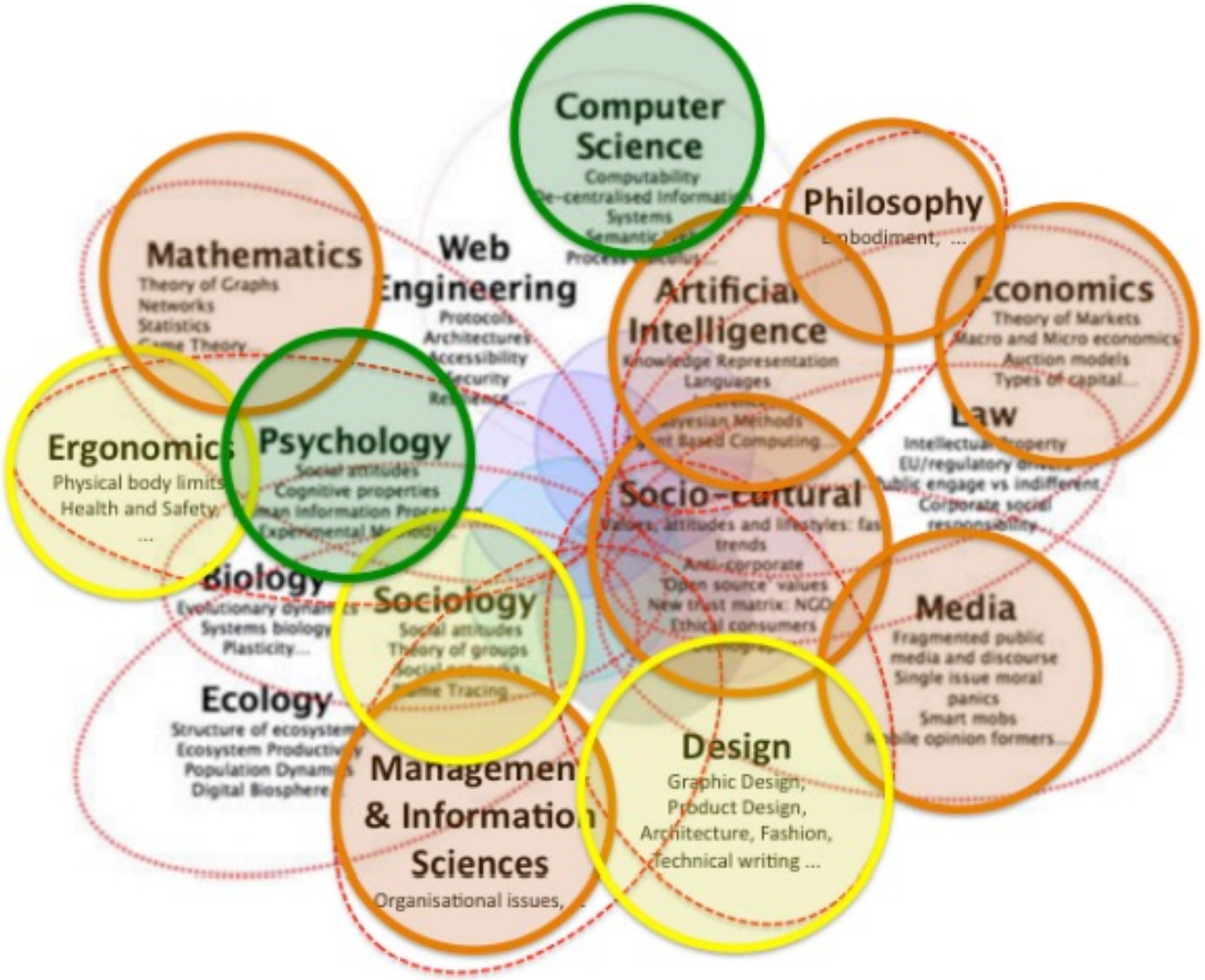
HCI as a “design” field

HCI as a term sometimes gets used to describe the phenomena of humans interacting with computers – it’s a thing that people do

But it’s also an interdisciplinary field, that involves “design” in a broadest possible sense….

- HCI researchers study groups of people to understand how technology might support them better, drawing out implications or directions for technology design
- HCI researchers study groups of people to understand how technology is currently used by them, how they struggle with it, what works well and what doesn’t – drawing out implications for (re)designing technology, and developing good or bad practices
- HCI researchers experiment with designing new ways of interacting with or using technologies – building novel prototypes with new interaction modes, and giving these to people to test them out
- HCI researchers come up with methods and techniques that enable “non-designers” to input into design processes - something we call co-design, or participatory design..

HCI as a field of research



Started off as the intermingling of computer science and cognitive science and cognitive psychology ….

… then started to draw more on social psychology, social science and sociology …

… and now connects many different disciplines together

Credit for diagram: Ruth Stalker-Firth, 2018.

The “Waves” and “Paradigms” of HCI

First Wave
1970s – 1990s

Expert users in workplaces

Focused on one-to-one interactions between a “user” and a “machine”

Heavily influenced by cognitive science and cognitive ergonomics

Lab-based studies

Modelling human behaviour

Bødker, S. 2006. When second wave HCI meets third wave challenges.

The “Waves” and “Paradigms” of HCI

First Wave 1970s – 1990s	Second Wave 1990s – 2000s
Expert users in workplaces	Many users in workplaces
Focused on one-to-one interactions between a “user” and a “machine”	Focused on groups of people working with a range of applications to get jobs done
Heavily influenced by cognitive science and cognitive ergonomics	Still a lot of cognitive science, but also influence of social psychology and ethnography
Lab-based studies	Human studies in contextual settings and understood as practices
Modelling human behaviour	

Bødker, S. 2006. When second wave HCI meets third wave challenges.

The “Waves” and “Paradigms” of HCI

First Wave
1970s – 1990s

Expert users in workplaces

Focused on one-to-one interactions between a “user” and a “machine”

Heavily influenced by cognitive science and cognitive ergonomics

Lab-based studies

Modelling human behaviour

Second Wave
1990s – 2000s

Many users in workplaces

Focused on groups of people working with a range of applications to get jobs done

Still a lot of cognitive science, but also influence of social psychology and ethnography

Human studies in contextual settings and understood as practices

Third Wave
2000s -

Technology in homes, in culture, in leisure, in everyday settings

Technology no longer just desktop computers – different modes of interaction, moving between contexts

Focus on understanding experience and emotion rather than just usability and tasks

More influence of ethnography, cultural theories, design research

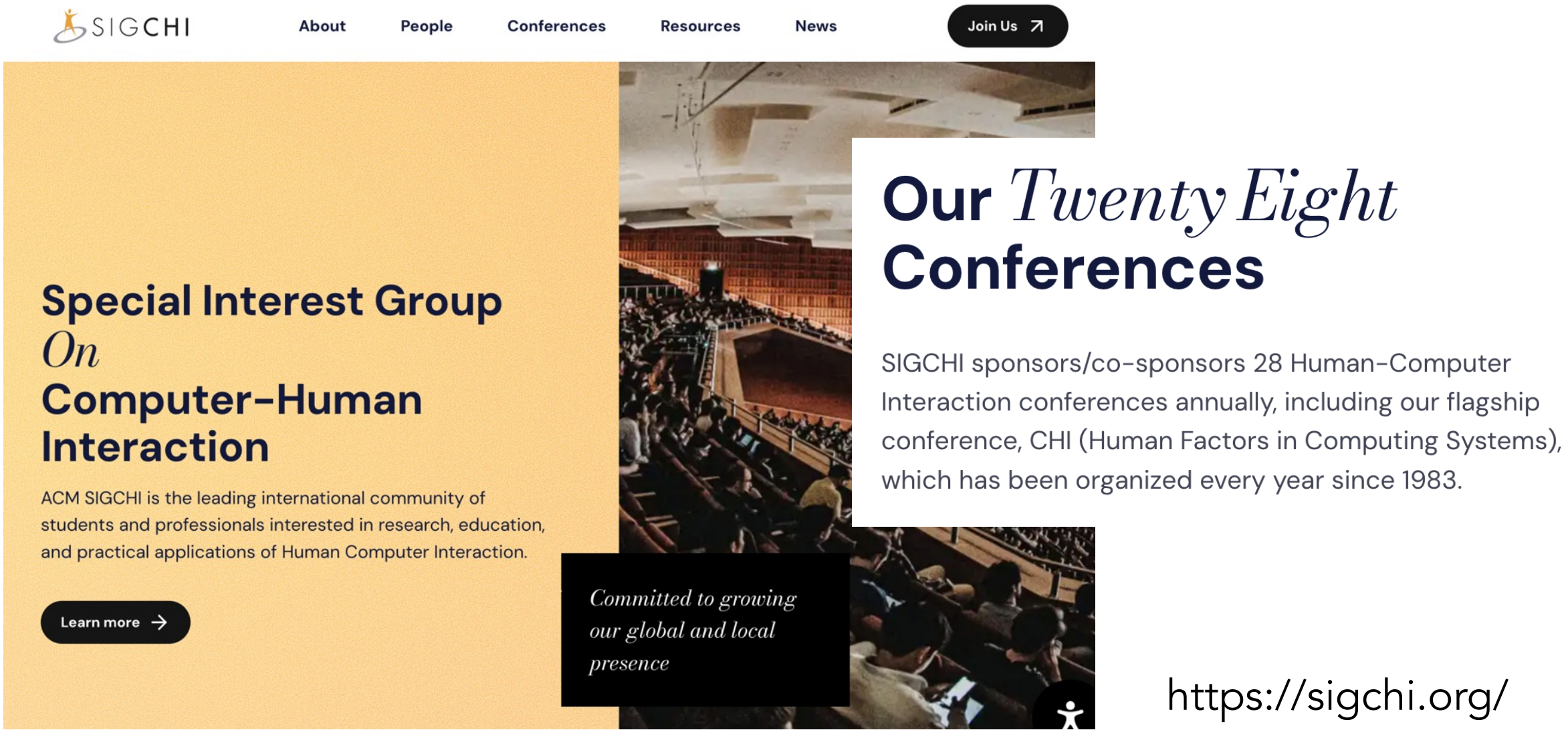
Bødker, S. 2006. When second wave HCI meets third wave challenges.

The “Waves” and “Paradigms” of HCI

	Paradigm 1	Paradigm 2	Paradigm 3
Metaphor of interaction	Interaction as man-machine coupling	Interaction as information communication	Interaction as phenomenologically situated
Central goal for interaction	Optimizing fit between man and machine	Optimizing accuracy and efficiency of information transfer	Support for situated action in the world
Typical questions of interest	How can we fix specific problems that arise in interaction?	▪ What mismatches come up in communication between computers and people? ▪ How can we accurately model what people do? ▪ How can we improve the efficiency of computer use?	▪ What existing situated activities in the world should we support? ▪ How do users appropriate technologies, and how can we support those appropriations? ▪ How can we support interaction without constraining it too strongly by what a computer can do or understand? ▪ What are the politics and values at the site of interaction, and how can we support those in design?

Harrison et al. 2007. The three paradigms of HCI.

Where to find HCI research - SIGCHI



https://sigchi.org/

Where to find HCI research – SIGCHI Conferences

CHI

ACM Conference on Human Factors in Computing Systems

CSCW

ACM Conference On Computer-Supported Cooperative Work And Social Computing

CUI

ACM Conversational User Interfaces

DIS

ACM Designing Interactive Systems

IUI

ACM Conference on Intelligent User Interfaces

UIST

ACM Symposium on User Interface Software and Technology

UMAP

ACM Conference on User Modeling, Adaptation and Personalization

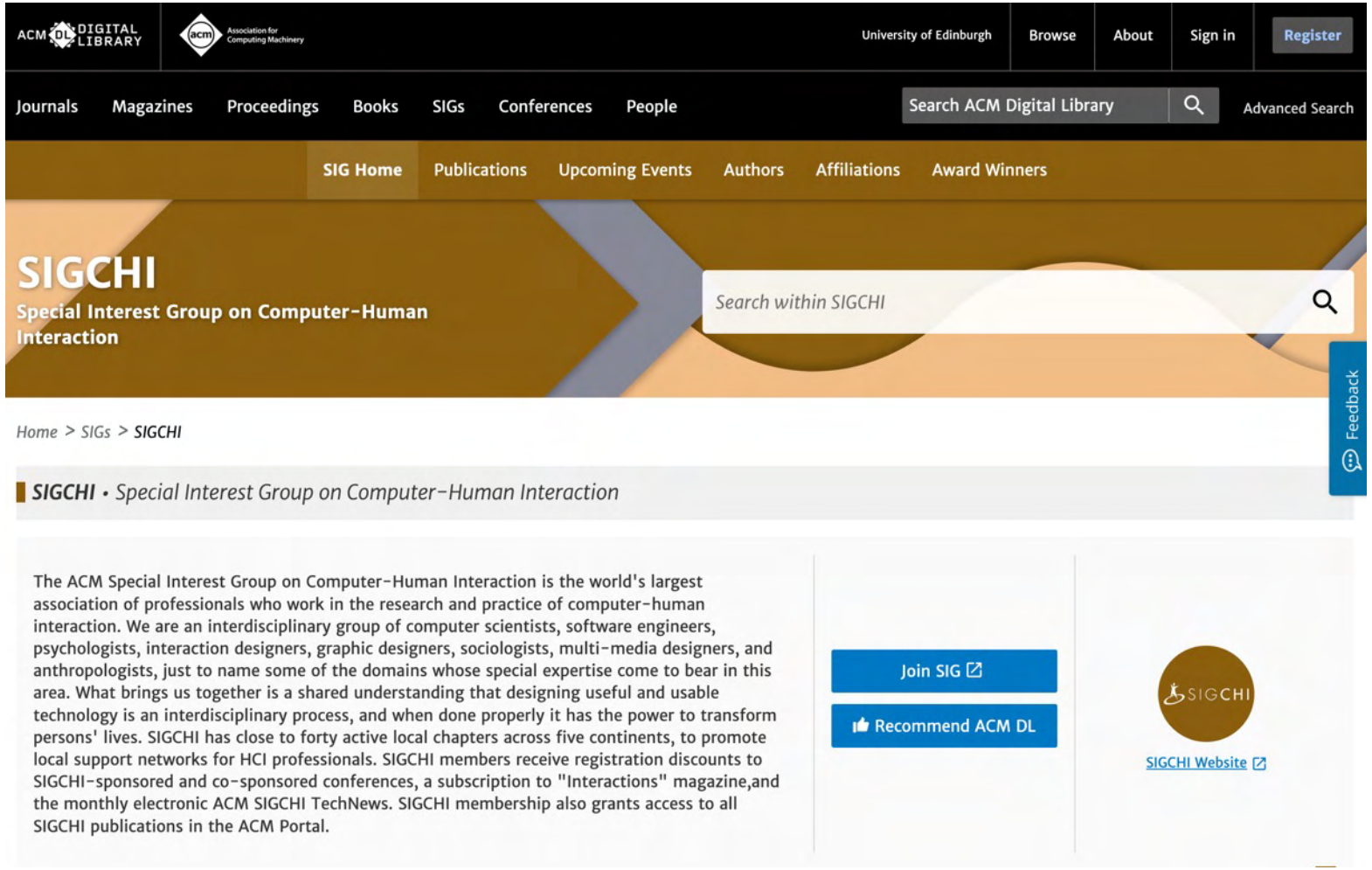
CI

ACM Collective Intelligence Conference

COMPASS

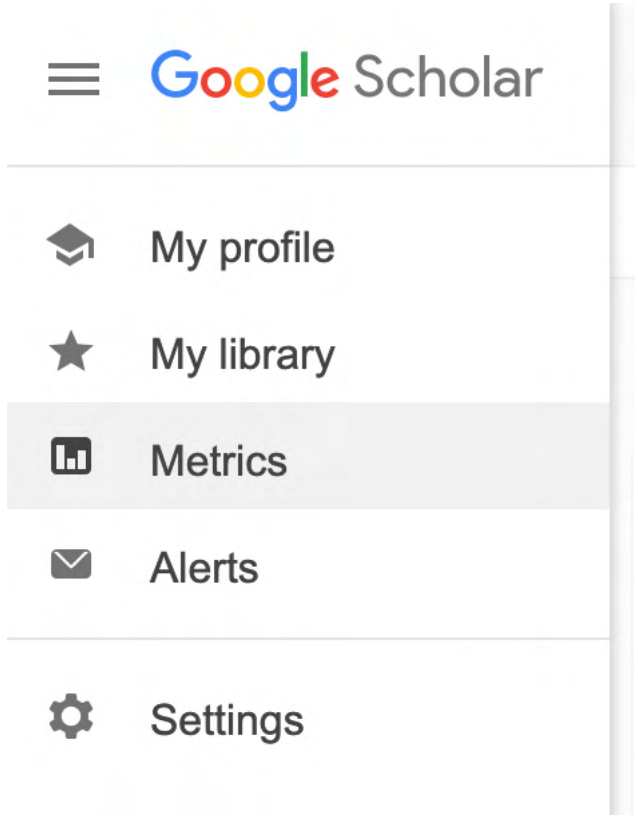
ACM SIGCAS/ SIGCHI Conference on Computing and Sustainable Societies

Where to find HCI research – SIGCHI on ACM DL



https://dl.acm.org/sig/sigchi

Where to find HCI research beyond SIGCHI!



Categories > Engineering & Computer Science > **Human Computer Interaction**

	Publication	h5-index	h5-median
1.	Computer Human Interaction (CHI)	139	185
2.	Proceedings of the ACM on Human-Computer Interaction	88	138
3.	International Journal of Human-Computer Interaction	83	124
4.	Behaviour & Information Technology	73	116
5.	IEEE Transactions on Affective Computing	72	115
6.	International Journal of Human-Computer Studies	64	104
7.	Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies	64	95
8.	Virtual Reality	62	115
9.	International Journal of Interactive Mobile Technologies	60	86
10.	ACM Transactions on Computer-Human Interaction	55	90
11.	ACM Designing Interactive Systems Conference	54	71
12.	ACM Symposium on User Interface Software and Technology	53	79
13.	ACM/IEEE International Conference on Human Robot Interaction	52	78
14.	Frontiers in Virtual Reality	52	76
15.	IEEE Virtual Reality Conference	52	66
16.	International Conference on Intelligent User Interfaces (IUI)	50	90
17.	Universal Access in the Information Society	48	74
18.	IEEE Transactions on Human-Machine Systems	45	71
19.	HCI International	45	64
20.	Multimodal Technologies and Interaction	42	62

https://scholar.google.co.uk/citations?view_op=top_venues&hl=en&vq=eng_humancomputerinteraction

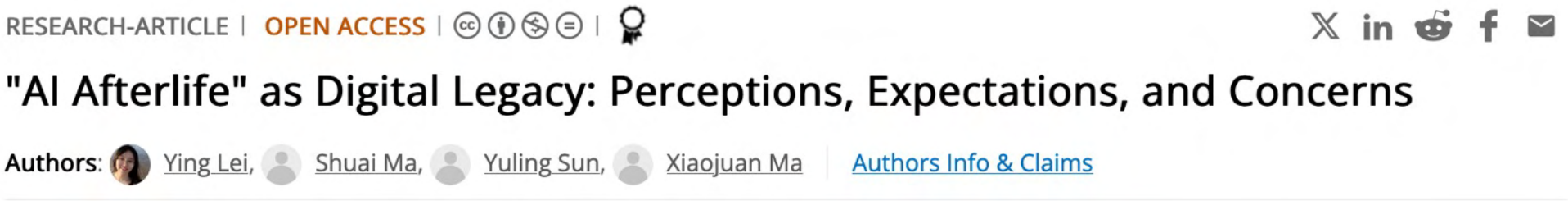
Over to you all …. 2 mins each!

Tell us…

- How did you find exploring the CHI 2025 proceedings?
- Was it challenging, or easy, or … ? …. Finding a paper for you and your team mate?
- Tell us about the papers you found for you and your team mate – why did you choose them?

Alex!

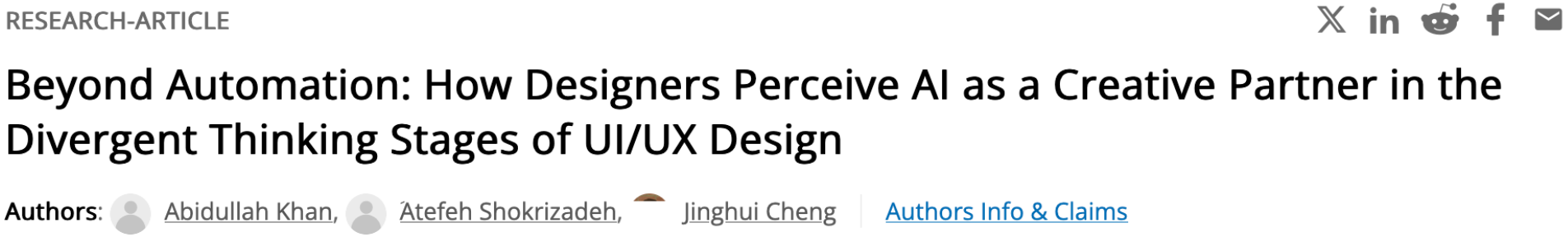
For me...


RESEARCH-ARTICLE | OPEN ACCESS |

"AI Afterlife" as Digital Legacy: Perceptions, Expectations, and Concerns

Authors: Ying Lei, Shuai Ma, Yuling Sun, Xiaojuan Ma | Authors Info & Claims

For Benoît...


RESEARCH-ARTICLE

Beyond Automation: How Designers Perceive AI as a Creative Partner in the Divergent Thinking Stages of UI/UX Design

Authors: Abidullah Khan, Atefeh Shokrizadeh, Jinghui Cheng | Authors Info & Claims

Benoît!

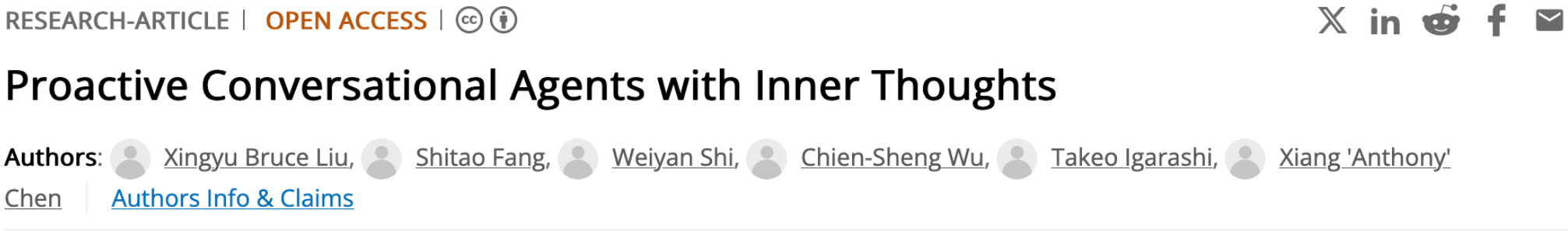
For me...

RESEARCH-ARTICLE | OPEN ACCESS |

AI, Help Me Think—but for Myself: Assisting People in Complex Decision-Making by Providing Different Kinds of Cognitive Support

Authors: Leon Reicherts, Zelun Tony Zhang, Elisabeth von Oswald, Yuanting Liu, Yvonne Rogers, Mariam Hassib | Authors Info & Claims

For Dayyán...



RESEARCH-ARTICLE | OPEN ACCESS |

Proactive Conversational Agents with Inner Thoughts

Authors: Xingyu Bruce Liu, Shitao Fang, Weiyan Shi, Chien-Sheng Wu, Takeo Igarashi, Xiang 'Anthony' Chen | Authors Info & Claims

Dayyán!

For me...

RESEARCH-ARTICLE | OPEN ACCESS |

Finding the Conversation: A Method for Scoring Documents for Natural Conversation Content

Authors: Robert John Moore, Sungeun An, Jay Pankaj Gala, Divyesh Jadav | Authors Info & Claims

For Alex...

RESEARCH-ARTICLE | OPEN ACCESS |

Digital Legacy Systems for Young Adults: Emphasizing Relationship-Oriented Perspectives and Physical Artifacts in Death Preparation

Authors: Soonho Kwon, Hyunah Jo, Sohee Ryu, Jihwan Ryan Do, HwaJung Lee, JooHyun Lee, Keeheon Lee, Younah Kang | Authors Info & Claims

Riyadh!

For me...

RESEARCH-ARTICLE | OPEN ACCESS |

Dango: A Mixed-Initiative Data Wrangling System using Large Language Model

Authors: Wei-Hao Chen, Weixi Tong, Amanda Case, Ph.D., Tianyi Zhang | Authors Info & Claims

For Cyril...

RESEARCH-ARTICLE | OPEN ACCESS |

Deceptive Explanations by Large Language Models Lead People to Change their Beliefs About Misinformation More Often than Honest Explanations

Authors: Valdemar Danry, Pat Pataranutaporn, Matthew Groh, Ziv Epstein | Authors Info & Claims

Cyril!

For me...

RESEARCH-ARTICLE | OPEN ACCESS |

Perceptions of Sentient AI and Other Digital Minds: Evidence from the AI, Morality, and Sentience (AIMS) Survey

Authors:  Jacy Reese Anthis, Janet V.T. Pauketat, Ali Ladak, Aikaterina Manoli | Authors Info & Claims

For Daisy...

RESEARCH-ARTICLE | OPEN ACCESS |

Towards AI-driven Sign Language Generation with Non-manual Markers

Authors: Han Zhang, Rotem Shalev-Arkushin, Vasileios Baltatzis, Connor Gillis,  Gierad Laput,  Raja Kushalnagar, Lorna C Quandt, Leah Findlater, Abdelkareem Bedri, Colin Lea | Authors Info & Claims

Daisy!

For me...

RESEARCH-ARTICLE | OPEN ACCESS |

Exploring Reduced Feature Sets for American Sign Language Dictionaries

Authors: Ben Kosa, Aashaka Desai, Alex X Lu,  Richard E. Ladner, Danielle Bragg | Authors Info & Claims

For Harvey...

RESEARCH-ARTICLE | OPEN ACCESS |

Unknown Word Detection for English as a Second Language (ESL) Learners using Gaze and Pre-trained Language Models

Authors: Jiexin Ding, Bowen Zhao,  Yuntao Wang, Xinyun Liu, Rui Hao,  Ishan Chatterjee, Yuanchun Shi | Authors Info & Claims

Harvey!

For me...

RESEARCH-ARTICLE | OPEN ACCESS |

Looking but Not Focusing: Defining Gaze-Based Indices of Attention Lapses and Classifying Attentional States

Authors: Eugene Hwang, Jeongmi Lee | Authors Info & Claims

For ?...

RESEARCH-ARTICLE |

Explanatory Debiasing: Involving Domain Experts in the Data Generation Process to Mitigate Representation Bias in AI Systems

Authors: Aditya Bhattacharya, Simone Stumpf, Robin De Croon, Katrien Verbert | Authors Info & Claims

Alessandra!

For me...

RESEARCH-ARTICLE | OPEN ACCESS |

Lost in Moderation: How Commercial Content Moderation APIs Over- and Under-Moderate Group-Targeted Hate Speech and Linguistic Variations

Authors: David Hartmann, Amin Oueslati, Dimitri Staufer, Lena Pohlmann, Simon Munzert, Hendrik Heuer | Authors Info & Claims

For Morgan...

RESEARCH-ARTICLE | OPEN ACCESS |

Towards Effective Human Intervention in Algorithmic Decision-Making: Understanding the Effect of Decision-Makers' Configuration on Decision-Subjects' Fairness Perceptions

Authors: Mireia Yurrita, Himanshu Verma, Agathe Balayn, Ujwal Gadiraju, Sylvia C. Pont, Alessandro Bozzon | Authors Info & Claims

Morgan!

For me...

RESEARCH-ARTICLE

Towards Human-AI Deliberation: Design and Evaluation of LLM-Empowered Deliberative AI for AI-Assisted Decision-Making

Authors: Shuai Ma, Qiaoyi Chen, Xinru Wang, Chengbo Zheng, Zhenhui Peng, Ming Yin, Xiaojuan Ma | Authors Info & Claims

For Yintao...

RESEARCH-ARTICLE

Assistance or Disruption? Exploring and Evaluating the Design and Trade-offs of Proactive AI Programming Support

Authors: Kevin Pu, Daniel Lazaro, Ian Arawjo,  Haijun Xia, Ziang Xiao,  Tovi Grossman, Yan Chen | Authors Info & Claims

Yintao!

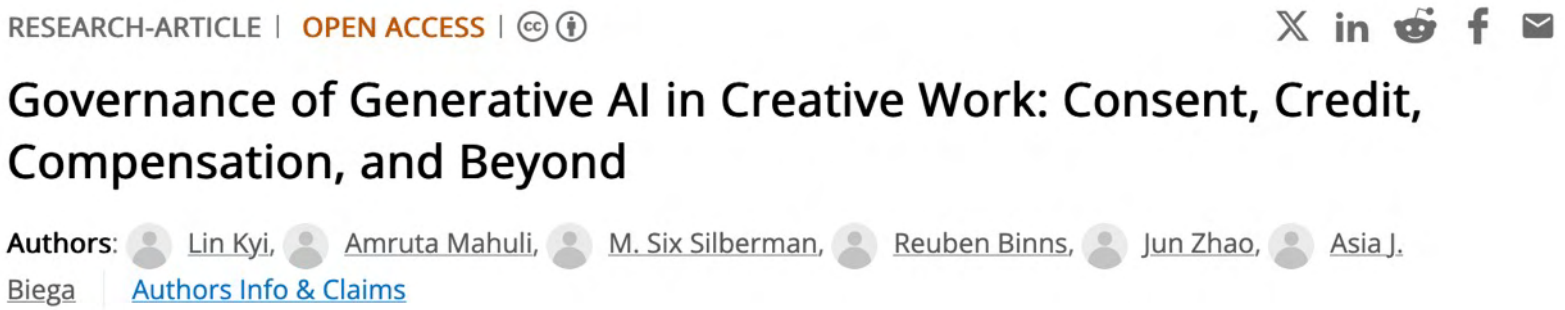
For me...

RESEARCH-ARTICLE | OPEN ACCESS |

Deceptive Explanations by Large Language Models Lead People to Change their Beliefs About Misinformation More Often than Honest Explanations

Authors: Valdemar Danry, Pat Pataranutaporn, Matthew Groh, Ziv Epstein | Authors Info & Claims

For Morgan ... ?



RESEARCH-ARTICLE | OPEN ACCESS |

Governance of Generative AI in Creative Work: Consent, Credit, Compensation, and Beyond

Authors: Lin Kyi, Amruta Mahuli, M. Six Silberman, Reuben Binns, Jun Zhao, Asia J. Biega | Authors Info & Claims

Sam!

For me...

RESEARCH-ARTICLE | OPEN ACCESS |

Fostering Appropriate Reliance on Large Language Models: The Role of Explanations, Sources, and Inconsistencies

Authors: Sunnie S. Y. Kim, Jennifer Wortman Vaughan, Q. Vera Liao, Tania Lombrozo, Olga Russakovsky | Authors Info & Claims

For JJ...

RESEARCH-ARTICLE | OPEN ACCESS |

Piecing Together Teamwork: A Responsible Approach to an LLM-based Educational Jigsaw Agent

Authors: Emily Doherty, E. Margaret Perkoff, Sean von Bayern, Rui Zhang, Indrani Dey, Michal Bodzianowski, Sadhana Puntambekar, Leanne Hirshfield | Authors Info & Claims

Billy!

For me...

RESEARCH-ARTICLE | OPEN ACCESS |

Understanding Adolescents' Perceptions of Benefits and Risks in Health AI Technologies through Design Fiction

Authors: Jamie Lee, Kyuha Jung, Erin Gregg Newman, Emilie Chow, Yunan Chen | Authors Info & Claims

For Sam...

RESEARCH-ARTICLE | OPEN ACCESS |

Playing Dumb to Get Smart: Creating and Evaluating an LLM-based Teachable Agent within University Computer Science Classes

Authors: Kantwon Rogers, Michael Davis, Mallesh Maharana, Pete Etheredge, Sonia Chernova | Authors Info & Claims

Mugdha!

For me...

RESEARCH-ARTICLE | OPEN ACCESS |

Lost in Moderation: How Commercial Content Moderation APIs Over- and Under-Moderate Group-Targeted Hate Speech and Linguistic Variations

Authors: David Hartmann, Amin Oueslati, Dimitri Staufer, Lena Pohlmann, Simon Munzert, Hendrik Heuer | Authors Info & Claims

For Billy...

RESEARCH-ARTICLE | OPEN ACCESS |

ProxiCycle: Passively Mapping Cyclist Safety Using Smart Handlebars for Near-Miss Detection

Authors: Joseph Breda, Keyu Chen, Thomas Plötz, Shwetak Patel | Authors Info & Claims

JJ!

For me...

For Mugdha ...

Break time – 5 mins

Types of research contributions in HCI

Jacob O. Wobbrock, University of Washington
Julie A. Kientz, University of Washington

Research Contributions in Human-Computer Interaction

Insights

- Knowledge generated by HCI research can be categorized into certain contribution types.
- Each contribution type has key characteristics that imply how it is judged.
- The contribution types used for submissions to the CHI conference have evolved over time to distill types of knowledge from other concerns.

IMAGE BY VARINTHORN MEKAARUN

All scholarly fields strive to contribute new knowledge. In the field of human-computer interaction (HCI), this new knowledge increasingly comes in rich forms like videos and demos, but the archival research paper remains the most widely used and accepted capture and delivery mechanism for research knowledge. The knowledge contribution made by a research paper—or more precisely, made by the work a research paper describes—is any research paper's central feature. For example, a theoretical physics paper may contribute a new mathematical model for the behavior of light near black holes. A civil engineering paper may contribute a new method for stress-testing bridges. A social anthropology paper may contribute an account of people's reactions to teen pregnancies in rural religious communities. Whatever the field of inquiry, whatever the phenomenon of interest, every research paper strives to make a research contribution by offering new knowledge. In an effort to distinguish this kind of knowledge from everyday know-how, some scholars even capitalize the term: Knowledge.

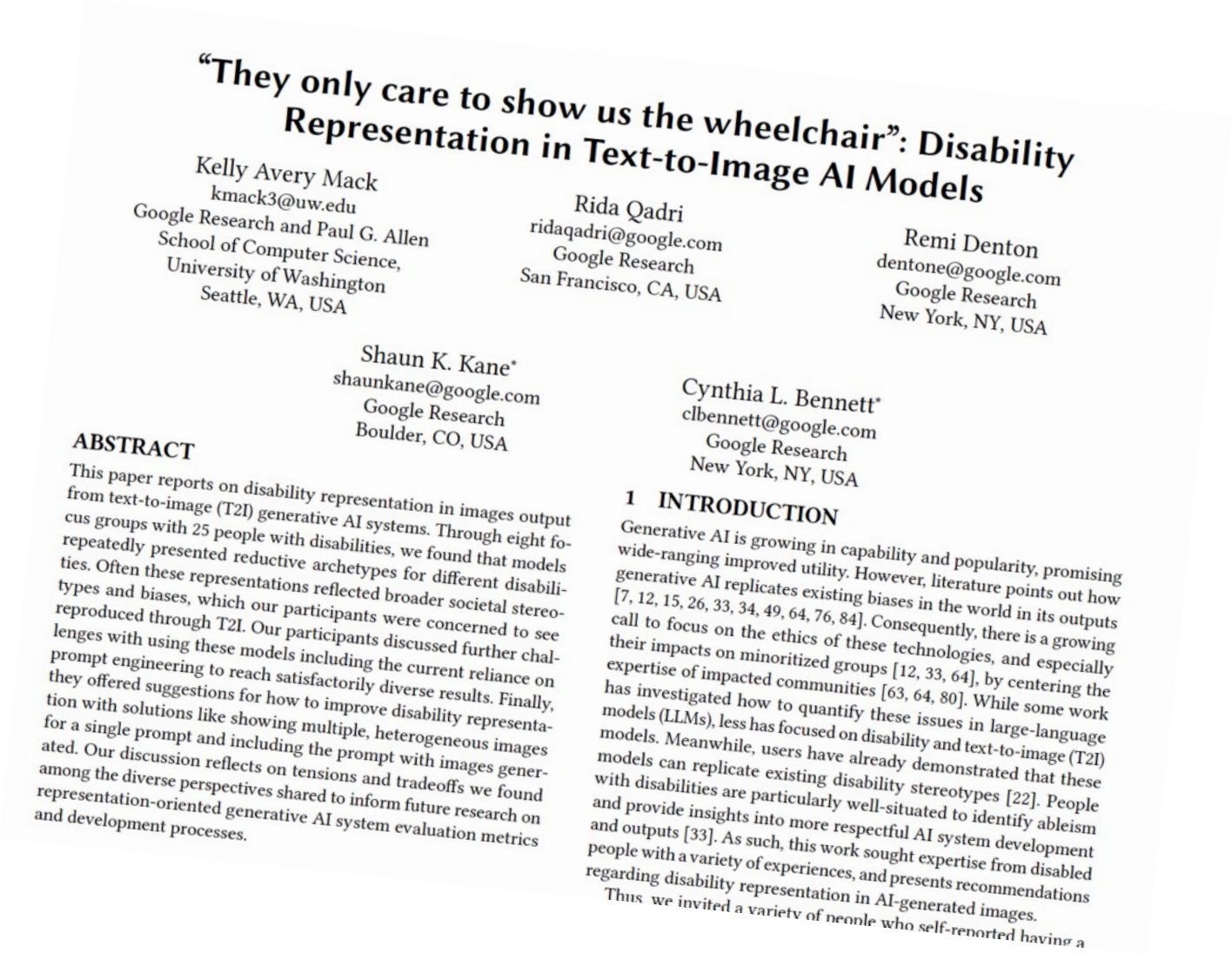
In the whole of human inquiry, there are, of course, countless specific research contributions to be made. But

Empirical Research

Studies that give insights about the design and use of technologies based on observation and data gathering.

"*In HCI, empirical contributions arise from a variety of sources, including experiments, user tests, field observations, interviews, surveys, focus groups, diaries, ethnographies, sensors, log files, and many others.*" *(p.40)*

Contributions are evaluated based on the level of insights they provide over prior work (e.g., deepened understanding of a user group, studying a novel context), and on the soundness of methods of data gathering and analysis.



"They only care to show us the wheelchair": Disability Representation in Text-to-Image AI Models

Kelly Avery Mack
kmack3@uw.edu
Google Research and Paul G. Allen School of Computer Science, University of Washington
Seattle, WA, USA

Rida Qadri
ridaqadri@google.com
Google Research
San Francisco, CA, USA

Remi Denton
dentone@google.com
Google Research
New York, NY, USA

Shaun K. Kane*
shaunkane@google.com
Google Research
Boulder, CO, USA

Cynthia L. Bennett*
clbennett@google.com
Google Research
New York, NY, USA

ABSTRACT

This paper reports on disability representation in images output from text-to-image (T2I) generative AI systems. Through eight focus groups with 25 people with disabilities, we found that models repeatedly presented reductive archetypes for different disabilities. Often these representations reflected broader societal stereotypes and biases, which our participants were concerned to see reproduced through T2I. Our participants discussed further challenges with using these models including the current reliance on prompt engineering to reach satisfactorily diverse results. Finally, they offered suggestions for how to improve disability representation with solutions like showing multiple, heterogeneous images for a single prompt and including the prompt with images generated. Our discussion reflects on tensions and tradeoffs we found among the diverse perspectives shared to inform future research on representation-oriented generative AI system evaluation metrics and development processes.

1 INTRODUCTION

Generative AI is growing in capability and popularity, promising wide-ranging improved utility. However, literature points out how generative AI replicates existing biases in the world in its outputs [7, 12, 15, 26, 33, 34, 49, 64, 76, 84]. Consequently, there is a growing call to focus on the ethics of these technologies, and especially their impacts on minoritized groups [12, 33, 64], by centering the expertise of impacted communities [63, 64, 80]. While some work has investigated how to quantify these issues in large-language models (LLMs), less has focused on disability and text-to-image (T2I) models. Meanwhile, users have already demonstrated that these models can replicate existing disability stereotypes [22]. People with disabilities are particularly well-situated to identify ableism and provide insights into more respectful AI system development and outputs [33]. As such, this work sought expertise from disabled people with a variety of experiences, and presents recommendations regarding disability representation in AI-generated images.

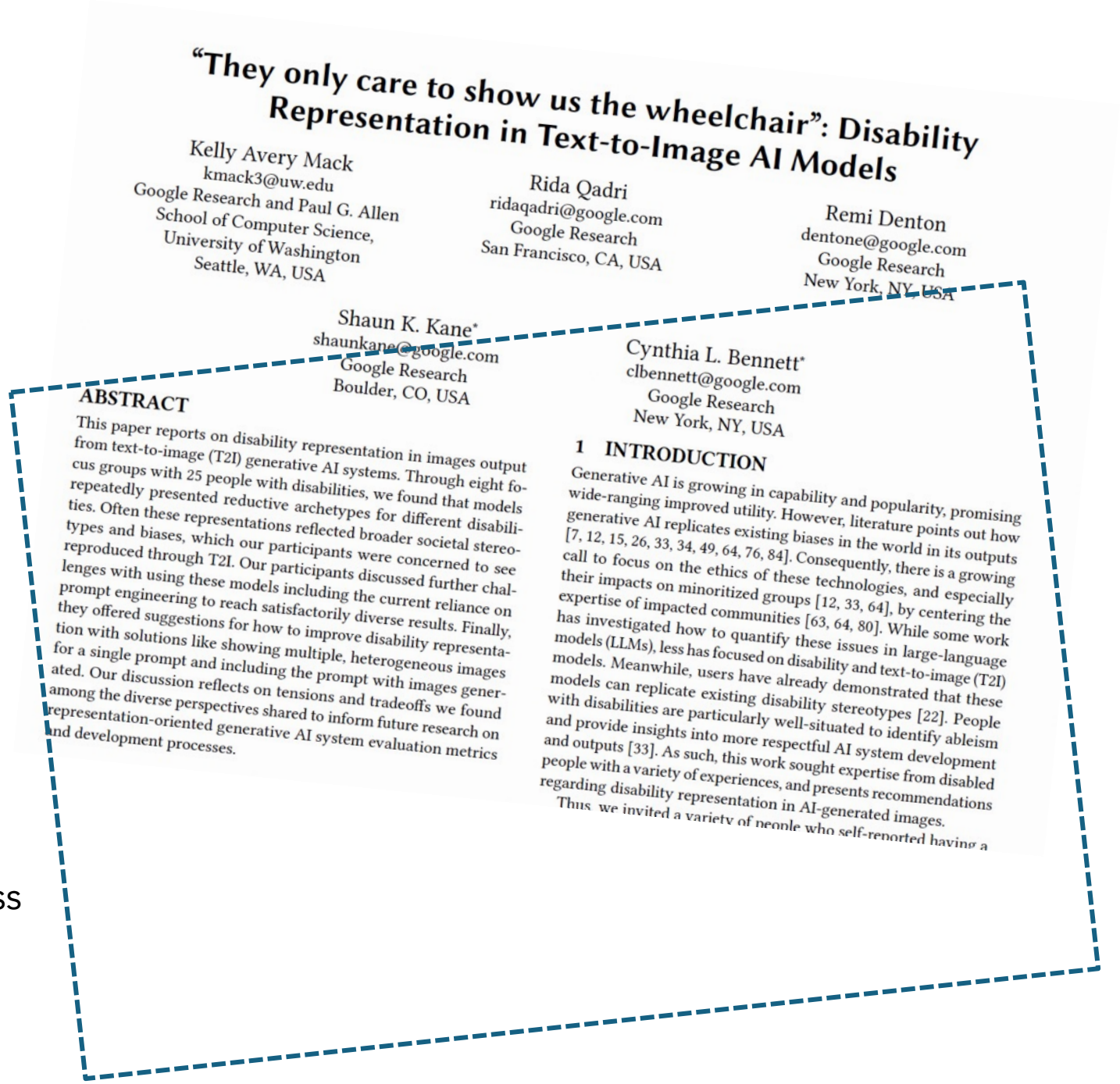
Thus, we invited a variety of people who self-reported having a

Empirical Research

Studies that give insights about the design and use of technologies based on observation and data gathering.

"*In HCI, empirical contributions arise from a variety of sources, including experiments, user tests, field observations, interviews, surveys, focus groups, diaries, ethnographies, sensors, log files, and many others.*" *(p.40)*

Contributions are evaluated based on the level of insights they provide over prior work (e.g., deepened understanding of a user group, studying a novel context), and on the soundness of methods of data gathering and analysis.



"They only care to show us the wheelchair": Disability Representation in Text-to-Image AI Models

Kelly Avery Mack
kmack3@uw.edu
Google Research and Paul G. Allen School of Computer Science, University of Washington
Seattle, WA, USA

Rida Qadri
ridaqadri@google.com
Google Research
San Francisco, CA, USA

Remi Denton
dentone@google.com
Google Research
New York, NY, USA

Shaun K. Kane*
shaunkane@google.com
Google Research
Boulder, CO, USA

Cynthia L. Bennett*
clbennett@google.com
Google Research
New York, NY, USA

ABSTRACT

This paper reports on disability representation in images output from text-to-image (T2I) generative AI systems. Through eight focus groups with 25 people with disabilities, we found that models repeatedly presented reductive archetypes for different disabilities. Often these representations reflected broader societal stereotypes and biases, which our participants were concerned to see reproduced through T2I. Our participants discussed further challenges with using these models including the current reliance on prompt engineering to reach satisfactorily diverse results. Finally, they offered suggestions for how to improve disability representation with solutions like showing multiple, heterogeneous images for a single prompt and including the prompt with images generated. Our discussion reflects on tensions and tradeoffs we found among the diverse perspectives shared to inform future research on representation-oriented generative AI system evaluation metrics and development processes.

1 INTRODUCTION

Generative AI is growing in capability and popularity, promising wide-ranging improved utility. However, literature points out how generative AI replicates existing biases in the world in its outputs [7, 12, 15, 26, 33, 34, 49, 64, 76, 84]. Consequently, there is a growing call to focus on the ethics of these technologies, and especially their impacts on minoritized groups [12, 33, 64], by centering the expertise of impacted communities [63, 64, 80]. While some work has investigated how to quantify these issues in large-language models (LLMs), less has focused on disability and text-to-image (T2I) models. Meanwhile, users have already demonstrated that these models can replicate existing disability stereotypes [22]. People with disabilities are particularly well-situated to identify ableism and provide insights into more respectful AI system development and outputs [33]. As such, this work sought expertise from disabled people with a variety of experiences, and presents recommendations regarding disability representation in AI-generated images.
Thus, we invited a variety of people who self-reported having a

Artefacts

"Artifacts, often prototypes, include new systems, architectures, tools, toolkits, techniques, sketches, mockups, and envisionments that reveal new possibilities, enable new explorations, facilitate new insights, or compel us to consider new possible futures. New knowledge is embedded in and manifested by artifacts and the supporting materials that describe them." (p40)

Artefact contributions might be evaluated as part of empirical research. But they do not always need to be – well documented, justified and described artefacts can be a contribution in themselves (see e.g., Gaver and Hook for a discussion!)

Artefacts

"Artifacts, often prototypes, include new systems, architectures, tools, toolkits, techniques, sketches, mockups, and envisionments that reveal new possibilities, enable new explorations, facilitate new insights or compel us to consider new possible futures. New knowledge is embedded in and manifested by artifacts and the supporting materials that describe them." (p40)

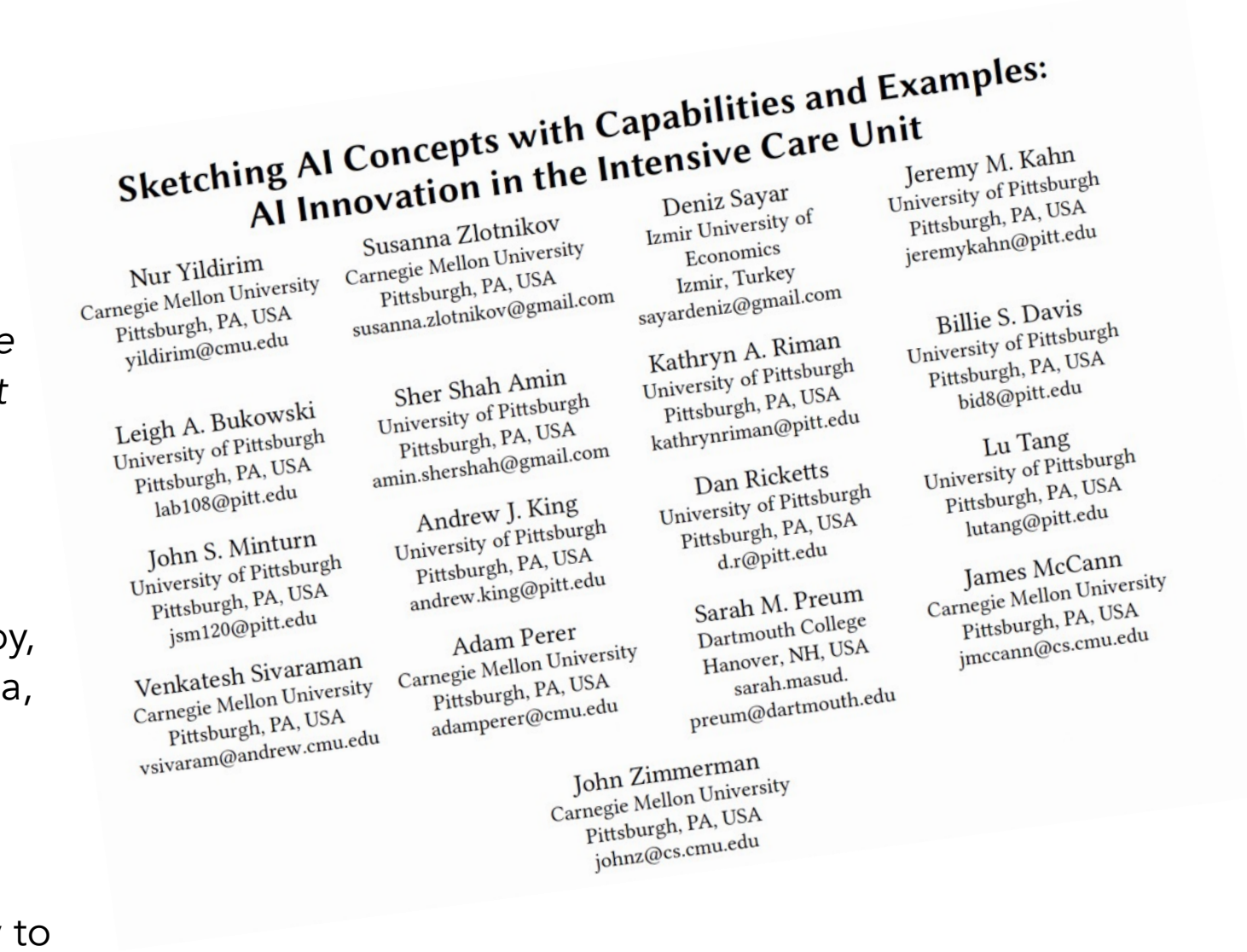
Artefact contributions might be evaluated as part of empirical research. But they do not always need to be – well documented, justified and described artefacts can be a contribution in themselves (see e.g., Gaver and Hook for a discussion!)

Methodological

"Methodological research contributions create new knowledge that informs how we carry out our work. Such contributions may improve research or practice." (p40)

Contributions that are methodological might help us study users of design better systems by, e.g., helping us make use of new forms of data, analyse data with new tools, help us generate better or more diverse ideas, etc.

Usually methodological contributions are evaluated based on its usefulness and novelty to other methods, and how reproducible or how straightforward to adopt it is for others.



Sketching AI Concepts with Capabilities and Examples: AI Innovation in the Intensive Care Unit

Nur Yildirim
Carnegie Mellon University
Pittsburgh, PA, USA
yildirim@cmu.edu

Susanna Zlotnikov
Carnegie Mellon University
Pittsburgh, PA, USA
susanna.zlotnikov@gmail.com

Deniz Sayar
Izmir University of Economics
Izmir, Turkey
sayardeniz@gmail.com

Jeremy M. Kahn
University of Pittsburgh
Pittsburgh, PA, USA
jeremykahn@pitt.edu

Leigh A. Bukowski
University of Pittsburgh
Pittsburgh, PA, USA
lab108@pitt.edu

Sher Shah Amin
University of Pittsburgh
Pittsburgh, PA, USA
amin.shershah@gmail.com

Kathryn A. Riman
University of Pittsburgh
Pittsburgh, PA, USA
kathrynriman@pitt.edu

Billie S. Davis
University of Pittsburgh
Pittsburgh, PA, USA
bid8@pitt.edu

John S. Minturn
University of Pittsburgh
Pittsburgh, PA, USA
jsm120@pitt.edu

Andrew J. King
University of Pittsburgh
Pittsburgh, PA, USA
andrew.king@pitt.edu

Dan Ricketts
University of Pittsburgh
Pittsburgh, PA, USA
d.r@pitt.edu

Lu Tang
University of Pittsburgh
Pittsburgh, PA, USA
lutang@pitt.edu

Venkatesh Sivaraman
Carnegie Mellon University
Pittsburgh, PA, USA
vsivaram@andrew.cmu.edu

Adam Perer
Carnegie Mellon University
Pittsburgh, PA, USA
adamperer@cmu.edu

Sarah M. Preum
Dartmouth College
Hanover, NH, USA
sarah.masud.preum@dartmouth.edu

James McCann
Carnegie Mellon University
Pittsburgh, PA, USA
jmccann@cs.cmu.edu

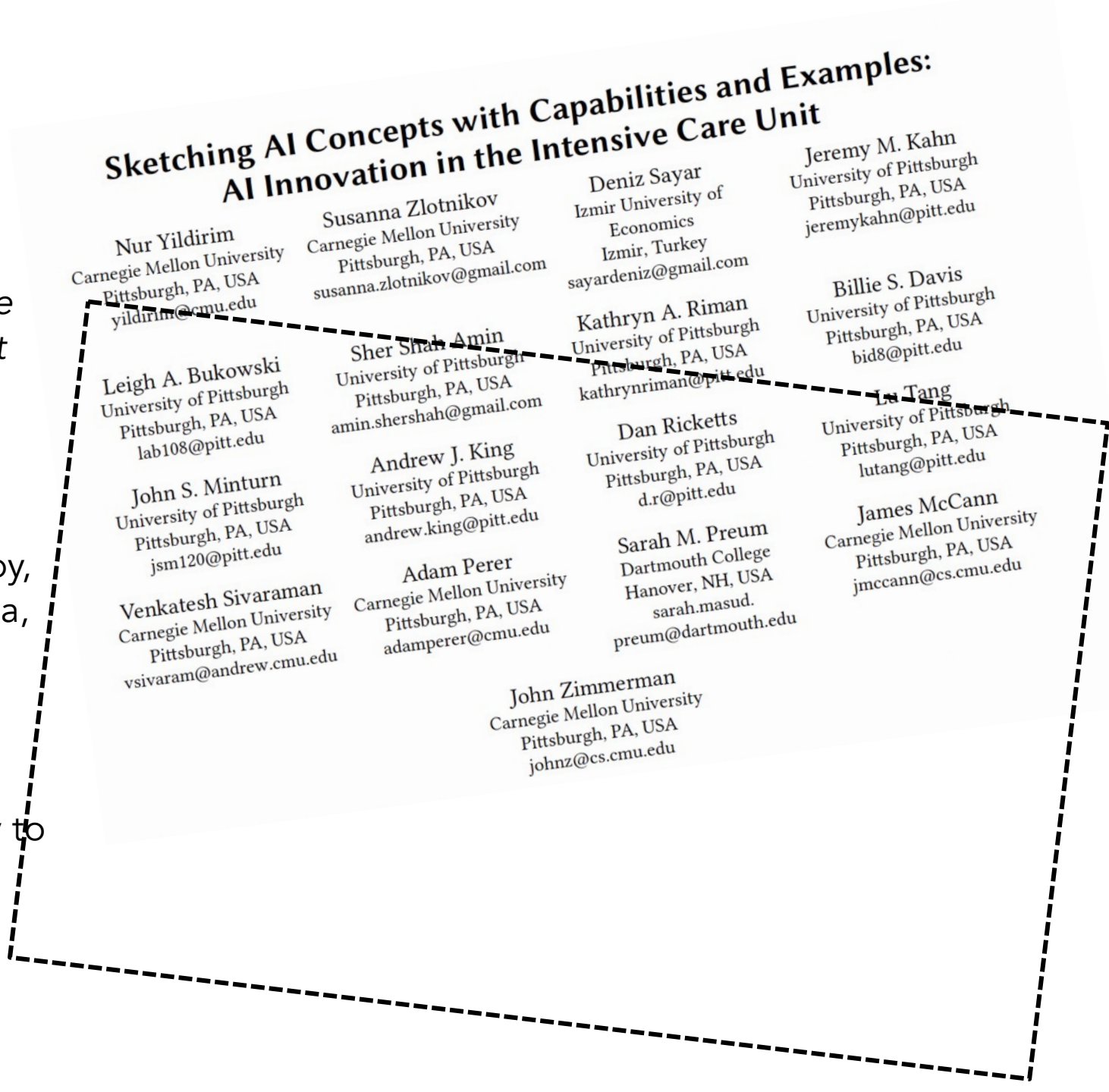
John Zimmerman
Carnegie Mellon University
Pittsburgh, PA, USA
johnz@cs.cmu.edu

Methodological

"Methodological research contributions create new knowledge that informs how we carry out our work. Such contributions may improve research or practice." (p40)

Contributions that are methodological might help us study users of design better systems by, e.g., helping us make use of new forms of data, analyse data with new tools, help us generate better or more diverse ideas, etc.

Usually methodological contributions are evaluated based on its usefulness and novelty to other methods, and how reproducible or how straightforward to adopt it is for others.



Sketching AI Concepts with Capabilities and Examples: AI Innovation in the Intensive Care Unit

Nur Yildirim
Carnegie Mellon University
Pittsburgh, PA, USA
yildirim@cmu.edu

Susanna Zlotnikov
Carnegie Mellon University
Pittsburgh, PA, USA
susanna.zlotnikov@gmail.com

Deniz Sayar
Izmir University of Economics
Izmir, Turkey
sayardeniz@gmail.com

Jeremy M. Kahn
University of Pittsburgh
Pittsburgh, PA, USA
jeremykahn@pitt.edu

Leigh A. Bukowski
University of Pittsburgh
Pittsburgh, PA, USA
lab108@pitt.edu

Sher Shah Amin
University of Pittsburgh
Pittsburgh, PA, USA
amin.shershah@gmail.com

Kathryn A. Riman
University of Pittsburgh
Pittsburgh, PA, USA
kathrynriman@pitt.edu

Billie S. Davis
University of Pittsburgh
Pittsburgh, PA, USA
bid8@pitt.edu

John S. Minturn
University of Pittsburgh
Pittsburgh, PA, USA
jsm120@pitt.edu

Andrew J. King
University of Pittsburgh
Pittsburgh, PA, USA
andrew.king@pitt.edu

Dan Ricketts
University of Pittsburgh
Pittsburgh, PA, USA
d.r@pitt.edu

Lu Tang
University of Pittsburgh
Pittsburgh, PA, USA
lutang@pitt.edu

Venkatesh Sivaraman
Carnegie Mellon University
Pittsburgh, PA, USA
vsivaram@andrew.cmu.edu

Adam Perer
Carnegie Mellon University
Pittsburgh, PA, USA
adamperer@cmu.edu

Sarah M. Preum
Dartmouth College
Hanover, NH, USA
sarah.masud.preum@dartmouth.edu

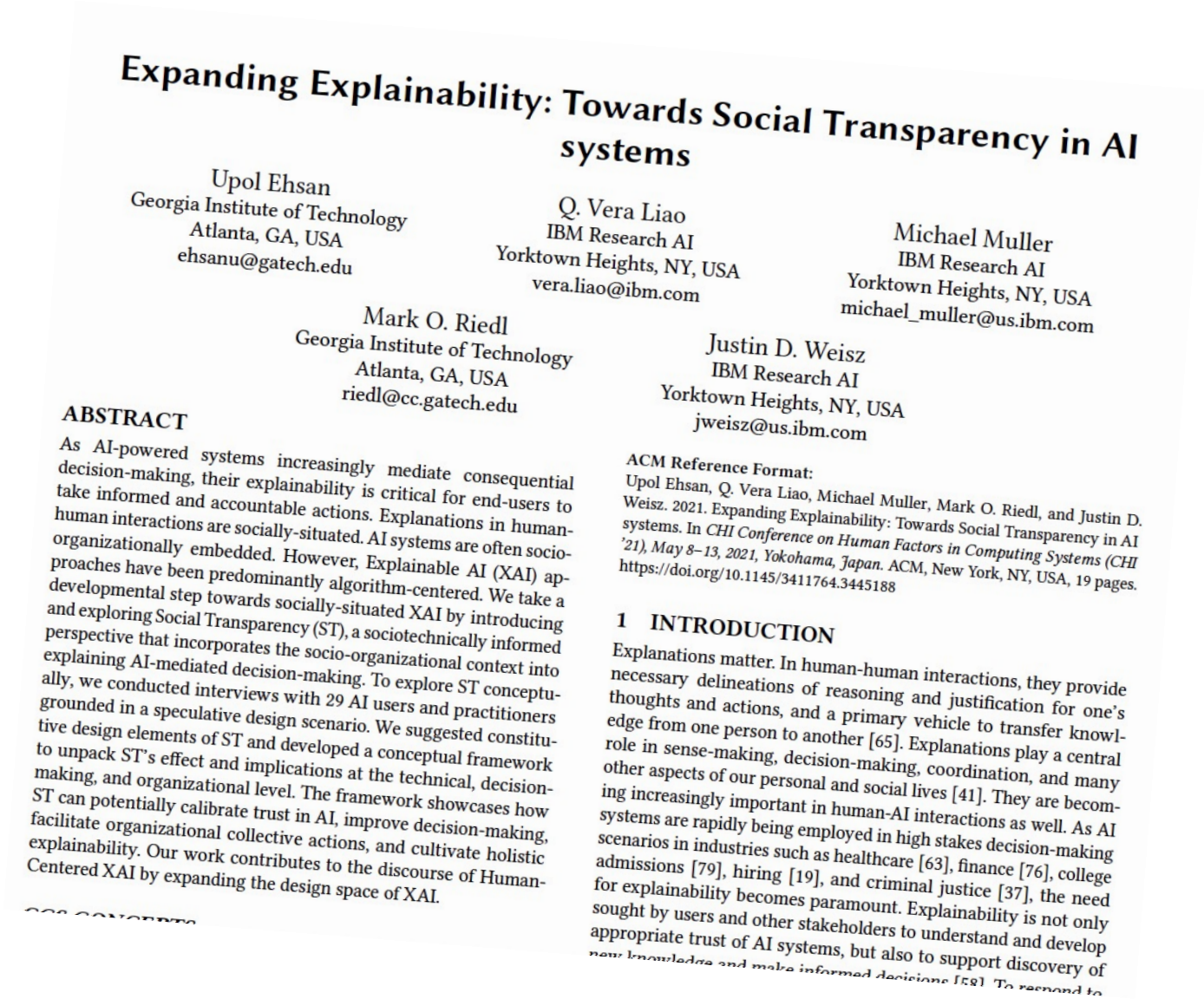
James McCann
Carnegie Mellon University
Pittsburgh, PA, USA
jmccann@cs.cmu.edu

John Zimmerman
Carnegie Mellon University
Pittsburgh, PA, USA
johnz@cs.cmu.edu

Theoretical

"Theoretical research contributions consist of new or improved concepts, definitions, models, principles, or frameworks. They are vehicles for thought. Whereas methodological contributions inform how we do things, theoretical contributions inform what we do, why we do it, and what we expect from it." (p41)

Theories can be descriptive in nature (i.e., they could describe how certain groups of users behave with certain systems) or predictive (i.e., if you introduce a certain design feature this is how people will behave). Good theoretical contributions tell us why things happen, not just what happens.



Expanding Explainability: Towards Social Transparency in AI systems

Upol Ehsan
Georgia Institute of Technology
Atlanta, GA, USA
ehsanu@gatech.edu

Q. Vera Liao
IBM Research AI
Yorktown Heights, NY, USA
vera.liao@ibm.com

Michael Muller
IBM Research AI
Yorktown Heights, NY, USA
michael_muller@us.ibm.com

Mark O. Riedl
Georgia Institute of Technology
Atlanta, GA, USA
riedl@cc.gatech.edu

Justin D. Weisz
IBM Research AI
Yorktown Heights, NY, USA
jweisz@us.ibm.com

ABSTRACT

As AI-powered systems increasingly mediate consequential decision-making, their explainability is critical for end-users to take informed and accountable actions. Explanations in human-human interactions are socially-situated. AI systems are often socio-organizationally embedded. However, Explainable AI (XAI) approaches have been predominantly algorithm-centered. We take a developmental step towards socially-situated XAI by introducing and exploring Social Transparency (ST), a sociotechnically informed perspective that incorporates the socio-organizational context into explaining AI-mediated decision-making. To explore ST conceptually, we conducted interviews with 29 AI users and practitioners grounded in a speculative design scenario. We suggested constitutive design elements of ST and developed a conceptual framework to unpack ST's effect and implications at the technical, decision-making, and organizational level. The framework showcases how ST can potentially calibrate trust in AI, improve decision-making, facilitate organizational collective actions, and cultivate holistic explainability. Our work contributes to the discourse of Human-Centered XAI by expanding the design space of XAI.

ACM Reference Format:
Upol Ehsan, Q. Vera Liao, Michael Muller, Mark O. Riedl, and Justin D. Weisz. 2021. Expanding Explainability: Towards Social Transparency in AI systems. In *CHI Conference on Human Factors in Computing Systems (CHI '21), May 8–13, 2021, Yokohama, Japan.* ACM, New York, NY, USA, 19 pages. https://doi.org/10.1145/3411764.3445188

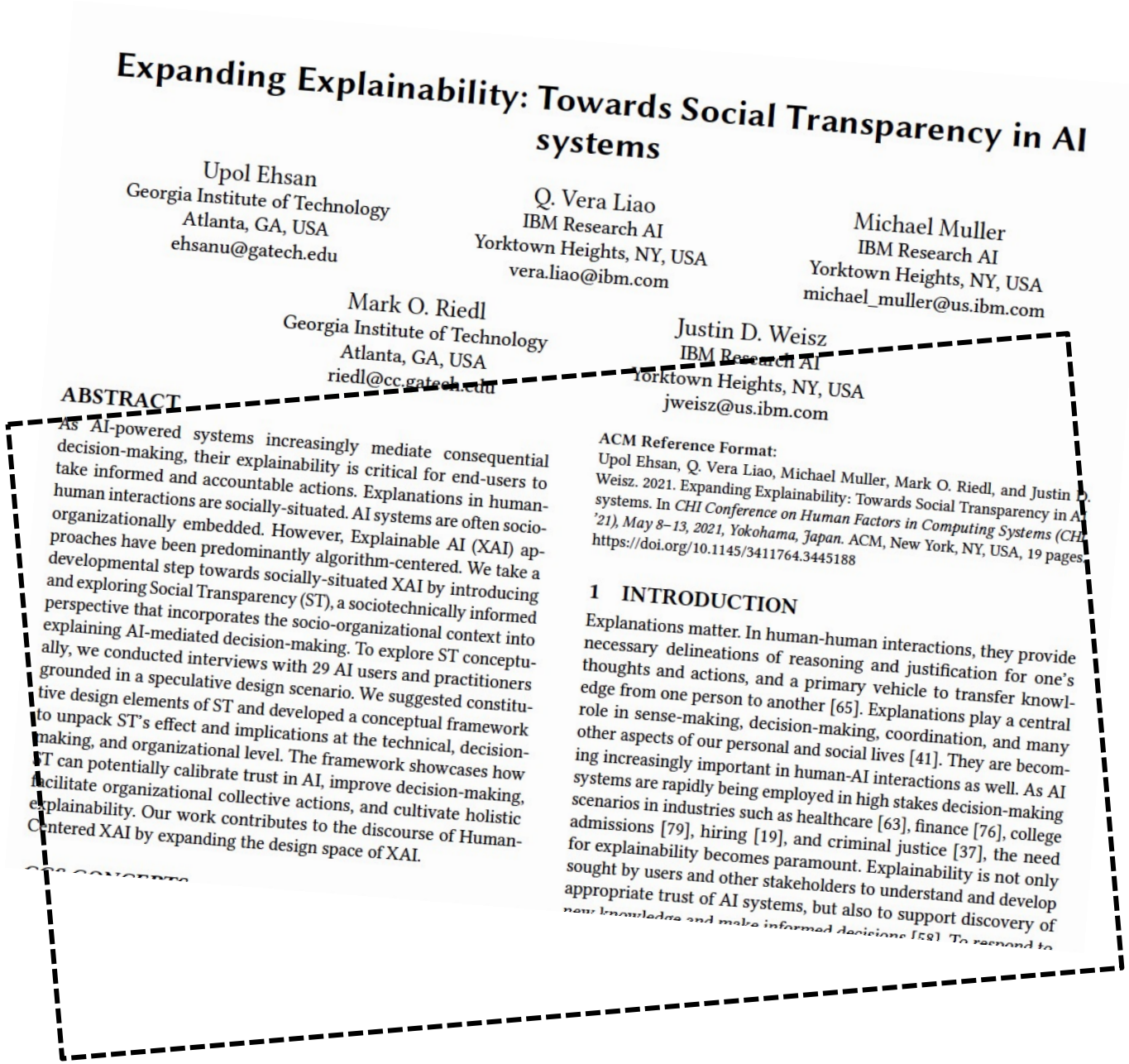
1 INTRODUCTION

Explanations matter. In human-human interactions, they provide necessary delineations of reasoning and justification for one's thoughts and actions, and a primary vehicle to transfer knowledge from one person to another [65]. Explanations play a central role in sense-making, decision-making, coordination, and many other aspects of our personal and social lives [41]. They are becoming increasingly important in human-AI interactions as well. As AI systems are rapidly being employed in high stakes decision-making scenarios in industries such as healthcare [63], finance [76], college admissions [79], hiring [19], and criminal justice [37], the need for explainability becomes paramount. Explainability is not only sought by users and other stakeholders to understand and develop appropriate trust of AI systems, but also to support discovery of

Theoretical

"Theoretical research contributions consist of new or improved concepts, definitions, models, principles, or frameworks. They are vehicles for thought. Whereas methodological contributions inform how we do things, theoretical contributions inform what we do, why we do it, and what we expect from it." (p41)

Theories can be descriptive in nature (i.e., they could describe how certain groups of users behave with certain systems) or predictive (i.e., if you introduce a certain design feature this is how people will behave). Good theoretical contributions tell us why things happen, not just what happens.



Expanding Explainability: Towards Social Transparency in AI systems

Upol Ehsan
Georgia Institute of Technology
Atlanta, GA, USA
ehsanu@gatech.edu

Q. Vera Liao
IBM Research AI
Yorktown Heights, NY, USA
vera.liao@ibm.com

Michael Muller
IBM Research AI
Yorktown Heights, NY, USA
michael_muller@us.ibm.com

Mark O. Riedl
Georgia Institute of Technology
Atlanta, GA, USA
riedl@cc.gatech.edu

Justin D. Weisz
IBM Research AI
Yorktown Heights, NY, USA
jweisz@us.ibm.com

ABSTRACT

As AI-powered systems increasingly mediate consequential decision-making, their explainability is critical for end-users to take informed and accountable actions. Explanations in human-human interactions are socially-situated. AI systems are often socio-organizationally embedded. However, Explainable AI (XAI) approaches have been predominantly algorithm-centered. We take a developmental step towards socially-situated XAI by introducing and exploring Social Transparency (ST), a sociotechnically informed perspective that incorporates the socio-organizational context into explaining AI-mediated decision-making. To explore ST conceptually, we conducted interviews with 29 AI users and practitioners grounded in a speculative design scenario. We suggested constitutive design elements of ST and developed a conceptual framework to unpack ST's effect and implications at the technical, decision-making, and organizational level. The framework showcases how ST can potentially calibrate trust in AI, improve decision-making, facilitate organizational collective actions, and cultivate holistic explainability. Our work contributes to the discourse of Human-Centered XAI by expanding the design space of XAI.

ACM Reference Format:
Upol Ehsan, Q. Vera Liao, Michael Muller, Mark O. Riedl, and Justin D. Weisz. 2021. Expanding Explainability: Towards Social Transparency in AI systems. In *CHI Conference on Human Factors in Computing Systems (CHI '21), May 8–13, 2021, Yokohama, Japan.* ACM, New York, NY, USA, 19 pages. https://doi.org/10.1145/3411764.3445188

1 INTRODUCTION

Explanations matter. In human-human interactions, they provide necessary delineations of reasoning and justification for one's thoughts and actions, and a primary vehicle to transfer knowledge from one person to another [65]. Explanations play a central role in sense-making, decision-making, coordination, and many other aspects of our personal and social lives [41]. They are becoming increasingly important in human-AI interactions as well. As AI systems are rapidly being employed in high stakes decision-making scenarios in industries such as healthcare [63], finance [76], college admissions [79], hiring [19], and criminal justice [37], the need for explainability becomes paramount. Explainability is not only sought by users and other stakeholders to understand and develop appropriate trust of AI systems, but also to support discovery of

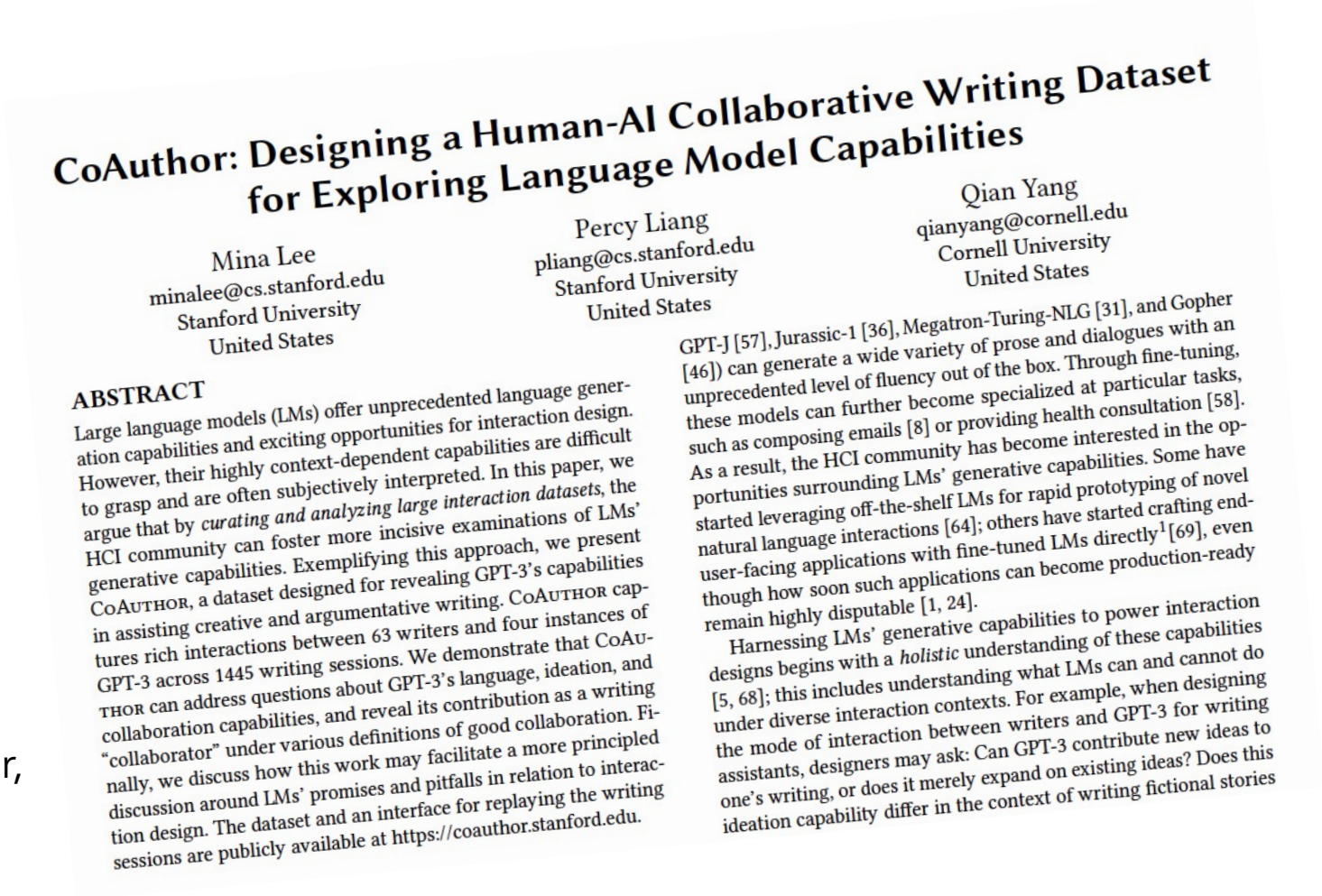
Datasets

"A dataset contribution provides a new and useful corpus, often accompanied by an analysis of its characteristics, for the benefit of the research community." (p41)

It's common to see a dataset be contributed alongside a new tool or method of analysis.

Datasets have been historically rare – however, we are increasingly seeing research funders pushing researchers to publish datasets from their research under open access policies, which is increasing their presence.

It's still very rare for a dataset to be a contribution on its own though.



CoAuthor: Designing a Human-AI Collaborative Writing Dataset for Exploring Language Model Capabilities

Mina Lee
minalee@cs.stanford.edu
Stanford University
United States

Percy Liang
pliang@cs.stanford.edu
Stanford University
United States

Qian Yang
qianyang@cornell.edu
Cornell University
United States

ABSTRACT

Large language models (LMs) offer unprecedented language generation capabilities and exciting opportunities for interaction design. However, their highly context-dependent capabilities are difficult to grasp and are often subjectively interpreted. In this paper, we argue that by *curating and analyzing large interaction datasets*, the HCI community can foster more incisive examinations of LMs' generative capabilities. Exemplifying this approach, we present CoAuthor, a dataset designed for revealing GPT-3's capabilities in assisting creative and argumentative writing. CoAuthor captures rich interactions between 63 writers and four instances of GPT-3 across 1445 writing sessions. We demonstrate that CoAuthor can address questions about GPT-3's language, ideation, and collaboration capabilities, and reveal its contribution as a writing "collaborator" under various definitions of good collaboration. Finally, we discuss how this work may facilitate a more principled discussion around LMs' promises and pitfalls in relation to interaction design. The dataset and an interface for replaying the writing sessions are publicly available at https://coauthor.stanford.edu.

GPT-J [57], Jurassic-1 [36], Megatron-Turing-NLG [31], and Gopher [46]) can generate a wide variety of prose and dialogues with an unprecedented level of fluency out of the box. Through fine-tuning, these models can further become specialized at particular tasks, such as composing emails [8] or providing health consultation [58]. As a result, the HCI community has become interested in the opportunities surrounding LMs' generative capabilities. Some have started leveraging off-the-shelf LMs for rapid prototyping of novel natural language interactions [64]; others have started crafting end-user-facing applications with fine-tuned LMs directly[1] [69], even though how soon such applications can become production-ready remain highly disputable [1, 24].

Harnessing LMs' generative capabilities to power interaction designs begins with a *holistic* understanding of these capabilities [5, 68]; this includes understanding what LMs can and cannot do under diverse interaction contexts. For example, when designing the mode of interaction between writers and GPT-3 for writing assistants, designers may ask: Can GPT-3 contribute new ideas to one's writing, or does it merely expand on existing ideas? Does this ideation capability differ in the context of writing fictional stories

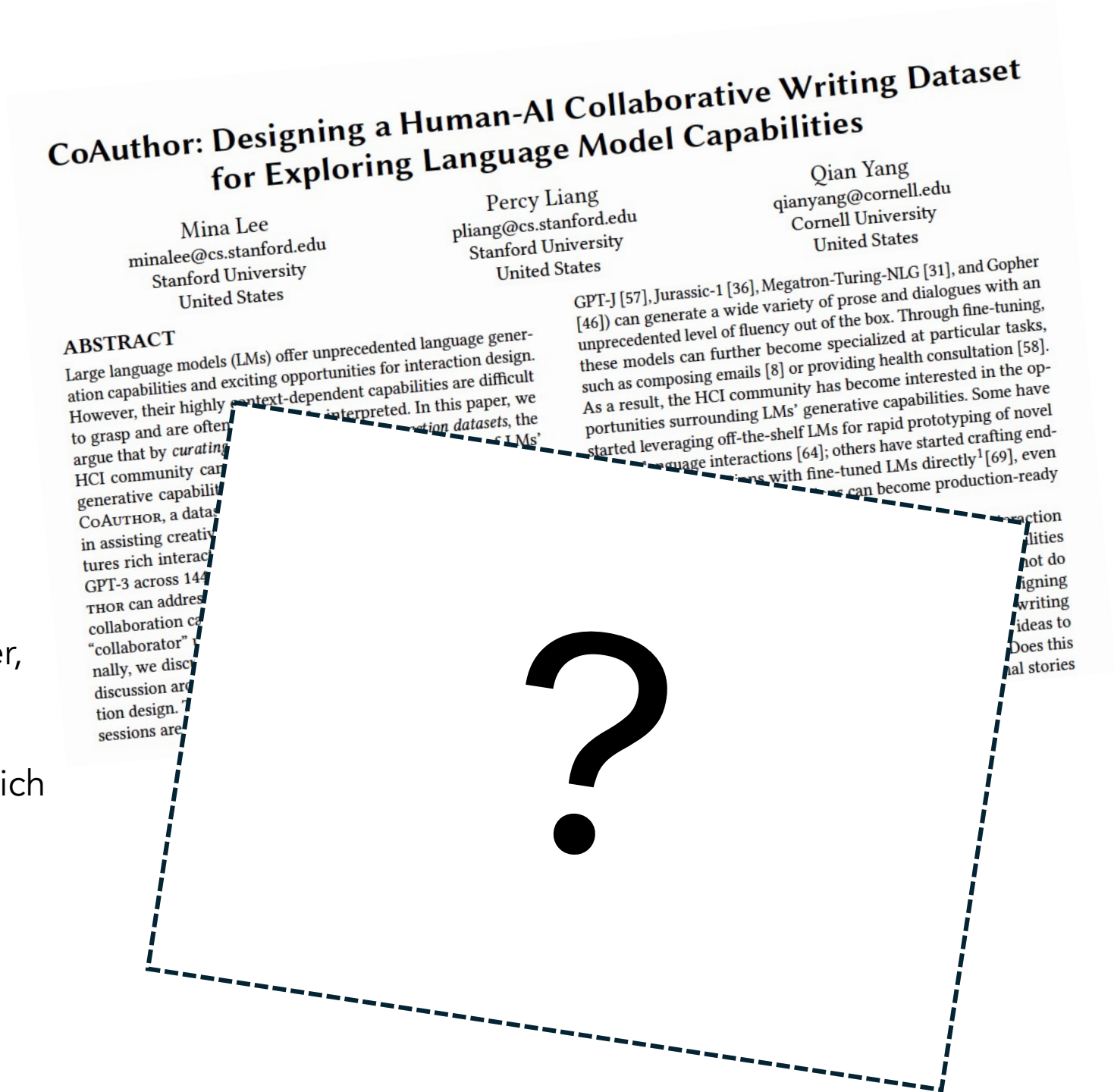
Datasets

"A dataset contribution provides a new and useful corpus, often accompanied by an analysis of its characteristics, for the benefit of the research community." (p41)

It's common to see a dataset be contributed alongside a new tool or method of analysis.

Datasets have been historically rare – however, we are increasingly seeing research funders pushing researchers to publish datasets from their research under open access policies, which is increasing their presence.

It's still very rare for a dataset to be a contribution on its own though.

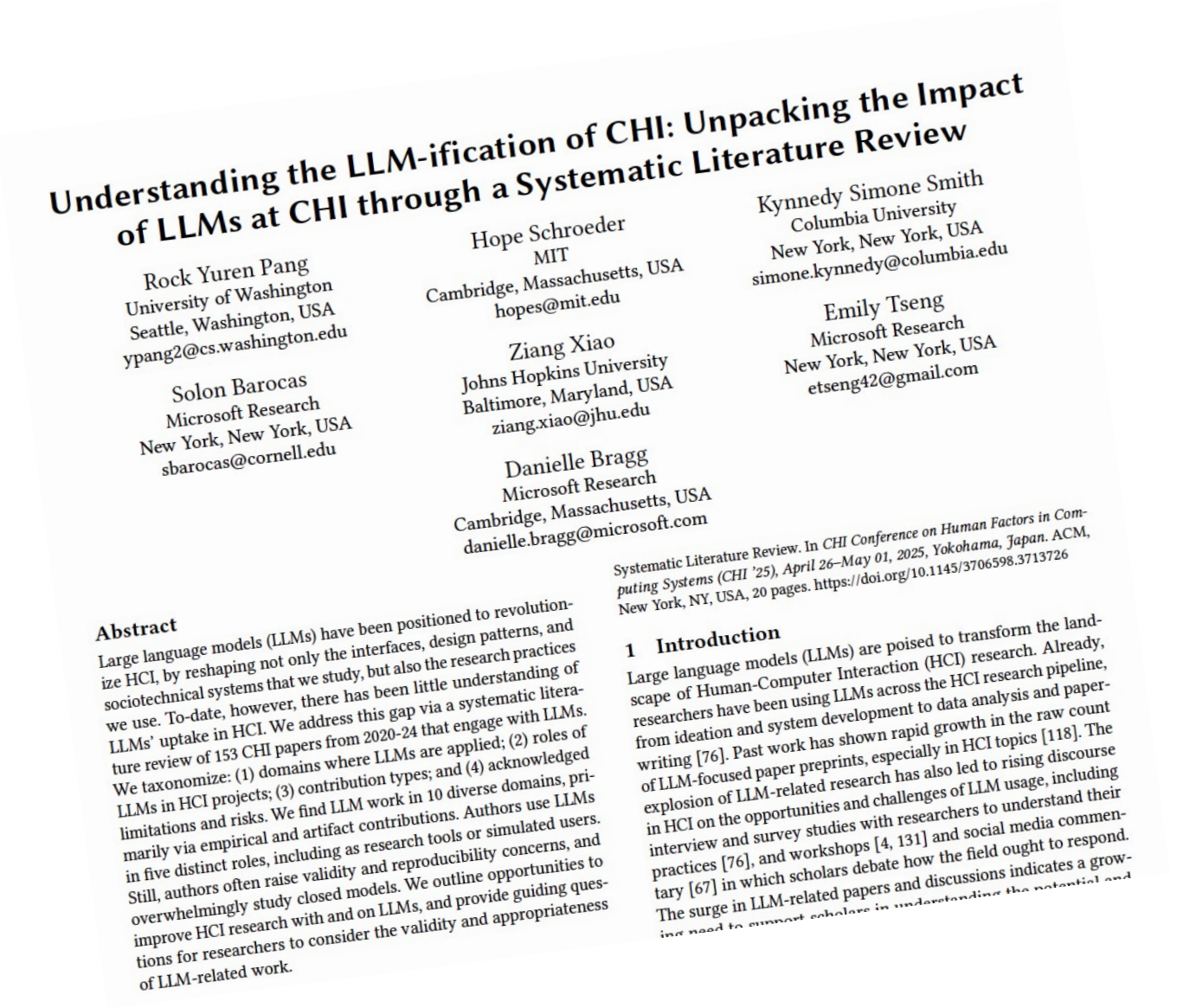


Survey

"Survey research contributions and other meta-analyses review and synthesize work done on a research topic with the goal of exposing trends and gaps. Survey contributions are appropriate after a topic has reached a certain level of maturity." (p42)

Notably, surveys were very rare in HCI conferences and journals until relatively recently – perhaps as they field is so diverse, and relatively new.

Surveys are evaluated on how well they analyse and organise existing work, and how well their analysis of literature reveals insights for new research and design work.



Understanding the LLM-ification of CHI: Unpacking the Impact of LLMs at CHI through a Systematic Literature Review

Rock Yuren Pang
University of Washington
Seattle, Washington, USA
ypang2@cs.washington.edu

Hope Schroeder
MIT
Cambridge, Massachusetts, USA
hopes@mit.edu

Kynnedy Simone Smith
Columbia University
New York, New York, USA
simone.kynnedy@columbia.edu

Solon Barocas
Microsoft Research
New York, New York, USA
sbarocas@cornell.edu

Ziang Xiao
Johns Hopkins University
Baltimore, Maryland, USA
ziang.xiao@jhu.edu

Emily Tseng
Microsoft Research
New York, New York, USA
etseng42@gmail.com

Danielle Bragg
Microsoft Research
Cambridge, Massachusetts, USA
danielle.bragg@microsoft.com

Systematic Literature Review. In *CHI Conference on Human Factors in Computing Systems (CHI '25), April 26–May 01, 2025, Yokohama, Japan.* ACM, New York, NY, USA, 20 pages. https://doi.org/10.1145/3706598.3713726

Abstract

Large language models (LLMs) have been positioned to revolutionize HCI, by reshaping not only the interfaces, design patterns, and sociotechnical systems that we study, but also the research practices we use. To-date, however, there has been little understanding of LLMs' uptake in HCI. We address this gap via a systematic literature review of 153 CHI papers from 2020-24 that engage with LLMs. We taxonomize: (1) domains where LLMs are applied; (2) roles of LLMs in HCI projects; (3) contribution types; and (4) acknowledged limitations and risks. We find LLM work in 10 diverse domains, primarily via empirical and artifact contributions. Authors use LLMs in five distinct roles, including as research tools or simulated users. Still, authors often raise validity and reproducibility concerns, and overwhelmingly study closed models. We outline opportunities to improve HCI research with and on LLMs, and provide guiding questions for researchers to consider the validity and appropriateness of LLM-related work.

1 Introduction

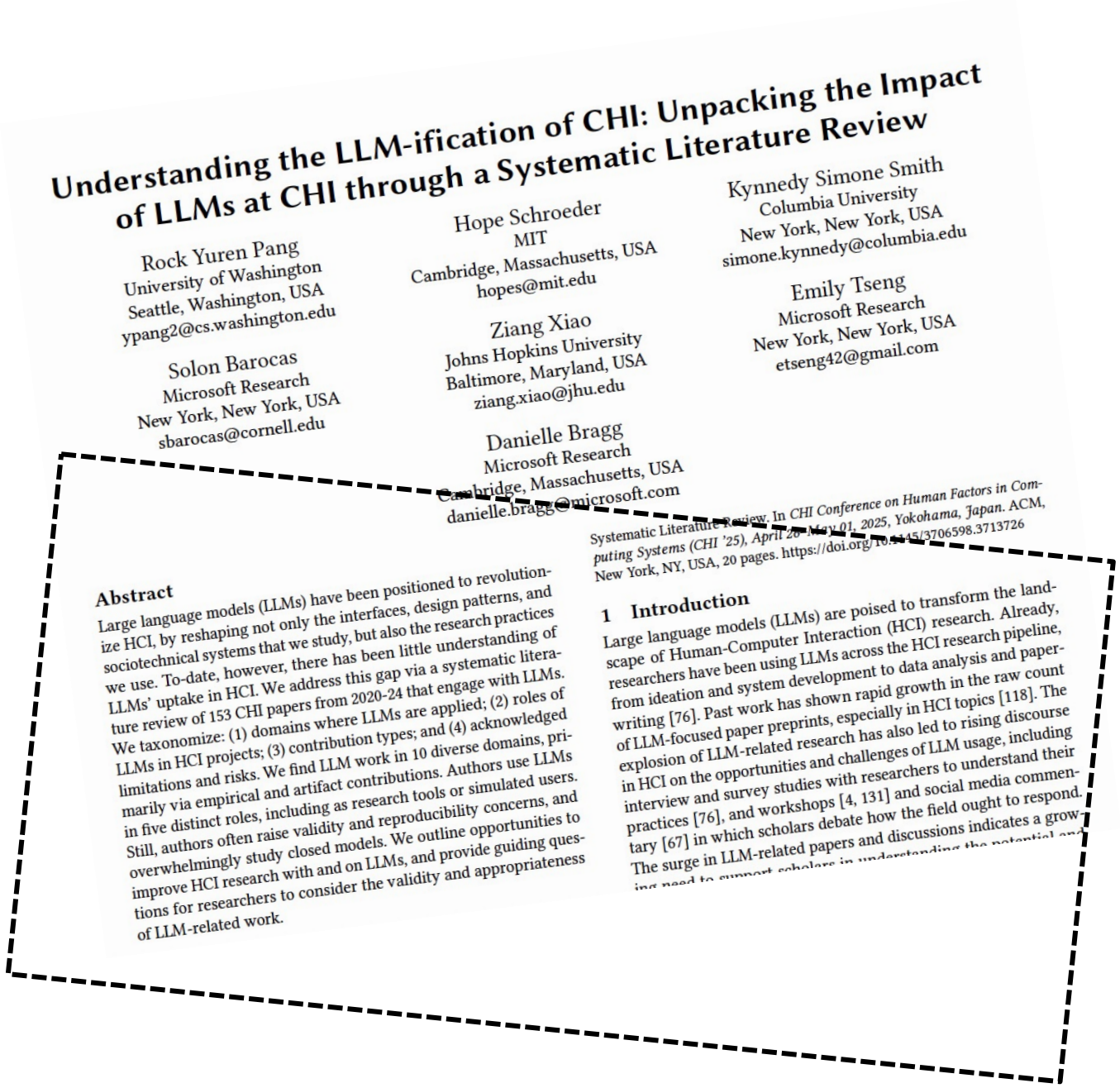
Large language models (LLMs) are poised to transform the landscape of Human-Computer Interaction (HCI) research. Already, researchers have been using LLMs across the HCI research pipeline, from ideation and system development to data analysis and paper-writing [76]. Past work has shown rapid growth in the raw count of LLM-focused paper preprints, especially in HCI topics [118]. The explosion of LLM-related research has also led to rising discourse in HCI on the opportunities and challenges of LLM usage, including interview and survey studies with researchers to understand their practices [76], and workshops [4, 131] and social media commentary [67] in which scholars debate how the field ought to respond. The surge in LLM-related papers and discussions indicates a grow-

Survey

"Survey research contributions and other meta-analyses review and synthesize work done on a research topic with the goal of exposing trends and gaps. Survey contributions are appropriate after a topic has reached a certain level of maturity." (p42)

Notably, surveys were very rare in HCI conferences and journals until relatively recently – perhaps as they field is so diverse, and relatively new.

Surveys are evaluated on how well they analyse and organise existing work, and how well their analysis of literature reveals insights for new research and design work.



Understanding the LLM-ification of CHI: Unpacking the Impact of LLMs at CHI through a Systematic Literature Review

Rock Yuren Pang
University of Washington
Seattle, Washington, USA
ypang2@cs.washington.edu

Hope Schroeder
MIT
Cambridge, Massachusetts, USA
hopes@mit.edu

Kynnedy Simone Smith
Columbia University
New York, New York, USA
simone.kynnedy@columbia.edu

Solon Barocas
Microsoft Research
New York, New York, USA
sbarocas@cornell.edu

Ziang Xiao
Johns Hopkins University
Baltimore, Maryland, USA
ziang.xiao@jhu.edu

Emily Tseng
Microsoft Research
New York, New York, USA
etseng42@gmail.com

Danielle Bragg
Microsoft Research
Cambridge, Massachusetts, USA
danielle.bragg@microsoft.com

Systematic Literature Review. In *CHI Conference on Human Factors in Computing Systems (CHI '25), April 26-May 01, 2025, Yokohama, Japan.* ACM, New York, NY, USA, 20 pages. https://doi.org/10.1145/3706598.3713726

Abstract

Large language models (LLMs) have been positioned to revolutionize HCI, by reshaping not only the interfaces, design patterns, and sociotechnical systems that we study, but also the research practices we use. To-date, however, there has been little understanding of LLMs' uptake in HCI. We address this gap via a systematic literature review of 153 CHI papers from 2020-24 that engage with LLMs. We taxonomize: (1) domains where LLMs are applied; (2) roles of LLMs in HCI projects; (3) contribution types; and (4) acknowledged limitations and risks. We find LLM work in 10 diverse domains, primarily via empirical and artifact contributions. Authors use LLMs in five distinct roles, including as research tools or simulated users. Still, authors often raise validity and reproducibility concerns, and overwhelmingly study closed models. We outline opportunities to improve HCI research with and on LLMs, and provide guiding questions for researchers to consider the validity and appropriateness of LLM-related work.

1 Introduction

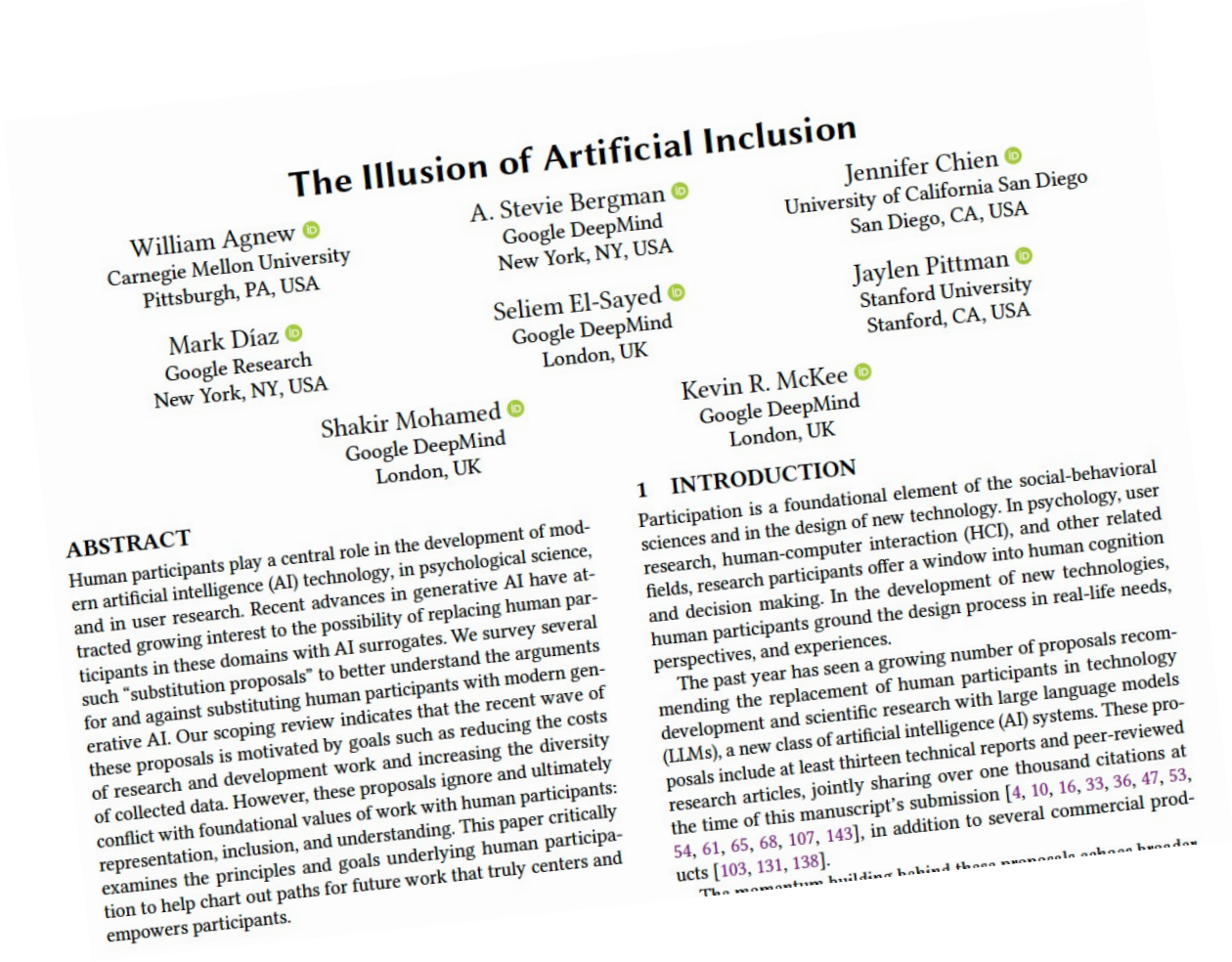
Large language models (LLMs) are poised to transform the landscape of Human-Computer Interaction (HCI) research. Already, researchers have been using LLMs across the HCI research pipeline, from ideation and system development to data analysis and paper-writing [76]. Past work has shown rapid growth in the raw count of LLM-focused paper preprints, especially in HCI topics [118]. The explosion of LLM-related research has also led to rising discourse in HCI on the opportunities and challenges of LLM usage, including interview and survey studies with researchers to understand their practices [76], and workshops [4, 131] and social media commentary [67] in which scholars debate how the field ought to respond. The surge in LLM-related papers and discussions indicates a grow-

Opinion

Opinion contributions might also be referred to as essays or arguments – their purpose tends to be to persuade the reader about the arguments the paper presents.

These types of contributions tend to tackle emergent or complex problems in the field, often drawing on ideas and concepts from outside of HCI and bring them into contact with challenges in the field of HCI. They are very hard papers to write – and just as hard to evaluate and review.

Opinion papers get evaluated on the strength of their argument – strong arguments typically are those that make use of credible evidence, and consider different perspectives on the topic.



The Illusion of Artificial Inclusion

William Agnew
Carnegie Mellon University
Pittsburgh, PA, USA

A. Stevie Bergman
Google DeepMind
New York, NY, USA

Jennifer Chien
University of California San Diego
San Diego, CA, USA

Mark Díaz
Google Research
New York, NY, USA

Seliem El-Sayed
Google DeepMind
London, UK

Jaylen Pittman
Stanford University
Stanford, CA, USA

Shakir Mohamed
Google DeepMind
London, UK

Kevin R. McKee
Google DeepMind
London, UK

ABSTRACT

Human participants play a central role in the development of modern artificial intelligence (AI) technology, in psychological science, and in user research. Recent advances in generative AI have attracted growing interest to the possibility of replacing human participants in these domains with AI surrogates. We survey several such "substitution proposals" to better understand the arguments for and against substituting human participants with modern generative AI. Our scoping review indicates that the recent wave of these proposals is motivated by goals such as reducing the costs of research and development work and increasing the diversity of collected data. However, these proposals ignore and ultimately conflict with foundational values of work with human participants: representation, inclusion, and understanding. This paper critically examines the principles and goals underlying human participation to help chart out paths for future work that truly centers and empowers participants.

1 INTRODUCTION

Participation is a foundational element of the social-behavioral sciences and in the design of new technology. In psychology, user research, human-computer interaction (HCI), and other related fields, research participants offer a window into human cognition and decision making. In the development of new technologies, human participants ground the design process in real-life needs, perspectives, and experiences.

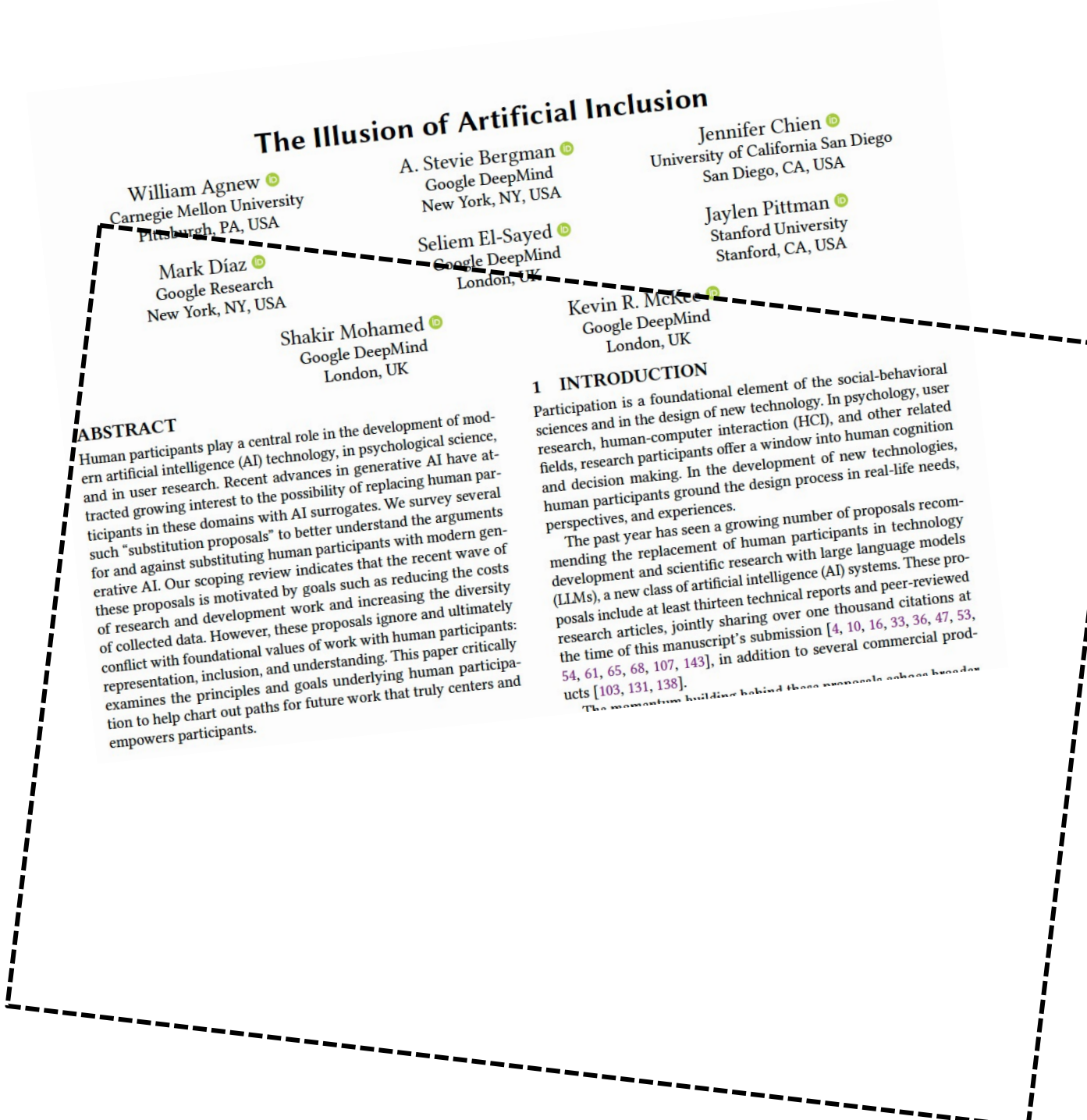
The past year has seen a growing number of proposals recommending the replacement of human participants in technology development and scientific research with large language models (LLMs), a new class of artificial intelligence (AI) systems. These proposals include at least thirteen technical reports and peer-reviewed research articles, jointly sharing over one thousand citations at the time of this manuscript's submission [4, 10, 16, 33, 36, 47, 53, 54, 61, 65, 68, 107, 143], in addition to several commercial products [103, 131, 138].

Opinion

Opinion contributions might also be referred to as essays or arguments – their purpose tends to be to persuade the reader about the arguments the paper presents.

These types of contributions tend to tackle emergent or complex problems in the field, often drawing on ideas and concepts from outside of HCI and bring them into contact with challenges in the field of HCI. They are very hard papers to write – and just as hard to evaluate and review.

Opinion papers get evaluated on the strength of their argument – strong arguments typically are those that make use of credible evidence, and consider different perspectives on the topic.



The Illusion of Artificial Inclusion

William Agnew
Carnegie Mellon University
Pittsburgh, PA, USA

A. Stevie Bergman
Google DeepMind
New York, NY, USA

Jennifer Chien
University of California San Diego
San Diego, CA, USA

Jaylen Pittman
Stanford University
Stanford, CA, USA

Mark Díaz
Google Research
New York, NY, USA

Seliem El-Sayed
Google DeepMind
London, UK

Shakir Mohamed
Google DeepMind
London, UK

Kevin R. McKee
Google DeepMind
London, UK

ABSTRACT

Human participants play a central role in the development of modern artificial intelligence (AI) technology, in psychological science, and in user research. Recent advances in generative AI have attracted growing interest to the possibility of replacing human participants in these domains with AI surrogates. We survey several such "substitution proposals" to better understand the arguments for and against substituting human participants with modern generative AI. Our scoping review indicates that the recent wave of these proposals is motivated by goals such as reducing the costs of research and development work and increasing the diversity of collected data. However, these proposals ignore and ultimately conflict with foundational values of work with human participants: representation, inclusion, and understanding. This paper critically examines the principles and goals underlying human participation to help chart out paths for future work that truly centers and empowers participants.

1 INTRODUCTION

Participation is a foundational element of the social-behavioral sciences and in the design of new technology. In psychology, user research, human-computer interaction (HCI), and other related fields, research participants offer a window into human cognition and decision making. In the development of new technologies, human participants ground the design process in real-life needs, perspectives, and experiences.

The past year has seen a growing number of proposals recommending the replacement of human participants in technology development and scientific research with large language models (LLMs), a new class of artificial intelligence (AI) systems. These proposals include at least thirteen technical reports and peer-reviewed research articles, jointly sharing over one thousand citations at the time of this manuscript's submission [4, 10, 16, 33, 36, 47, 53, 54, 61, 65, 68, 107, 143], in addition to several commercial products [103, 131, 138].

What are the contributions of your papers?

Working on your own, choose one of your papers:

Read the introduction, and skim the sections of the rest of the paper:

- How does it motivate the research – what is it responding to? Does it refer to a "lack" or "gap" in knowledge?
- How does it convey its "novelty"?
- How is it structured – what types of sections does it have?
- How does it convey its "contribution"?
- Is there anything unusual about its content based on papers in other fields you may have read?

Any questions?

If you have any questions about this week, contact me at :
john.vines@ed.ac.uk

THE UNIVERSITY *of* EDINBURGH

design informatics

Designing Responsible Natural Language Processing