

Chapter 3

Beyond Human-in-the-Loop: Empowering End-Users with Transparent Machine Learning

Patrick C. Shih

Abstract Advances in data analytics and human computation are transforming how researchers conduct science in domains like bioinformatics, computational social science, and digital humanities. However, data analytics requires significant programming knowledge or access to technical experts, while human computation requires in-depth knowledge of crowd management and is error-prone due to lack of scientific domain expertise. The goal of this research is to empower a broader range of scientists and end-users to conduct data analytics by adopting the End-User Development (EUD) models commonly found in today’s commercial software platforms like Microsoft Excel, Wikipedia and WordPress. These EUD platforms enable people to focus on producing content rather than struggling with a development environment and new programming syntax or relying on disciplinary non-experts for essential technical help. This research explores a similar paradigm for scientists and end-users that can be thought of as End-User Data Analytics (EUDA), or Transparent Machine Learning (TML).

3.1 Introduction

The scientific method is based on empirical measures that provide evidence for hypothesis formation and reasoning. The process typically involves “systematic observation, measurement, and experiment, and the formulation, testing, and modification of hypotheses” [21]. Critical thinking—“the intellectually disciplined process of actively and skillfully conceptualizing, applying, analyzing, synthesizing, and/or evaluating information gathered from, or generated by, observation, experience, re-

P. C. Shih
Department of Informatics, School of Informatics, Computing, and Engineering
Indiana University Bloomington, Bloomington, IN, 47408, USA
e-mail: patshih@indiana.edu

flection, reasoning, or communication, as a guide to belief and action” [84]—is key to the process.

In empirical research, the scientific method typically involves a scientist collecting data based on interviews, observations, surveys, or sampling of specimens. Once the raw data sample is collected, a data cleaning and coding process identifies outliers and erroneous data resulting from sampling error, and the researcher synthesises raw data points into aggregated clusters or themes that suit the research focus. A data analytics and validation process typically follows, involving statistical or inter-coder reliability checks to ensure the quality of the findings. Finally, the results are formatted in a fashion appropriate for the intended audience, be it a research or public community. Fig. 3.1 provides a simplified view of the traditional scientific inquiry process.

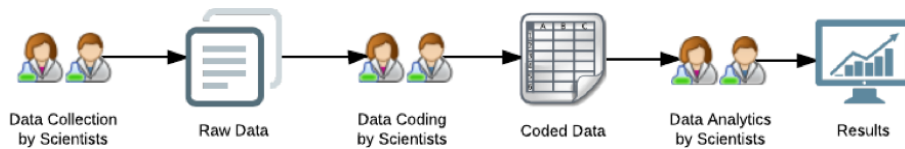


Fig. 3.1 Traditional Process of Scientific Inquiry.

With the improvement of data collection instruments (e.g., space imaging for astrophysicists, environmental sampling for climate scientists, etc.) and the emergence and wide adoption of consumer Information and Communication Technologies (ICTs), researchers are turning to a broad variety of data sources to infer sample population characteristics and patterns [47]. Although improvements in data collection have enabled scientists to make more accurate generalisations and ask novel questions, the sheer amount of available data can exceed scientists’ ability to utilise or process it. Some have described this as the “Big Data” phenomena, defined by the three V’s: volume, variety, and velocity [68]. To cope, the scientific community has enlisted the help of citizen science and crowdsourcing platforms to engage the public in both data collection and data analysis [109]. However, this naturally results in a crowd management problem in which factors like task modulation, task coordination, and data verification have added to the issues that scientists must actively manage [58]. Advances in computational infrastructure and the availability of big datasets have also led to a new set of computational techniques and data analytical tools capable of processing and visualising large scale datasets [16]. This imposes a further burden on scientists, however, in the form of having to constantly learn new computational techniques and manage new visualisation tools.

Thus, crowd management and computational data analytics have become vital skillsets that the scientific workforce is starting to develop as basic building blocks of the modern day scientific method. Scientists using Big Data are increasingly dependent on knowledge of computational skillsets or on having access to technical experts in all aspects of the scientific method (e.g., data gathering, data generation, data collection, data storage, data processing, data analysis, data verification,

data representation, data sharing, data preservation, etc.). They also find themselves leveraging crowd workers who may not possess relevant scientific knowledge to provide a ground truth label of large datasets, known as “Human-in-the-Loop” (HITL) machine learning [18, 82], and scientists to correct data errors and fine-tune algorithms, known as “Interactive Machine Learning” (IML) [26, 107]. In fact, the incumbent skillsets have become so necessary and in such high demand that the White House has issued a call for a Science, Technology, Engineering, and Mathematics (STEM) initiative to make these areas of inquiry and practice more accessible to the general public [45]. Fig. 3.2 provides an overview of this newer, emerging process of scientific inquiry.

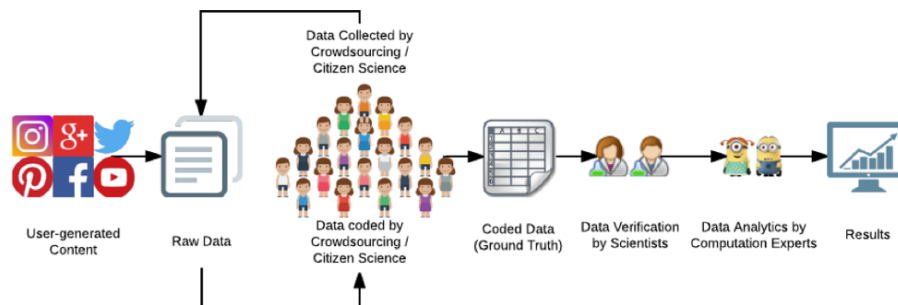


Fig. 3.2 Emerging Process of Scientific Inquiry.

Although the demand for STEM skillsets is increasing, enrolment in computer science has remained stagnant [2], which may be attributable to perceived race and gender stereotypes, or unequal access to computer science education [44, 108]. Computer education researchers have investigated how to effectively integrate computational thinking into education in order to craft, “the thought processes involved in formulating a problem and expressing its solution(s) in such a way that a computer-human or machine-can effectively carry out” [112, 111]. One approach involves motivating student interests with gamification [23, 33]. Another approach focuses on removing the technical barrier to content creation with user-friendly End-User Development (EUD) platforms [28, 63]. The latter view includes the belief that end-users with little or no technical expertise will be more willing to participate in tinkering, hacking, or other STEM activities if the barrier to entry is lowered. This research follows this second approach by proposing an end-user data analytics paradigm to broaden the population of researchers involved in this work, extending prior efforts to make computationally complex data analytics algorithms more accessible to end-users. This exploratory study focuses on examining the impact of interface design for eliciting data input from end-users as a segue into future work that will generate insights for designing end-user data analytics mechanisms.

The initial goal of this research is to create a transparent machine learning platform prototype to assist the scientists and end-users in processing and analysing real-time data streams and to understand opportunities and challenges of develop-

ing an end-user data analytics paradigm for future scientific workforces. Ultimately, the goal is to empower scientists and end-users to train supervised machine learning models to pre-process other sensor and device data streams along with those from cameras, and interactively provide feedback to improve model prediction accuracy. In this sense, the proposed end-user data analytics paradigm replaces human observers taking and coding data by hand with computational labour, where scientists or trained observers become end-users training the system by providing the system with ground truth labels for the data. In the process, the system frees the scientists from having to depend on highly technical programming expertise. In the context of a scientific workforce, this could potentially replace the onerous, labour-intensive system commonly used in observation research around the world. The same applies to domain applications with similar care and monitoring mandates, such as nursing homes, hospital intensive care units, certain security and military-related environments, and space and deep sea exploration vessels. Fig. 3.3 provides an overview of the proposed end-user data analytics paradigm.

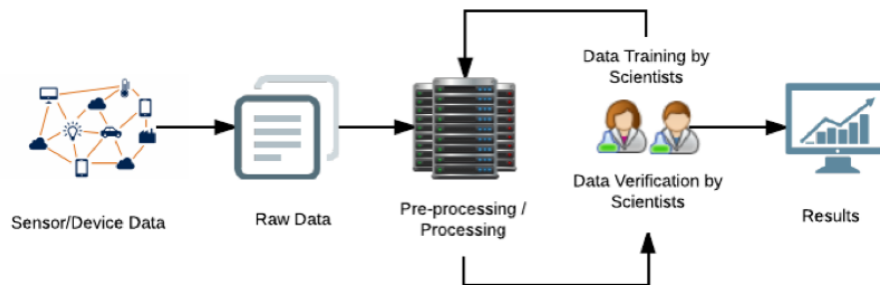


Fig. 3.3 Proposed Process of Scientific Inquiry (End-User Data Analytics).

3.2 Background

The emergence, adoption, and advances of ICTs in the past several decades have revolutionised the scientific method and the process of scientific inquiry. This section provides a general overview of the roles of ICTs in scientific inquiry along two dimensions: the scientific domain expertise of the users and the technical functions of the ICT platforms. ICT use in the scientific workforce evolved from collaboratories in the late 1980s that created communication infrastructures for scientists to share resources and early results, to citizen science platforms in the 1990s that allowed the public to contribute to scientific data collection, analysis, and interpretation. The citizen science platforms have led more recently to crowdsourcing platforms that allow online crowd workers to analyse modularised datasets (e.g., human computation and HITL machine learning). The proposed end-user data analytics platform—a transparent machine learning platform prototype that assists animal behavioural sci-

entists to analyse multi-channel high-definition video camera data-is an effort to now provide scientists with computational capabilities to process and analyse large datasets. Fig. 3.4 shows the overview of ICT use in scientific inquiry.

		Functions of ICTs	
		Infrastructure & Information Sharing	Coordination & Computation
Scientific Domain Expertise	Low Expertise	Citizen Science (1990's)	Crowdsourcing, Human Computation, & HITL Machine learning (2000's)
	High Expertise	<u>Collaboratory</u> (1980's)	End-user Data Analytics (2010's)




Fig. 3.4 Overview of ICT Use in Scientific Inquiry.

3.2.1 Collaboratory and Large-scale Scientific Workforce

The term “collaboratory” was coined by William Wulf while he worked for the National Science Foundation by merging the notion of traditional laboratory and collaboration that is afforded by ICT platforms that emerged in the late 1980s [60]. The shift in scientific inquiry occurred naturally out of the need to overcome physical limitations of instrument, infrastructure, and information sharing, such as results collected by scarce research instruments [1], or annotated electronic editions of 16th-century manuscripts [50]. Bos et al. (2007) describes a taxonomy of seven types of collaboratories that are differentiated by the nature of activities (loose coupling & asynchronous vs. tight coupling, synchronous) and resource needs (infrastructure and research instruments, open data, and virtual learning and knowledge communities) [9]. The early collaboratory platforms typically included functionalities such as electronic whiteboards, electronic notebooks, chatrooms, and video conferencing to facilitate effective coordination and interactions between dispersed scientists in astrophysics, physics, biology, medicine, chemistry, and the humanities [27].

3.2.2 Citizen Science

Citizen science is a two-part concept that focuses on (1) opening science and science policy processes to the public and (2) public participation in scientific projects under the direction of professional scientists [80]. Unfortunately, discussions of the public understanding of science tend to dismiss citizen expertise as uninformed or irrational, and some have advocated for involving the public in citizen projects to facilitate more sustainable development of the relationship between science, society,

and the environment [52, 53]. Although research has attempted to involve the public in citizen science projects, without proper research framing and training prior to the project, most people will not recognise scientifically relevant findings [17, 96]. Citizen science projects are also limited to those that could be broken down into modular efforts in which laypeople could reasonably participate [96]. This limits the complexity of the projects that citizens could participate in. There have been reports of mild success in terms of scientific discoveries, but the actual impact of involving citizens in scientific projects remains fairly minimal [10]. The intent in many citizen science projects is to involve volunteers in data collection or interpretation, such as the large volume of video data of animals at the zoo, that are difficult for scientists to process. These citizen science efforts are viewed as “complementary to more localized, hypothesis-driven research” [20]. Nonetheless, citizen science is generally seen as a positive factor in raising awareness of science and is frequently used as a mechanism for engaging people in civic-related projects [8, 25]. Earlier citizen science platforms typically employed traditional technologies that are commonly found in asynchronous collaboratories mentioned in the previous section [9, 27]. Modern citizen scientist platforms are starting to incorporate features found in common crowdsourcing platforms [99], and those will be described in the section below.

3.2.3 Crowdsourcing, Human Computation, and Human-in-the-Loop

Although citizen science taps into people’s intrinsic motivation to learn and contribute to science by providing labour for scientific inquiry, other crowdsourcing platforms have emerged as a way for people to outsource other kinds of labour at an affordable cost [46]. Research has linked gamification to crowdsourcing projects—if people can be incentivised to spend countless hours on playing highly interactive and engaging video games, this motivation can be harnessed as free work using progress achievement and social recognition [30, 31, 33, 85, 98]. Proponents also argue that if a task can be broken down finely enough, anyone can spend just a short moment to complete a simple task while also making a little bit of extra income. As such, crowdsourcing and human computation platforms primarily focus on task structure and worker coordination relating to workflow, task assignment, hierarchy, and quality control [58], whereas communication features between clients and workers and among workers themselves are practically nonexistent [51].

In terms of getting citizens to contribute to science projects, research has leveraged crowd workers on crowdsourcing platforms to provide ground truth label of large datasets to improve HITL prediction models [18, 26, 82, 107]. In contrast with the citizen science platforms that typically fulfil workers’ desires for educational or civic engagement activities, workers on crowdsourcing platforms are typically underpaid and have no opportunity to learn or become more engaged with the project after task completion [58]. The ethics of crowdsourcing platforms are heavily de-

bated for these reasons [42, 51, 81]. These platforms have also sparked a growth of peer-to-peer economy platforms that undercut existing worker wages [67].

3.2.4 From Human-in-the-Loop to Transparent Machine Learning

With the increase in available user-generated content and sensor data along with significant improvement in computing infrastructure, machine learning algorithms are being used to create prediction models that both recognise and analyse data. HITL machine learning attempts to leverage the benefits of human observation and categorisation skills as well as machine computation abilities to create better prediction models [18, 26, 82, 107]. In this approach, humans provide affordable ground truth labels while the machine creates models based on the humans' labels that accurately categorise the observations. However, HITL machine learning suffers similar issues of crowdsourcing and citizen science platforms. For example, similar to the workers on crowdsourcing platforms, the human agents in these cases are typically used to simply complete mundane work without deriving any benefits from participation in the project. In addition, human labels suffer from errors and biases [61, 62]. Similar to the participants of the citizen science program, crowd workers are prone to making incorrect labels without domain knowledge and proper research training and framing. Accuracy in the correct identification of data and the training of the system remain two major issues in the field of HITL machine learning and machine learning as a field in general [5, 61, 62, 78]. To empower scientists with mitigating the aforementioned issues, a research agenda on an end-user data analytics paradigm is necessary for investigating issues relating to the design, implementation, and use of a transparent machine learning platform prototype to make computationally complex data analytics algorithms more accessible to end-users with little or no technical expertise.

3.3 Impact of Interface Design for Eliciting Data Input from End-Users

The goal of this research is to learn about the barriers that scientists and end-users face in conducting data analytics and to discover what kinds of interaction techniques and end-user technological platforms will help them overcome these barriers. As an initial step to understand current problems and practices that scientists and end-users encounter throughout the data analytics process, the following experiment was conducted to demonstrate the impact of interface design for eliciting data input from end-users.

The experiment uses NeuralTalk2 [57, 102], a deep learning image caption generator, to generate 5 most likely captions for each of 9 images. In a between-subject experiment, a total of 88 college students were randomly assigned to one of the

three interface groups—Yes/No (31 students), multiple-selection (34), and open-ended questions (23).

- In the Yes/No group, participants answered whether the generated caption accurately describes an image. This was repeated for all 5 captions for each of the 9 images, totalling 45 questions.
- In the multiple-selection group, all five captions were presented to the participants at the same time. The participants were asked to select all the captions that accurately described an image. This was repeated for all 9 images, totalling 9 questions.
- In the open-ended group, participants were asked to describe what they saw in an image. This was repeated for all 9 images, totalling 9 questions.

Participants were asked to rate their confidence level after answering each question. Participants' feedback accuracy was assessed manually after the experiment was conducted. Selection consensus across participants and time spent were also compared and analysed. Fig. 3.5 illustrates the design of the experimental conditions. Below are results that detail how different feedback interfaces influence feedback accuracy, feedback consensus, confidence level, and time spent in how participants provide feedback to machine learning models.

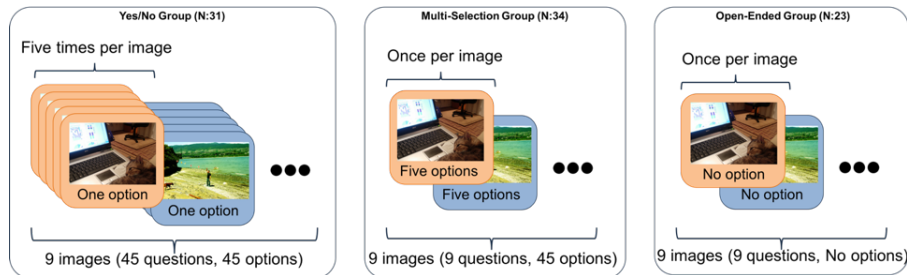


Fig. 3.5 Experimental Design.

Fig. 3.6 illustrates the feedback accuracy of the captions selected by the participation. An ANOVA test followed by post-hoc comparisons revealed that the open-ended group produced higher feedback accuracy than both the Yes/No group and the multiple-selection group, and the Yes/No group outperformed the multiple-selection group ($F(2,85) = 20.44, p < .0001$).

Although the feedback accuracy varied significantly across groups, participants achieved similarly high within-group consensus across all 3 conditions (non-sig., see Fig. 3.7). This indicates that the differences in the feedback provided by the participants were indeed caused by the interface design conditions.

In terms of feedback confidence, although the open-ended group provided the highest level of feedback accuracy, their self-perceived confidence level ($U = 372.5, p < 0.05$) is as low as the multiple-selection group ($U = 197.5, p < 0.01$) when compared to the Yes/No group. Fig. 3.8 shows that the Yes/No group reported the highest

self-perceived confidence level. This is likely due to the fact that there leaves less room for self-doubt when the participants are presented with only Yes/No options.

Fig. 3.9 illustrates the difference in time spent for providing feedback across the 3 groups. It took the Yes/No group significantly more time to rate 45 captions (5 per 9 images) than the multiple-selection group ($F(2,85) = 6.15, p < 0.05$), whereas there is no significant difference between the open-ended and the multiple-selection groups. This is likely due to the fact that the captions in the Yes/No group were presented across a series of 45 questions instead of 9 questions presented to the multiple-selection and the open-ended groups.

Based on the results presented above, future transparent machine learning research should account for the following trade-offs when eliciting user feedback.

- The open-ended group achieved the highest level of feedback accuracy, and the participants also reported the highest level of confidence in their feedback. The fact that this can be accomplished within a similarly short time frame as the multiple-selection group points to the potential of utilising an open-ended form to elicit user feedback when the task demands such high level of accuracy. The biggest drawback is that the open-ended feedback requires active human involvement to interpret the data. A future transparent machine learning model could utilise current state-of-the-art natural language processing efforts to pre-process the open-ended responses to generate a list of possible labels before a second round of human coding. This essentially reduces the effort of analysing open-

Fig. 3.6 Feedback Accuracy: Open-ended > Yes/No > Multiple-selection.

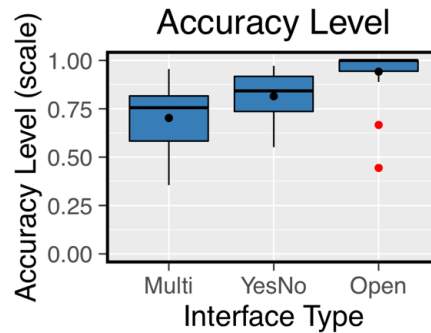


Fig. 3.7 Feedback Consensus.

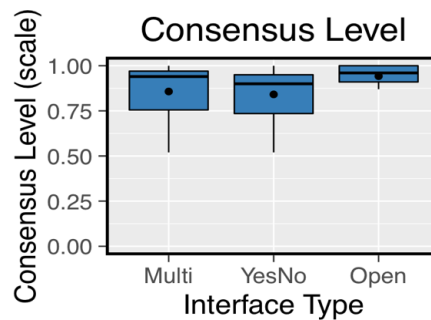


Fig. 3.8 Feedback Confidence: Yes/No > Open-ended and Multiple-selection.

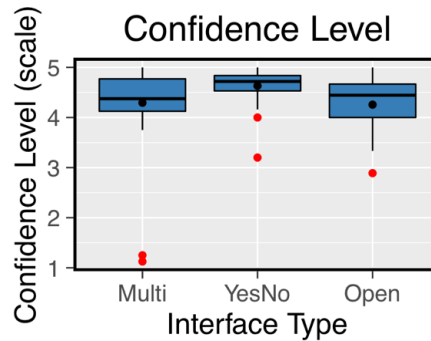
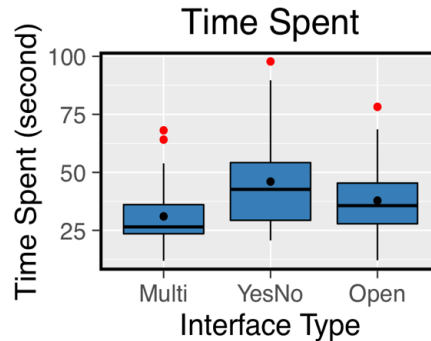


Fig. 3.9 Time Spent on Providing Feedback: Yes/No > Multiple-selection.



ended responses into two rounds of Yes/No or multiple-selection coding efforts for the users. The cumulative time spent in the proposed multi-round effort will not greatly exceed that of the Yes/No group based on the results demonstrated in this experiment, and the superb accuracy may justify the usage of the multi-round effort in some cases.

- While the multiple-selection group may appear to be promising due to the ease of data processing of user feedback relative to the open-ended group, the results show that it produced the lowest feedback accuracy and the participants are less confident of their feedback. One advantage of this user feedback elicitation method is that it gives the users the ability to view and provide feedback on multiple machine-generated labels at the same time, which results in the lowest cumulative time spent for the participants in our experiment. This method may be desirable in situations where feedback accuracy is less critical and the goal is to process through a large amount of data in a short period of time.
- The Yes/No group produced the medium level of feedback accuracy. Although the participants in the Yes/No group spent the highest cumulative time to provide feedback, it took the participants much less time to rate the individual options with the highest self-reported confidence level compared to the multiple-selection and the open-ended groups. The flexibility of adjusting the number of options that the users rate at any given time (e.g., users can stop after rating through 2 options instead of having to view all of the options at once in the

multiple-selection group) can be especially desirable when user commitment is unknown and the intention is to minimise user burden to provide feedback. The human-labelled results are also easy for machine learning models to process, making the Yes/No group the most flexible and adaptable method.

These experimental findings show that interface design significantly affects how end-users transform information from raw data into codified data that can be processed using data analytics tools, and the insights can inform the design, implementation, and evaluation of a usable transparent machine learning platform in the future. Future transparent machine learning could expand the study to different user feedback scenarios and contexts that require human feedback.

3.4 Design for End-User Data Analytics

Currently, there are many popular, general-purpose open-source scientific numerical computation software libraries such as NumPy [103], Matplotlib [48], and Pandas [70] that users can import into their software development environment to conduct numerical analysis programmatically. However, the use of these software libraries requires significant programming knowledge. To make data analytics more user-friendly, popular machine learning and data mining software suites such as Weka [32, 113], Orange [19], KNIME [7], and Caffe [56] provide users with command-line and/or graphical user interfaces to access a collection of visualisation tools and algorithms for data analysis and predictive modelling. Yet these software suites do not provide label suggestions based on the currently trained model, typically operating under the assumption that ground truth labels are error-free. Functionalities of these software suites are typically limited to training static rather than real-time live-stream datasets and lack the ability to allow users to interactively train machine learning models in order to more effectively explore data trends and correct label errors. In other words, these platforms neglect data collection and data (pre-)processing phases, both of which are essential steps throughout data analytics. A new paradigm is needed to more effectively disseminate the data science mindset more holistically and make data analytics more accessible to learners and end-users.

To realise intuitive, easy-to-learn, and user-friendly interfaces for data collection, processing, and analytics, it is necessary to create a series of software front-end prototypes, increasing in complexity but all sharing the same basic framework for interaction. The goal of the prototypes will be to learn about how different interaction techniques can replace or enhance the current paradigm of data processing by scientists and end-users. In the spirit of end-user development paradigms such as Scratch [79], combining interaction techniques used in interactive machine learning [26, 107] and direct manipulation interfaces [49] to create a novel interface to ease the training process of supervised learning models could potentially yield a more usable transparent machine learning platform. The goal is to create a system that allows the user to smoothly move between data and a list of inferred behaviours, allowing scientists and end-users to visually preview and make corrections to the

prediction model. Although the prototypes will vary, the interactions will share the same basic features. Users will use the platform to select the input data streams to be worked and then overlay these with behavioural data previously coded by trained scientists and end-users.

3.5 Conclusion

Successful investigations of transparent machine learning require multidisciplinary expertise in (1) human-computer interaction and end-user oriented design processes such as participatory design, interaction design, and scenario-based design [12, 14, 59, 64, 83, 92, 95, 97, 104, 3, 4, 66, 65, 24, 29, 74], (2) human computation and crowdsourcing [6, 37, 86, 90, 100, 105, 106, 114, 38, 22, 75, 11, 36, 40, 41, 13], (3) end-user visualisation interfaces and computational data analytics [35, 39, 54, 55, 73, 88, 89, 91, 93, 110, 116, 101, 72, 71, 43, 34, 87, 94], and (4) computer science education [15, 76, 77, 115, 117, 118, 69]. This research reveals the initial insights on how to make data analytics more accessible to end-users, to empower researchers in scientific inquiry, and to involve the public in citizen science. This research also will provide trained end-users opportunities to participate in citizen science efforts, allowing them to contribute directly to citizen science as well as become more familiar with the scientific method and data literacy, heightening awareness of how STEM impacts the world.

There are numerous potential applications of this work. Sensor and surveillance technologies have made great strides in behaviour profiling and behavioural anomaly detection. Such technologies may allow scientists and end-users to closely observe real-time data streams around the clock. Although the proposed end-user data analytic and transparent machine learning platform is currently targeted toward scientists and end-users, the platform and the resulting knowledge could be used most immediately to make data analytics more accessible for other domain applications with similar care and monitoring mandates, such as nursing homes, hospital intensive care units, certain security and military-related environments, and space and deep sea exploration vessels.

References

1. Abramovici, A., Althouse, W.E., Drever, R.W., Gürsel, Y., Kawamura, S., Raab, F.J., Shoemaker, D., Sievers, L., Spero, R.E., Thorne, K.S., et al.: Ligo: The laser interferometer gravitational-wave observatory. *Science* **256**(5055), 325–333 (1992)
2. Association, C.R., et al.: Taulbee survey. *Comput. Res. News* **28**(5), 19 (2015)
3. Baglione, A.N., Girard, M.M., Price, M., Clawson, J., Shih, P.C.: Mobile technologies for grief support: Prototyping an application to support the bereaved. In: *Workshop on Interactive Systems in Health Care* (2017)

4. Baglione, A.N., Girard, M.M., Price, M., Clawson, J., Shih, P.C.: Modern bereavement: A model for complicated grief in the digital age. In: ACM Conference on Human Factors in Computing Systems (2018)
5. Batista, G.E., Prati, R.C., Monard, M.C.: A study of the behavior of several methods for balancing machine learning training data. *ACM SIGKDD explorations newsletter* **6**(1), 20–29 (2004)
6. Bellotti, V.M., Cambridge, S., Hoy, K., Shih, P.C., Handalian, L.R., Han, K., Carroll, J.M.: Towards community-centered support for peer-to-peer service exchange: rethinking the time-banking metaphor. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 2975–2984. ACM (2014)
7. Berthold, M.R., Cebron, N., Dill, F., Gabriel, T.R., Kötter, T., Meinl, T., Ohl, P., Thiel, K., Wiswedel, B.: Knime-the konstanz information miner: version 2.0 and beyond. *AcM SIGKDD explorations Newsletter* **11**(1), 26–31 (2009)
8. Bonney, R., Cooper, C.B., Dickinson, J., Kelling, S., Phillips, T., Rosenberg, K.V., Shirk, J.: Citizen science: a developing tool for expanding science knowledge and scientific literacy. *BioScience* **59**(11), 977–984 (2009)
9. Bos, N., Zimmerman, A., Olson, J., Yew, J., Yerkie, J., Dahl, E., Olson, G.: From shared databases to communities of practice: A taxonomy of collaboratories. *Journal of Computer-Mediated Communication* **12**(2), 652–672 (2007)
10. Brossard, D., Lewenstein, B., Bonney, R.: Scientific knowledge and attitude change: The impact of a citizen science project. *International Journal of Science Education* **27**(9), 1099–1121 (2005)
11. Carroll, J.M., Shih, P.C., Han, K., Kropczynski, J.: Coordinating community cooperation: Integrating timebanks and nonprofit volunteering by design. *International Journal of Design* **11**(1), 51–63 (2017)
12. Carroll, J.M., Shih, P.C., Hoffman, B., Wang, J., Han, K.: Presence and hyperpresence: implications for community awareness. *Interacting with presence: HCI and the sense of presence in computer-mediated environments* pp. 70–82 (2014)
13. Carroll, J.M., Shih, P.C., Kropczynski, J.: Community informatics as innovation in sociotechnical infrastructures. *Journal of Community Informatics* **11**(2) (2015)
14. Carroll, J.M., Shih, P.C., Kropczynski, J., Cai, G., Rosson, M.B., Han, K.: The internet of places at community-scale: Design scenarios for hyperlocal. *Enriching Urban Spaces with Ambient Computing, the Internet of Things, and Smart City Design* (2016) **1** (2016)
15. Carroll, J.M., Wu, Y., Shih, P.C., Zheng, S.: Re-appropriating a question/answer system to support dialectical constructivist learning activity. *Educational Technology Research and Development* **64**(1), 137–156 (2016)
16. Chen, H., Chiang, R.H., Storey, V.C.: Business intelligence and analytics: from big data to big impact. *MIS quarterly* pp. 1165–1188 (2012)
17. Cohn, J.P.: Citizen science: Can volunteers do real research? *AIBS Bulletin* **58**(3), 192–197 (2008)
18. Dautenhahn, K.: The art of designing socially intelligent agents: Science, fiction, and the human in the loop. *Applied artificial intelligence* **12**(7-8), 573–617 (1998)
19. Demšar, J., Curk, T., Erjavec, A., Gorup, Č., Hočevar, T., Milutinovič, M., Možina, M., Polajnar, M., Toplak, M., Starič, A., et al.: Orange: data mining toolbox in python. *The Journal of Machine Learning Research* **14**(1), 2349–2353 (2013)
20. Dickinson, J.L., Zuckerberg, B., Bonter, D.N.: Citizen science as an ecological research tool: challenges and benefits. *Annual review of ecology, evolution, and systematics* **41**, 149–172 (2010)
21. Dictionary, O.E.: Oed online (2015)
22. Ding, X., Shih, P.C., Gu, N.: Socially embedded work: A study of wheelchair users performing online crowd work in china. In: Proceedings of the ACM Conference on Computer Supported Cooperative Work & Social Computing, pp. 642–654. ACM (2017)
23. Domínguez, A., Saenz-De-Navarrete, J., De-Marcos, L., Fernández-Sanz, L., Pagés, C., Martínez-Herráiz, J.J.: Gamifying learning experiences: Practical implications and outcomes. *Computers & Education* **63**, 380–392 (2013)

24. Dunbar, J.C., Connelly, C.L., Maestre, J.F., MacLeod, H., Siek, K., Shih, P.C.: Considerations for using the asynchronous remote communities (arc) method in health informatics research. In: *Workshop on Interactive Systems in Health Care* (2017)
25. Evans, C., Abrams, E., Reitsma, R., Roux, K., Salmonsens, L., Marra, P.P.: The neighborhood nestwatch program: Participant outcomes of a citizen-science ecological research project. *Conservation Biology* **19**(3), 589–594 (2005)
26. Fails, J.A., Olsen Jr, D.R.: Interactive machine learning. In: *Proceedings of the 8th international conference on Intelligent user interfaces*, pp. 39–45. ACM (2003)
27. Finholt, T.A., Olson, G.M.: From laboratories to collaboratories: A new organizational form for scientific collaboration. *Psychological Science* **8**(1), 28–36 (1997)
28. Fischer, G., Giaccardi, E., Ye, Y., Sutcliffe, A.G., Mehandjiev, N.: Meta-design: a manifesto for end-user development. *Communications of the ACM* **47**(9), 33–37 (2004)
29. Gao, G., Min, A., Shih, P.C.: Gendered design bias: Gender differences of in-game character choice and playing style in league of legends. In: *Australian Conference on Computer-Human Interaction*. ACM Press (2017)
30. Ghose, A., Ipeirotis, P.G., Li, B.: Designing ranking systems for hotels on travel search engines by mining user-generated and crowdsourced content. *Marketing Science* **31**(3), 493–520 (2012)
31. Goncalves, J., Hosio, S., Ferreira, D., Kostakos, V.: Game of words: tagging places through crowdsourcing on public displays. In: *Proceedings of the 2014 conference on Designing interactive systems*, pp. 705–714. ACM (2014)
32. Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., Witten, I.H.: The weka data mining software: an update. *ACM SIGKDD explorations newsletter* **11**(1), 10–18 (2009)
33. Hamari, J., Koivisto, J., Sarsa, H.: Does gamification work?—a literature review of empirical studies on gamification. In: *System Sciences (HICSS), 2014 47th Hawaii International Conference on*, pp. 3025–3034. IEEE (2014)
34. Han, K., Cook, K., Shih, P.C.: Exploring effective decision making through human-centered and computational intelligence methods. In: *ACM Conference on Human Factors in Computing Systems: Workshop on Human-Centred Machine Learning* (2016)
35. Han, K., Shih, P., Carroll, J.M.: Aggregating community information to explore social connections. In: *When The City Meets The Citizen Workshop, ICWSM 2013: 7th International AAAI Conference On Weblogs And Social Media*, pp. 8–11 (2013)
36. Han, K., Shih, P.C., Bellotti, V., Carroll, J.M.: Timebanking with a smartphone application. In: *Collective Intelligence Conference* (2014)
37. Han, K., Shih, P.C., Bellotti, V., Carroll, J.M.: It’s time there was an app for that too: A usability study of mobile timebanking. *International Journal of Mobile Human Computer Interaction (IJMHCI)* **7**(2), 1–22 (2015)
38. Han, K., Shih, P.C., Beth Rosson, M., Carroll, J.M.: Understanding local community attachment, engagement and social support networks mediated by mobile technology. *Interacting with Computers* **28**(3), 220–237 (2014)
39. Han, K., Shih, P.C., Carroll, J.M.: Local news chatter: augmenting community news by aggregating hyperlocal microblog content in a tag cloud. *International Journal of Human-Computer Interaction* **30**(12), 1003–1014 (2014)
40. Han, K., Shih, P.C., Rosson, M.B., Carroll, J.M.: Enhancing community awareness of and participation in local heritage with a mobile application. In: *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing*, pp. 1144–1155. ACM (2014)
41. Hanna, S.A., Kropczynski, J., Shih, P.C., Carroll, J.M.: Using a mobile application to encourage community interactions at a local event. In: *ACM Richard Tapia Celebration of Diversity in Computing Conference* (2015)
42. Hansson, K., Muller, M., Aitamurto, T., Irani, L., Mazarakis, A., Gupta, N., Ludwig, T.: Crowd dynamics: Exploring conflicts and contradictions in crowdsourcing. In: *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, pp. 3604–3611. ACM (2016)

43. Hiler, L., Foulk, B., Nippert-Eng, C., Shih, P.C.: Detecting biological samples using olfactory sensors. In: *Animal Behavior Conference* (2017)
44. Hill, C., Corbett, C., St Rose, A.: *Why so few? Women in science, technology, engineering, and mathematics*. ERIC (2010)
45. House, W.: President obama to announce major expansion of educate to innovate campaign to improve science, technology, engineering and math (stem) education. Office of the Press Secretary (2010)
46. Howe, J.: The rise of crowdsourcing. *Wired magazine* **14**(6), 1–4 (2006)
47. Humphreys, P.: *Extending ourselves: Computational science, empiricism, and scientific method*. Oxford University Press (2004)
48. Hunter, J.D.: Matplotlib: A 2d graphics environment. *Computing in science & engineering* **9**(3), 90–95 (2007)
49. Hutchins, E.L., Hollan, J.D., Norman, D.A.: Direct manipulation interfaces. *Human-computer interaction* **1**(4), 311–338 (1985)
50. Ide, N., Véronis, J.: *Text encoding initiative: Background and contexts*, vol. 29. Springer Science & Business Media (1995)
51. Irani, L.C., Silberman, M.: Turkopticon: Interrupting worker invisibility in amazon mechanical turk. In: *Proceedings of the SIGCHI conference on human factors in computing systems*, pp. 611–620. ACM (2013)
52. Irwin, A.: *Citizen science: A study of people, expertise and sustainable development*. Psychology Press (1995)
53. Irwin, A.: Constructing the scientific citizen: science and democracy in the biosciences. *Public understanding of science* **10**(1), 1–18 (2001)
54. Jang, J.Y., Han, K., Lee, D., Jia, H., Shih, P.C.: Teens engage more with fewer photos: Temporal and comparative analysis on behaviors in instagram. In: *Proceedings of the 27th ACM Conference on Hypertext and Social Media*, pp. 71–81. ACM (2016)
55. Jang, J.Y., Han, K., Shih, P.C., Lee, D.: Generation like: comparative characteristics in instagram. In: *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, pp. 4039–4042. ACM (2015)
56. Jia, Y., Shelhamer, E., Donahue, J., Karayev, S., Long, J., Girshick, R., Guadarrama, S., Darrell, T.: Caffe: Convolutional architecture for fast feature embedding. In: *Proceedings of the 22nd ACM international conference on Multimedia*, pp. 675–678. ACM (2014)
57. Karpathy, A., Fei-Fei, L.: Deep visual-semantic alignments for generating image descriptions. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 3128–3137 (2015)
58. Kittur, A., Nickerson, J.V., Bernstein, M., Gerber, E., Shaw, A., Zimmerman, J., Lease, M., Horton, J.: The future of crowd work. In: *Proceedings of the 2013 conference on Computer supported cooperative work*, pp. 1301–1318. ACM (2013)
59. Koehne, B., Shih, P.C., Olson, J.S.: Remote and alone: coping with being the remote member on the team. In: *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work*, pp. 1257–1266. ACM (2012)
60. Kouzes, R.T., Myers, J.D., Wulf, W.A.: Collaboratories: Doing science on the internet. *Computer* **29**(8), 40–46 (1996)
61. Le, J., Edmonds, A., Hester, V., Biewald, L.: Ensuring quality in crowdsourced search relevance evaluation: The effects of training question distribution. In: *SIGIR 2010 workshop on crowdsourcing for search evaluation*, vol. 2126 (2010)
62. Lease, M.: On quality control and machine learning in crowdsourcing. *Human Computation* **11**(11) (2011)
63. Lieberman, H., Paternò, F., Klann, M., Wulf, V.: End-user development: An emerging paradigm. In: *End user development*, pp. 1–8. Springer (2006)
64. Liu, L.S., Shih, P.C., Hayes, G.R.: Barriers to the adoption and use of personal health record systems. In: *Proceedings of the 2011 iConference*, pp. 363–370. ACM (2011)
65. Maestre, J.F., MacLeod, H., Connelly, C.L., Dunbar, J.C., Beck, J., Siek, K., Shih, P.C.: Defining through expansion: Conducting asynchronous remote communities (arc) research with stigmatized groups. In: *ACM Conference on Human Factors in Computing Systems* (2018)

66. Maestre, J.F., Shih, P.C.: Impact of initial trust on video-mediated social support. In: Australian Conference on Computer-Human Interaction. ACM Press (2017)
67. Malhotra, A., Van Alstyne, M.: The dark side of the sharing economy and how to lighten it. *Communications of the ACM* **57**(11), 24–27 (2014)
68. McAfee, A., Brynjolfsson, E., Davenport, T.H., Patil, D., Barton, D.: Big data: the management revolution. *Harvard business review* **90**(10), 60–68 (2012)
69. McCoy, C., Shih, P.C.: Teachers as producers of data analytics: A case study of a teacher-focused educational data science program. *Journal of Learning Analytics* **3**(3), 193–214 (2016)
70. McKinney, W., et al.: Data structures for statistical computing in python. In: Proceedings of the 9th Python in Science Conference, vol. 445, pp. 51–56. Austin, TX (2010)
71. Min, A., Lee, D., Shih, P.C.: Potentials of smart breathalyzer: An interventions for excessive drinking among college students. In: Proceedings of the iConference (2018)
72. Min, A., Shih, P.C.: Exploring new design factors for electronic interventions to prevent college students from excessive drinking by using personal breathalyzers. In: Workshop on Interactive Systems in Health Care (2017)
73. Nelson, J.K., Shih, P.C.: Companionviz: Mediated platform for gauging canine health and enhancing human–pet interactions. *International Journal of Human-Computer Studies* **98**, 169–178 (2017)
74. Ongwere, T., Cantor, G., Martin, S.R., Shih, P.C., Clawson, J., Connelly, K.: Too many conditions, too little time: Designing technological intervention for patients with type-2 diabetes and discordant chronic comorbidities. In: Workshop on Interactive Systems in Health Care (2017)
75. Parry-Hill, J., Shih, P.C., Mankoff, J., Ashbrook, D.: Understanding volunteer at fabricators: Opportunities and challenges in diy-at for others in e-nable. In: ACM Conference on Human Factors in Computing Systems, pp. 6184–6194. ACM (2017)
76. Pena, J., Shih, P.C., Rosson, M.B.: Instructors as end-user developers: Technology usage opportunities in the inverted classroom. In: Handbook of Research on Applied Learning Theory and Design in Modern Education, pp. 560–571. IGI Global (2016)
77. Peña, J., Shih, P.C., Rosson, M.B.: Scenarior-based design of technology to support teaching in inverted classes. *ICConference 2016 Proceedings* (2016)
78. Rani, P., Liu, C., Sarkar, N., Vanman, E.: An empirical study of machine learning techniques for affect recognition in human–robot interaction. *Pattern Analysis and Applications* **9**(1), 58–69 (2006)
79. Resnick, M., Maloney, J., Monroy-Hernández, A., Rusk, N., Eastmond, E., Brennan, K., Millner, A., Rosenbaum, E., Silver, J., Silverman, B., et al.: Scratch: programming for all. *Communications of the ACM* **52**(11), 60–67 (2009)
80. Riesch, H., Potter, C.: Citizen science as seen by scientists: Methodological, epistemological and ethical dimensions. *Public understanding of science* **23**(1), 107–120 (2014)
81. Salehi, N., Irani, L.C., Bernstein, M.S., Alkhatib, A., Ogbe, E., Milland, K., et al.: We are dynamo: Overcoming stalling and friction in collective action for crowd workers. In: Proceedings of the 33rd annual ACM conference on human factors in computing systems, pp. 1621–1630. ACM (2015)
82. Schirner, G., Erdogmus, D., Chowdhury, K., Padir, T.: The future of human-in-the-loop cyber-physical systems. *Computer* **46**(1), 36–45 (2013)
83. Schumann, J., Shih, P.C., Redmiles, D.F., Horton, G.: Supporting initial trust in distributed idea generation and idea evaluation. In: Proceedings of the 17th ACM international conference on Supporting group work, pp. 199–208. ACM (2012)
84. Scriven, M., Paul, R.: Critical thinking as defined by the national council for excellence in critical thinking. In: 8th Annual International Conference on Critical Thinking and Education Reform, Rohnert Park, CA, pp. 25–30 (1987)
85. Seaborn, K., Fels, D.I.: Gamification in theory and action: A survey. *International Journal of human-computer studies* **74**, 14–31 (2015)

86. Shih, P.C., Bellotti, V., Han, K., Carroll, J.M.: Unequal time for unequal value: Implications of differing motivations for participation in timebanking. In: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, pp. 1075–1084. ACM (2015)
87. Shih, P.C., Christena, N.E.: From quantified self to quantified other: Engaging the public on promoting animal well-being. In: ACM Conference on Human Factors in Computing Systems: Workshop on HCI Goes to the Zoo (2016)
88. Shih, P.C., Han, K., Carroll, J.M.: Community incident chatter: Informing local incidents by aggregating local news and social media content. In: ISCRAM (2014)
89. Shih, P.C., Han, K., Carroll, J.M.: Community poll: Externalizing public sentiments in social media in a local community context. In: Second AAAI Conference on Human Computation and Crowdsourcing (2014)
90. Shih, P.C., Han, K., Carroll, J.M.: Engaging community members with digitally curated social media content at an arts festival. In: Digital Heritage, 2015, vol. 1, pp. 321–324. IEEE (2015)
91. Shih, P.C., Han, K., Carroll, J.M.: Using social multimedia content to inform emergency planning of recurring and cyclical events in local communities. *Journal of Homeland Security and Emergency Management* **12**(3), 627–652 (2015)
92. Shih, P.C., Han, K., Poole, E.S., Rosson, M.B., Carroll, J.M.: Use and adoption challenges of wearable activity trackers. *ICoNference 2015 Proceedings* (2015)
93. Shih, P.C., Nguyen, D.H., Hirano, S.H., Redmiles, D.F., Hayes, G.R.: Groupmind: supporting idea generation through a collaborative mind-mapping tool. In: Proceedings of the ACM 2009 international conference on Supporting group work, pp. 139–148. ACM (2009)
94. Shih, P.C., Olson, G.M.: Using visualization to support idea generation in context. In: ACM Creativity and Cognition Conference Workshop: Creativity and Cognition in Engineering Design. ACM (2009)
95. Shih, P.C., Venolia, G., Olson, G.M.: Brainstorming under constraints: why software developers brainstorm in groups. In: Proceedings of the 25th BCS conference on human-computer interaction, pp. 74–83. British Computer Society (2011)
96. Silvertown, J.: A new dawn for citizen science. *Trends in ecology & evolution* **24**(9), 467–471 (2009)
97. Su, N.M., Shih, P.C.: Virtual spectating: hearing beyond the video arcade. In: Proceedings of the 25th BCS conference on human-computer interaction, pp. 269–278. British Computer Society (2011)
98. Terveen, L., Hill, W.: Beyond recommender systems: Helping people help each other. *HCI in the New Millennium* **1**(2001), 487–509 (2001)
99. Tinati, R., Van Kleek, M., Simperl, E., Luczak-Rösch, M., Simpson, R., Shadbolt, N.: Designing for citizen data analysis: A cross-sectional case study of a multi-domain citizen science platform. In: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, pp. 4069–4078. ACM (2015)
100. Tomlinson, B., Ross, J., Andre, P., Baumer, E., Patterson, D., Corneli, J., Mahaux, M., Nobarany, S., Lazzari, M., Penzenstadler, B., et al.: Massively distributed authorship of academic papers. In: CHI'12 Extended Abstracts on Human Factors in Computing Systems, pp. 11–20. ACM (2012)
101. Vaghela, S.J.D., Shih, P.C.: Walksafe: College campus safety app. In: International Conference on Information Systems for Crisis Response and Management (2018)
102. Vinyals, O., Toshev, A., Bengio, S., Erhan, D.: Show and tell: A neural image caption generator. In: Computer Vision and Pattern Recognition (CVPR), 2015 IEEE Conference on, pp. 3156–3164. IEEE (2015)
103. Walt, S.v.d., Colbert, S.C., Varoquaux, G.: The numpy array: a structure for efficient numerical computation. *Computing in Science & Engineering* **13**(2), 22–30 (2011)
104. Wang, J., Shih, P.C., Carroll, J.M.: Life after weight loss: design implications for community-based long-term weight management. *Computer Supported Cooperative Work (CSCW)* **24**(4), 353–384 (2015)

105. Wang, J., Shih, P.C., Carroll, J.M.: Revisiting linuss law: Benefits and challenges of open source software peer review. *International Journal of Human-Computer Studies* **77**, 52–65 (2015)
106. Wang, J., Shih, P.C., Wu, Y., Carroll, J.M.: Comparative case studies of open source software peer review practices. *Information and Software Technology* **67**, 1–12 (2015)
107. Ware, M., Frank, E., Holmes, G., Hall, M., Witten, I.H.: Interactive machine learning: letting users build classifiers. *International Journal of Human-Computer Studies* **55**(3), 281–292 (2001)
108. Warschauer, M., Matuchniak, T.: New technology and digital worlds: Analyzing evidence of equity in access, use, and outcomes. *Review of research in education* **34**(1), 179–225 (2010)
109. Wiggins, A., Crowston, K.: From conservation to crowdsourcing: A typology of citizen science. In: *System Sciences (HICSS)*, 2011 44th Hawaii international conference on, pp. 1–10. IEEE (2011)
110. Williams, K., Li, L., Khabsa, M., Wu, J., Shih, P.C., Giles, C.L.: A web service for scholarly big data information extraction. In: *Web Services (ICWS)*, 2014 IEEE International Conference on, pp. 105–112. IEEE (2014)
111. Wing, J.: Computational thinking benefits society. 40th Anniversary Blog of Social Issues in Computing **2014** (2014)
112. Wing, J.M.: Computational thinking. *Communications of the ACM* **49**(3), 33–35 (2006)
113. Witten, I.H., Frank, E., Hall, M.A., Pal, C.J.: *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann (2016)
114. Wu, Y., Kropczynski, J., Shih, P.C., Carroll, J.M.: Exploring the ecosystem of software developers on github and other platforms. In: *Proceedings of the companion publication of the 17th ACM conference on Computer supported cooperative work & social computing*, pp. 265–268. ACM (2014)
115. Wu, Y., Shih, P.C., Carroll, J.M.: Design for supporting dialectical constructivist learning activities. In: *EDULEARN14 Proceedings*, pp. 4156–4164. IATED (2014)
116. Yang, S., Chen, P.Y., Shih, P.C., Bardzell, J., Bardzell, S.: Cross-strait frenemies: Chinese netizens vpn in to facebook taiwan. *Proceedings of the ACM on Human-Computer Interaction* **1**(CSCW), Article–115 (2017)
117. Zheng, S., Rosson, M.B., Shih, P.C., Carroll, J.M.: Designing moocs as interactive places for collaborative learning. In: *Proceedings of the Second (2015) ACM Conference on Learning@ Scale*, pp. 343–346. ACM (2015)
118. Zheng, S., Rosson, M.B., Shih, P.C., Carroll, J.M.: Understanding student motivation, behaviors and perceptions in moocs. In: *Proceedings of the 18th ACM conference on computer supported cooperative work & social computing*, pp. 1882–1895. ACM (2015)