Crowdsourcing and Human-in-the-loop for IoT

Luis-Daniel Ibáñez Neal Reeves Elena Simperl

January 2019

1 Introduction

Internet of Things (IoT) networks of sensors, mobile phones and other smart devices are providing researchers, practitioners, and end users with an unprecedented amount of data to enable new services, inform decisions and create added value. According to [36] the number of smart phone users was predicted to top three billion by the end of 2018. Other wearable devices such as watches, eyewear, and garments have become increasingly ubiquitous, with a projected 245 million units expected be sold in 2019 alone [22]. In the public sector, *smart cities* leverage IoT technologies to design better policies, create efficiencies, and manage growth sustainably [31]. Urban areas around the world have made substantial investments to deploy 'smart connections' for everything from buses to street lights to buildings, which fuel data analytics.

While developers have focused on improving sensor accuracy and devising advanced methods to store, manage and analyse IoT data, public authorities soon realised that technology is just one, albeit a crucial component of their smart city strategy, which could help them achieve their wider development goals and be more responsive towards residents' needs [35]. A smart city is hence commonly understood as a people-centric city, delivering services that matter to citizens and empowering communities and businesses to engage in decisions that will affect them.

Human involvement enhances technology as well. More than a decade of development in big data and data science, experts agree that the best solutions employ a combination of machine and manual processes [9] - for example, a state-of-the-art machine learning model can handle roughly 80% of a problem, while approximately 19% of cases require some form of human input, and the remaining 1% is random [4].

Augmenting technology is particularly helpful to:

1. Provide context to a device measurement, easing the detection of false positives and outliers. For example, for healthcare applications, considering biological measures alongside a description of the patient's activity at measurement time can be critical for taking the right treatment decision. In a more abstract sense, people *complement* devices with their own sensing capabilities.

2. Produce ground-truth datasets to train machine learning models. Machine learning techniques require large volumes of training data to be effectively calibrated. For example, an algorithm cannot recognize cars in traffic camera feeds unless it has enough examples of images with cars, and images without them. Manual feedback is also useful for verifying that a machine judgment is correct, especially in scenarios with a low margin of error, such as self-driving cars.

3. Gather data that is not available through other means. This can be achieved through *citizen sensing*, an approach which crowdsources sensor deployment and data collection to city residents, for example through apps that track locations or other smart devices, or through community projects, which reach out to residents to encourage them to participate in specific activities, such as OpenStreetMap.[1]

In this chapter we first introduce *crowdsourcing* (Section 2) and *human-in-the-loop* (Section 3), two related approaches for realising these three use cases and devise data science pipelines that seamlessly combine machine with human and collective intelligence. We then discuss two instances of crowdsourcing for location data: *spatial crowdsourcing* (Section 4), and *citizen sensing* (Section 5), which are particularly relevant in a smart city context.

2 Crowdsourcing

The term *crowdsourcing* is a portmanteau of the words *crowd* and *outsourcing*. Brabham [3] defines it as a production model that leverages the collective intelligence of online communities for specific purposes set forth by a 'requester' organisation. *Collective intelligence* is a capability that "emerges from the collaboration, collective efforts, and competition of many individuals and appears in consensus decision making" [30]. While crowdsourcing predates the digital age, the hyper-connectivity brought about by the rise of web and internet technologies has made it possible to mobilise large numbers of people in almost real-time, prompting the creation of on-demand platforms where people register to contribute to crowdsourcing projects, often in exchange of a reward.

Broadly speaking, there are two main categories of crowdsourcing activity: *microtask* and *macrotask* crowdsourcing. These can be distinguished based on task granularity, or the amount of work required for – and by extension, the complexity of – the task assigned to individual workers [43]. Microtasks are relatively quick, simple and repeatable activities that can be – and often are – completed by volunteers in parallel, without the need for specific training or specialist knowledge [11, 43]. Microtask crowdsourcing is particularly valuable as a means to combine human and machine intelligence, for example in the context of improving the performance of algorithms [9]. In contrast, macrotask crowdsourcing is a much more involved process, with tasks taking many

[1] https://www.openstreetmap.org

hours to complete and requiring specialist knowledge of the context in which the task is intended [20]. We present here *Kaggle* as an example of a macrotask crowdsourcing platform and *FigureEight* and *EyeWire* as examples of microtask crowdsourcing platforms.

In many areas, crowdsourcing has emerged as a suitable alternative to more established problem-solving approaches relying on experts. It can generate useful results quickly and at scale, provided it reaches a crowd with relevant skills and resources, enables them to contribute and coordinate effectively, and considers their motivations [21]. For example, Kaggle[2] houses a large community of data science and machine learning enthusiasts. Companies wanting to solve a problem in this space post it as a *competition* on Kaggle together with an evaluation metric and a number of rewards, usually monetary. Members of the community can participate in the challenge (individually or in teams), and are encouraged to engage in discussion forums. At the end of the competition, the winners earn prizes, while the organisation receives a solution to their problem.

Another example is FigureEight,[3], which targets the crowdsourcing of shorter, less complex tasks (called *microtasks*) that are nevertheless costly for an organisation to undertake with their own resources. For instance, an organisation wanting to train a machine learning model to recognize bikes in images needs a sizable amount of ground-truth data in the form of annotated images. In a similar way to Kaggle, an organisation can upload a dataset of non-annotated images together with human-readable instructions on what is required for each image. The task is posted on the platform, where any registered member can provide answers, with each answer being rewarded with a few cents. To provide a degree of certainty about the quality of the answers, the platform may ask different people to annotate the same image, and report the final answer based on the answer on which most contributors agreed.

Not all crowdsourcing incentives rely on monetary or physical rewards. In *virtual citizen science*, projects commonly rely on the intrinsic motivations of volunteers – their interest in science, altruism and desire to contribute to research. For example, EyeWire[4] is a VCS project that recruits volunteers to trace neuron pathways in Magnetic Resconance Imaging scans of the optic nerve [39]. EyeWire uses a range of incentives to recruit and retain volunteers, including gamification features such as points, badges and leaderboards; integrated discussion features such as an instant messenger chat service; feedback and discussion sessions with project scientists and regular competition events where volunteers work together or compete with one another to solve narratives and win in-game rewards such as bonus points [39, 44]. In terms of crowdsourcing mechanics, however, VCS projects are similar to other forms of crowdsourcing, with multiple volunteers independently making classifications and a final answer based on majority voting from volunteers [38].

[30] identified four dimensions of collective intelligence. They can be used as a checklist for any organisation interested in using crowdsourcing:

[2]`www.kaggle.com`

[3]`www.figureeight.com`

[4]`https://eyewire.org/`

What needs to be done?, and therefore What will people be asked to do?: [13] classified tasks published on microtask crowdsourcing platforms according to six types: information finding, verification and validation, interpretation and analysis of text or figures, content creation, surveying (customer satisfaction or demographic studies), and content access (e.g., to test a service). There is no hard constraint about what can be asked, but it is important to be aware of how difficult it is, how much time it would take on average, and if anything more than an ordinary PC or a mobile phone is needed for solving it. Macrotask crowdsourcing has slightly different challenges. While it can be used in any open-innovation context, practice has shown that it is important to be clear about the way the solutions are going to be evaluated [28], and consider how the most promising ones are going to be used.[5]

Who will do it?: what is the profile of the people that woudl potentially undertake the task? What skills are required? This is important to ensure the task is advertised with consideration for the appropriate challenges and reaches the right audience, just as when hiring a contractor. Crowdsourcing platforms manage profiles of contributors where they can update the description of their skillsets and track their performance. This information is available to organisations, giving them - to a certain extent - the ability to choose the most appropriate contributors to the task.

Why would someone do it?: in other words, what is the *incentive* for someone to do the task? [45] defines a typology of incentives, from financial to altruistic to reputation to enjoyment. In Kaggle, for instance, people receive points and badges based on their participation in machine learning challenges, which can be referred to in the LinkedIn CVs. These are examples of a broader set of techniques called *gamification* [41], which use game elements in non-game contexts to drive participation. **How to ensure the quality of the solutions?**: sometimes people err when performing microtasks, or deliberately give random answers to requests to receive rewards more quickly. Strategies to minimise the impact of unexpected or malevolent human behaviour are essential to ensure answer quality. As noted earlier, in the context of more complex *macrotasks*, a clear measure of how answers are evaluated gives the crowd a sense that they are participating under fair conditions and may increase participation [14, 53]. [7] presents a comprehensive survey of different quality control measures in crowdsourcing, separating them into *individual assessments*, that is, when an individual rates or evaluates a contribution; *group assessments*, when quality is assessed by a vote, peer-review aggregation, or through consensus in a group of people; and *computing assessments* when the answers can be checked automatically.

Once these questions have been answered, an organisation is ready to implement a workflow such as the one depicted in figure 1. The answers to the *What* questions should lead to a clear definition of the problem/task to be proposed, and to the design and implementation of a tool or interface through which peo-

[5] `https://www.forbes.com/sites/ryanholiday/2012/04/16/what-the-failed-1m-netflix-prize-tells-us-about-business-advice/`, retrieved January 14th 2019.

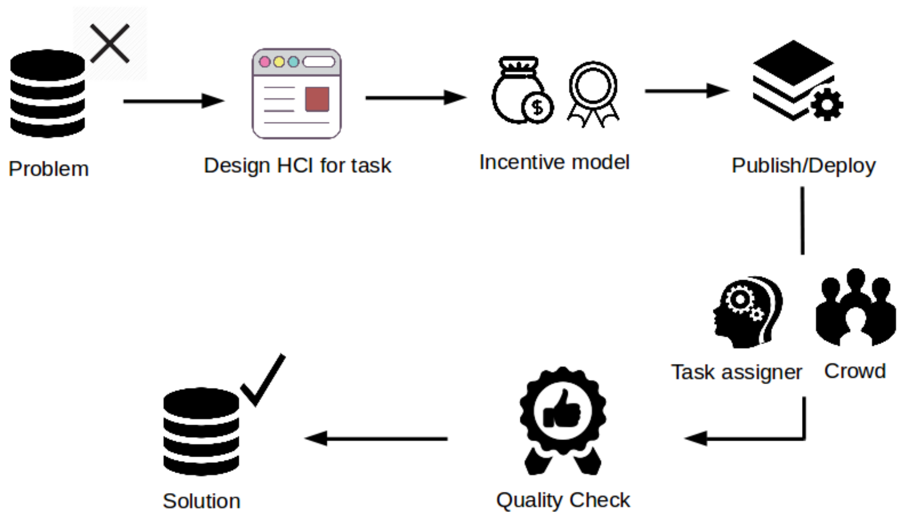

Figure 1: Workflow to crowdsource a task

ple can contribute. This may mean using existing crowdsourcing platforms, developing a solution in-house, for example a mobile app, or using social media and other marketing channels. The answers to the *Why* questions help define an appropriate incentive model for attracting contributions. The *Who* part determines where and when to launch or deploy the task, to which crowd, and how to execute the assignment of tasks to individual members of the crowd. Assignment can be as simple as simply making the task available to any member of the crowd, or consider performance on previous tasks, and people's preferences and availability. Finally, the *How* questions assist an organisation in building a better understanding of how success would be measured for the task and suggest developing tools and methods to manage quality effectively.

3 Human-in-the-loop

Human-in-the-loop (HITL) refers to a systems architecture which meets the following one of the criteria:

1. human interaction is a fundamental part of the workflow being implemented - in other words, the process cannot, for technical, legal, ethical or other reasons, be fully automated. There are many manifestations of this in IoT, for example smart city control rooms - they display large amounts of data and complex analyses to aid people to make decisions.

2. there is a case for the creation of a *loop* between machine output and human input and vice-versa. For instance, a health monitoring app would

commonly receive measurements from smart wearables alongside user-defined goals and suggest behavioural changes based on both. The 'human' in HITL refers to the user setting their health targets. The 'loop' consists of suggestions generated by a machine learning algorithm which are assessed by the user.

HITL and crowdsourcing are related, but there are important differences: crowdsourcing is a distributed problem-solving approach. It can be applied to computational or autonomous systems, but most forms of manual input assume that the participants are part of a large, unknown crowd and that they solve the problem collectively. As such, an important part of a crowdsourcing project is how to allocate the tasks to participants, and how to validate and aggregate their contributions. HITL does not necessarily involve decentralisation.

The work of [37] surveys HITL applications for IoT and cyber-physical systems. It proposes a taxonomy that divides applications based on whether they rely on *human control*, where the human either directly controls the system (e.g, a self-driving car) or supervises it (e.g, control in a factory); or *human monitoring* without direct control. The latter can further be classified into: *open-loop* and *closed-loop applications*. Open-loop refers to situations in which the system does not take any proactive action after collecting the data (e.g. a healthcare application that reports to medical staff). Closed-loop, by contrast, is about systems in which results are processed towards a common goal (e.g., exercise machines in a gym that monitor the body temperature of the people exercising, in order to adjust the temperature of the room). Hybrid systems combine control and monitoring in a single unit.

The first studies of HITL originated in control theory, where a large body of research has been devoted to human factors in complex systems (see for example [10]). Typically, humans are modeled as system components that introduce a degree of noise that the system needs to adapt to (e.g, driving-assistance systems), or as components that need to be given control of the system under certain conditions. The field of *human-robot interaction* [17] studies scenarios where people engage with robots, either remotely or in proximity. Robot interactions have challenges that do not exist with digital systems, such as robot autonomy, information exchange, and team-work both among robots, and among robots and people. HITL is also a common pattern in assistive technologies for helping people with disabilities, where the main question is how to derive intent from the sensor measures received from human participants [40]. In the machine learning community, a lot of attention has been recently put into integrating humans into the learning process, a technique called *interactive machine learning* [2], which builds upon the theoretical foundations of active, preferential, and reinforcement learning, and the theory and practice of HCI to speed up earning cycles and reduce the involvement of machine learning experts. This is an extension of the machine learning scenarios introduced in the previous section, where an undefined crowd, unrelated to the end-users of the classifier, is typically used to generate a gold standard, without any further interactions or 'loops'.

The four dimensions discussed in Section 2 apply to this case as well. In

HITL it is less common to seek input from multiple parties for the same task, a technique used in microtask and macrotask crowdsourcing. Many control or monitoring scenarios are designed for specific user groups, with their own motivations. The user often has an intrinsic interest in the system producing the correct output or a need to control the system. Microtask crowdsourcing has been used as a source of human input in interactive machine learning [8]. In those cases, participants are rewarded financially, just like for types of tasks e.g. completing surveys or curating databases. The challenge in HITL is how to effectively represent and integrate the inputs and outputs of human and machine components respectively to ensure a smooth operation of the overall system.

Figure 2 depicts a high-level view of the workflow of a human-in-the-loop application. It is consistent with the class of hybrid IoT applications from the taxonomy from [37]. Starting with data and sensor inputs, and possibly user input, collected by the automated part of the system, there is first a machine processing stage that produces an output. This output needs to be communicated to the user for verification, and the user needs to have the means to provide a meaningful output to the machine. For interactive machine learning, the output may be a simple validation of the classifier result. When crowdsourcing is used, the output would be collected from multiple *crowd workers* and then aggregated using automatic inference techniques that compute the most likely correct answer [7].

4 Spatial crowdsourcing

Sometimes a crowdsourcing task requires the presence at a particular location - in those cases, we talk about *spatial crowdsourcing* [55]. The crowd is equipped with smart devices, or if applicable use their own mobile phones. The goal is to collect or curate a geospatial dataset in a decentralised and distributed way. A typical task would be to visit specific locations and take measurements, e.g. take a picture, or to explore an area looking for items or events of interest which trigger measurements.

There are several platforms that offer bespoke support for spatial crowdsourcing projects, both at the macro and micro levels e.g., taskRabbit,[6] GigWalk,[7]) gmission, [5] i-Log [54] etc.

Figure 3 shows a spatial crowdsourcing workflow. Starting from a geospatial task, the first step that needs to be taken is deciding on the appropriate platform for the crowd to accomplish the task. For many types of task, existing phones with an internet connection, camera and sound recording capabilities will suffice. However, for other scenarios, special devices might be needed, making it important to consider how the devices would be distributed to the members of the crowd and financed. The platform affects the UX design of the task. For devices that are very specific to one task (e.g. radiation counters), one might consider

[6] `https://www.taskrabbit.com/`
[7] `http://www.gigwalk.com`

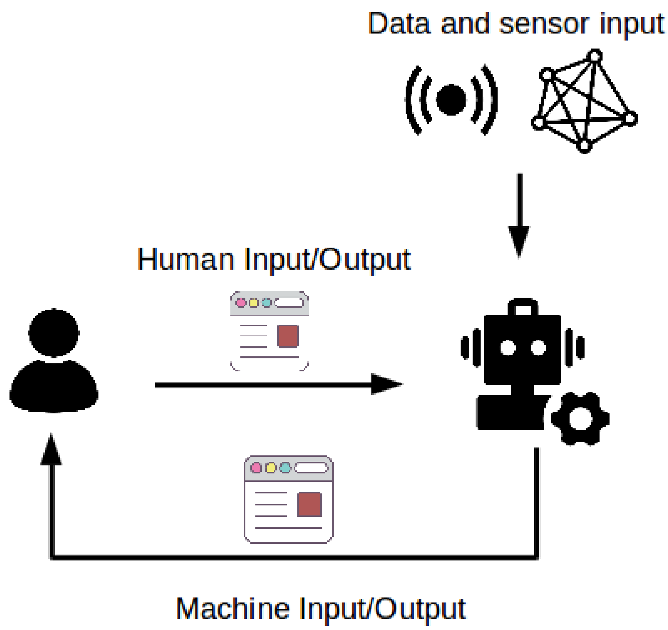

Figure 2: Human-in-the-loop workflow

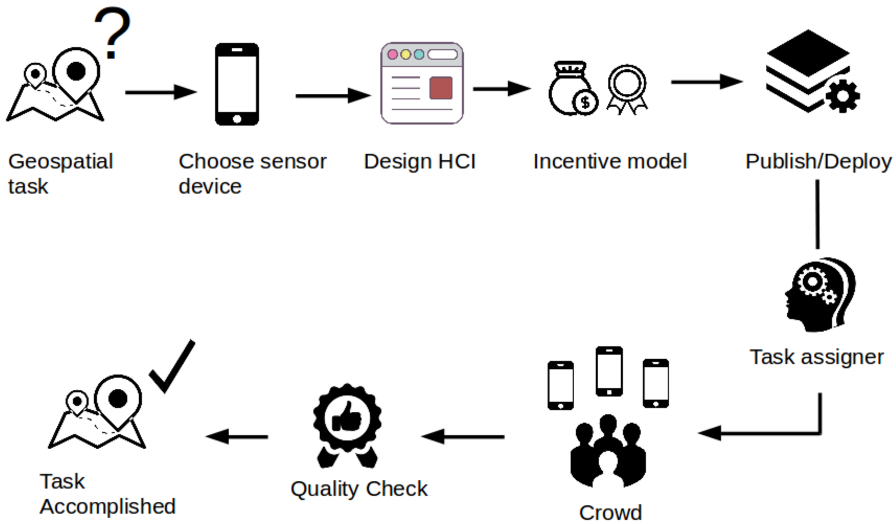

Figure 3: Spatial crowdsourcing workflow

the provision of an accompanying mobile app or website to better interact with contributors.

The workflow in Figure 3 shares many similarities to the one introduced in Section 2. The *allocation of tasks* has specific challenges in spatial crowdsourcing. First, the need to be at a location limits the size of the crowd substantially. Second, the relative location of the contributor to the location where the tasks need to be carried out influences the motivation of the contributor. To tackle these challenges, recent research has aimed at adapting well-known multi-agent and optimisation algorithms to handle uncertainty in the location of the crowd and their performance. The work of [48] assumes contributors send their locations to the server and thereafter the server assigns each of them tasks close by, with the aim of maximizing the overall number of assigned tasks and minimizing the effort required. However, unlike robots that can be programmed to stay put when idle, humans move around when they are waiting for a new task. The work of [49] addresses this issue by leveraging historical location traces to predict future spatio-temporal distributions of tasks, which are then used to guide idle contributors in a way that optimizes the overall allocation. A study by [26] takes into consideration the performance of individual members of the crowd and their task acceptance rates. By adding contextual tracking to the devices used to accomplish the task, a machine learning model is trained to predict both the likelihood of acceptance and the performance of a task before assigning it.

Spatial crowdsourcing has data privacy implications, which in the European Union are handled by the General Data Protection Regulation (GDPR). Contributors constantly report their location to the system that allocates tasks to them. Recent research has made use of a combination of differential privacy and geo-casting to allow a trusted cellular data provider to generate a *Private Spatial Decomposition* (PSD) of contributors' locations that is passed to the allocation unit [47]. Tasks are assigned based on the locations reported in the PSD by geo-casting to the zone with the highest probability of having the required level of crowd resources. The trade-off lies in the size of the zone to geo-cast versus the probability of broadcasting to an insufficient number of contributors, or to contributors who are too far away from the required location. A different approach resorts to obfuscating trajectories and locations based on global popularity and user preferences [24], following the same rationale as the randomised response method for handling sensitive questions in surveys, where the contributor chooses a number of erroneous locations proposed by the system. Compared to [47], this is more suited for crowdsourcing scenarios in relatively small areas (such as a university campus).

5 Participatory sensing

Participatory or *citizen* sensing describes the deployment of networks of mobile and other sensing devices to collect and subsequently share and analyse data [42]. It is a form of crowdsourcing which focuses on a specific type of activity and set of technologies and puts more emphasis on the role of the crowd and the participatory frameworks, tools and best practice [34].

Specific definitions and manifestations of participatory sensing are varied. In its original form, the term *participatory sensing* referred to a form of crowdsourced data gathering through the use of sensors and human observations. It is similar to spatial crowdsourcing, but with the distinction that participants would focus on gathering data within their local area - places they live, work and frequently visit [34]. Nevertheless, the field has subsequently grown to describe a wide variety of processes, both formal and informal. Initiatives vary from explicit to implicit activities from individual participants [15]. Rather than requiring formal data gathering processes, data may be gathered from web 2.0 services such as social networks to which individuals have unknowingly or unintentionally published valuable insights [6]. Sensing also does not necessarily involve *sensors* in the conventional technical sense, as participants may personally make observations themselves, with little or no help from technology.

Participatory sensing follows similar principles and approaches to *citizen science*, which engages volunteers in scientific research [52]. In fact, Haklay describes participatory sensing as a form of citizen science in which participant activity is passive and potentially implicit [34]. This requires a lower level of engagement from volunteers than some more complex forms of citizen science where participants have more agency to influence the way the data is collected and used. There are also differences in the incentives associated with these activities. Citizen science generally relies on participants' intrinsic motivations such as interest in science and altruistic desire to help scientists. It broadly operates on *volunteer* rather than paid participation [38]. Monetary rewards in citizen science have been demonstrated to be *demotivating*, encouraging negative behaviours and raising tensions around adequate rewards for the level of effort offered by volunteers [44]. The use of gamification is not widely spread, though some citizen science projects apply it widely [39]. Participatory sensing models and platforms do not share constraints - the crowd is sometimes financially motivated and previous studies have shown that such incentives could prove critical to maintain engagement in the long term [16].

Participatory sensing has been applied to a wide variety of contexts as a means of gathering data on a larger scale than would be feasible through other methods. In the context of environmental monitoring, sensor devices and participants have proven effective at identifying the presence of pollutants, monitoring the activities of potentially damaging corporations and recording species observations for conservation purposes [12]. Participatory sensing is particularly suited for contexts in which task assigners wish to understand or simulate the experiences of individuals within a given location or environment, or to garner feedback from them [18]. As well as being cost-effective, the more human-centred approach leads to a social contract that has been associated with more reliable and timely submission of data [18, 27].

Participatory sensing functions similarly to the spatial crowdsourcing workflow demonstrated in Figure 3 [51]. A task is assigned by the assigner and published to the crowd, who then gather the necessary data, which is sent to the task assigner. During the quality check process, feedback can be provided by the assigner in the form of additional incentives and rewards dependent on the quality of submissions [51]. However, there may be no formal crowd, nor a pool of workers on whom to draw. If social media is involved as a source of data, then instead of designing, launching and managing a crowdsourcing task, the assigner must instead engage in mining to source and aggregate data [50]. The intended usage and nature of the original data must also be considered. Where data are identified through social networks or other web 2.0 channels, the task assigner will be unable to influence - or perhaps even identify - the sensor devices used to gather data. This may influence their ability to use the data with confidence.

In relation to the Internet of Things, participatory sensing models have been employed to enable smart-city processes, distributing the data gathering and technological processes over crowds of residents and visitors to lower the load placed on any one individual and device [23]. In theory, in a smart cities context, any device could serve as a source of data and the data gathering process for individuals need not require active engagement. Instead, participants fully control the sensors and the data which they publish to the wider network, engendering voluntary participation in a similar sense to that suggested within the context of citizen science [18]. One particular advantage of participatory sensing is that it can gather data from more 'opportunistic' sources [19], expanding on the scope of existing smart information systems deployed for transport, utilities and other areas of smart cities [23].

However, participatory sensing is not without its issues. As it relies on human involvement and voluntary participation, there are no up-front guarantees on data coverage and accuracy [23]. Even when participation is implicit or opportunistic, careful consideration must be given to the specifications of the sensor devices - for example, the battery life of portable sensors such as mobile and wearable devices - to ensure that volunteers are not discouraged or prevented from gathering data [32]. Estimating how long the activity would take to reach critical mass is challenging - incentives are critical to encourage timely delivery and maximise coverage [25]. Furthermore, recent research has focused on the ethical and privacy implications of a participatory-fuelled smart cities approach, balancing the need for live feedback and open sharing of data for the common good with privacy and security concerns [46]. More broadly, modelling trust and reputation in participatory sensing data continues to be a key direction of the research landscape that must be addressed if participatory sensing data is to be applied not only in smart cities contexts, but in scientific research and beyond [33].

6 Conclusion

Most advanced IoT solutions today are more than just technology. They leverage human and social capital in interesting, effective and ethical ways to enhance, extend and oversee technical systems. Participatory sensing complements digital sensor networks. Citizen science and paid microtask platforms help create ground truth to train machine learning models. Human-in-the-loop architectures help design and operate complex systems that bring together people and IoT technology, for instance in smart city control rooms or interactive machine learning.

To leverage human and collective knowledge and creativity, an organisation should consider fundamental questions around: what will people be asked to do, what is the most appropriate available audience who could be engaged to undertake the task, why would they be interested to participate and how would the results be quality-assured and aggregated for further use. Answers to these questions are critical to ensure organisations apply crowdsourcing in all its forms and purposes effectively. This requires expertise in a range of fields, technical as well as non-technical, and raises particular challenges around the definition of tools and experiences to maximise the value of manual efforts, the choice of channels to recruit and communicate with participants, the incentive models that drive participant behaviour, and the ways outcomes are used in existing contexts. In addition, crowdsourcing raises important ethical questions around ownership and fair rewards, in addition to privacy and data protection when citizens' data is collected and shared with others.

Within the crowdsourcing paradigm, we described two areas of study that are particularly useful for IoT scenarios. First there is spatial crowdsourcing, as a special case where contributors travel to specific locations to perform a task. Compared to non-spatial crowdsourcing, the main challenge is effective task allocation, as the effort required of contributors to move around needs to be accounted for in the incentive model, and the pool of available workers near to a particular area may be too small. Second, participatory sensing, which are widely used in smart cities projects, and can tap into pre-existing citizen platforms to recruit participants and implement more opportunistic forms of data collection. In both cases, understanding the motivation and incentives of the likely participants is key - a wide range of models exist that make their own assumptions about what would drive 'the crowd' to engage, but these assumptions need to be complemented by empirical studies and a readiness to consider a mix of crowdsourcing forms to maximize outcomes where needed. For instance, approaches such as [1] and [29] show how different crowdsourcing platforms, incentives and engagement models could be streamlined into a coherent workflow to deliver more complex data collection or analysis objectives and reduce costs.

With sensible tools and guidance, careful coordination, and fair and relevant incentive models, the power of human and collective intelligence can be seamlessly integrated into automated processes, to get the best of both worlds.

References

[1] Maribel Acosta, Amrapali Zaveri, Elena Simperl, Dimitris Kontokostas, Fabian Flöck, and Jens Lehmann. Detecting linked data quality issues via crowdsourcing: A dbpedia study. *Semantic Web*, 9(3):303–335, 2018.

[2] Saleema Amershi, Maya Cakmak, William Bradley Knox, and Todd Kulesza. Power to the People: The Role of Humans in Interactive Machine Learning. *AI Magazine*, 35(4):105–120, December 2014.

[3] Daren C. Brabham. *Crowdsourcing*. The MIT Press essential knowledge series. The MIT Press, Cambridge, Massachusetts ; London, England, 2013.

[4] Adrian Bridgwater. Machine Learning Needs A Human-In-The-Loop. https://www.forbes.com/sites/adrianbridgwater/2016/03/07/machine-learning-needs-a-human-in-the-loop/.

[5] Zhao Chen, Rui Fu, Ziyuan Zhao, Zheng Liu, Leihao Xia, Lei Chen, Peng Cheng, Caleb Chen Cao, Yongxin Tong, and Chen Jason Zhang. gMission: A General Spatial Crowdsourcing Platform. *Proc. VLDB Endow.*, 7(13):1629–1632, August 2014.

[6] D. N. Crowley, E. Curry, and J. G. Breslin. Closing the loop — From citizen sensing to citizen actuation. In *2013 7th IEEE International Conference on Digital Ecosystems and Technologies (DEST)*, pages 108–113, July 2013.

[7] Florian Daniel, Pavel Kucherbaev, Cinzia Cappiello, Boualem Benatallah, and Mohammad Allahbakhsh. Quality Control in Crowdsourcing: A Survey of Quality Attributes, Assessment Techniques, and Assurance Actions. *ACM Comput. Surv.*, 51(1):7:1–7:40, January 2018.

[8] Gianluca Demartini. Hybrid human–machine information systems: Challenges and opportunities. *Computer Networks*, 90:5–13, 2015.

[9] Djellel Eddine Difallah, Michele Catasta, Gianluca Demartini, Panagiotis G Ipeirotis, and Philippe Cudré-Mauroux. The dynamics of micro-task crowdsourcing: The case of amazon mturk. In *Proceedings of the 24th international conference on world wide web*, pages 238–247. International World Wide Web Conferences Steering Committee, 2015.

[10] L. Feng, C. Wiltsche, L. Humphrey, and U. Topcu. Synthesis of Human-in-the-Loop Control Protocols for Autonomous Systems. *IEEE Transactions on Automation Science and Engineering*, 13(2):450–462, April 2016.

[11] Michael J Franklin, Donald Kossmann, Tim Kraska, Sukriti Ramesh, and Reynold Xin. Crowddb: answering queries with crowdsourcing. In *Proceedings of the 2011 ACM SIGMOD International Conference on Management of data*, pages 61–72. ACM, 2011.

[12] Jennifer Gabrys. Citizen sensing, air pollution and fracking: From 'caring about your air' to speculative practices of evidencing harm. *The Sociological Review*, 65(2_suppl):172–192, July 2017.

[13] Ujwal Gadiraju, Ricardo Kawase, and Stefan Dietze. A Taxonomy of Microtasks on the Web. In *Proceedings of the 25th ACM Conference on Hypertext and Social Media*, HT '14, pages 218–223, New York, NY, USA, 2014. ACM.

[14] Ujwal Gadiraju, Jie Yang, and Alessandro Bozzon. Clarity is a worthwhile quality: On the role of task clarity in microtask crowdsourcing. In *Proceedings of the 28th ACM Conference on Hypertext and Social Media*, pages 5–14. ACM, 2017.

[15] Diego S. Gallo, Carlos Cardonha, Priscilla Avegliano, and Tereza Cristina Carvalho. Taxonomy of Citizen Sensing for Intelligent Urban Infrastructures. *IEEE Sensors Journal*, 14(12):4154–4164, December 2014.

[16] L. Gao, F. Hou, and J. Huang. Providing long-term participation incentive in participatory sensing. In *2015 IEEE Conference on Computer Communications (INFOCOM)*, pages 2803–2811, April 2015.

[17] Michael A. Goodrich and Alan C. Schultz. Human–Robot Interaction: A Survey. *Foundations and Trends® in Human–Computer Interaction*, 1(3):203–275, January 2008.

[18] Jayavardhana Gubbi, Rajkumar Buyya, Slaven Marusic, and Marimuthu Palaniswami. Internet of Things (IoT): A vision, architectural elements, and future directions. *Future Generation Computer Systems*, 29(7):1645–1660, September 2013.

[19] Bin Guo, Daqing Zhang, Zhu Wang, Zhiwen Yu, and Xingshe Zhou. Opportunistic IoT: Exploring the harmonious interaction between human and the internet of things. *Journal of Network and Computer Applications*, 36(6):1531–1539, November 2013.

[20] Daniel Haas, Jason Ansel, Lydia Gu, and Adam Marcus. Argonaut: macrotask crowdsourcing for complex data processing. *Proceedings of the VLDB Endowment*, 8(12):1642–1653, 2015.

[21] Christoph Hienerth and Frederik Riar. The wisdom of the crowd vs. expert evaluation: A conceptualization of evaluation validity. In *35th DRUID Celebration Conference, Barcelona, Spain, June*, pages 17–19, 2013.

[22] CCS Insight. Global Wearables Forecast. Technical report, CCS Insight, 2018.

[23] J. Jin, J. Gubbi, S. Marusic, and M. Palaniswami. An Information Framework for Creating a Smart City Through Internet of Things. *IEEE Internet of Things Journal*, 1(2):112–121, April 2014.

[24] Thivya Kandappu, Archan Misra, Shih-Fen Cheng, Randy Tandriansyah, and Hoong Chuin Lau. Obfuscation At-Source: Privacy in Context-Aware Mobile Crowd-Sourcing. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 2(1):16:1–16:24, March 2018.

[25] Ngo Khoi, Sven Casteleyn, M. Moradi, and Edzer Pebesma. Do Monetary Incentives Influence Users' Behavior in Participatory Sensing? *Sensors*, 18(5):1426, May 2018.

[26] Sabrina Klos nee Muller, Cem Tekin, Mihaela van der Schaar, and Anja Klein. Context-Aware Hierarchical Online Learning for Performance Maximization in Mobile Crowdsourcing. *IEEE/ACM Trans. Netw.*, 26(3):1334–1347, June 2018.

[27] Stacey Kuznetsov and Eric Paulos. Participatory Sensing in Public Spaces: Activating Urban Surfaces with Sensor Probes. In *Proceedings of the 8th ACM Conference on Designing Interactive Systems*, DIS '10, pages 21–30, New York, NY, USA, 2010. ACM.

[28] Guoliang Li, Yudian Zheng, Ju Fan, Jiannan Wang, and Reynold Cheng. Crowdsourced data management: Overview and challenges. In *Proceedings of the 2017 ACM International Conference on Management of Data*, pages 1711–1716. ACM, 2017.

[29] Eddy Maddalena, Luis-Daniel Ibáñez, and Elena Simperl. On the mapping of points of interest through streetview imagery and paid crowdsourcing. submitted.

[30] Thomas Malone, Robert Laubacher, and Chrysanthos Dellarocas. The collective intelligence genome. *IEEE Engineering Management Review*, 38(3):38–52, 2010.

[31] S. P. Mohanty, U. Choppali, and E. Kougianos. Everything you wanted to know about smart cities: The Internet of things is the backbone. *IEEE Consumer Electronics Magazine*, 5(3):60–70, July 2016.

[32] Federico Montori, Prem Prakash Jayaraman, Ali Yavari, Alireza Hassani, and Dimitrios Georgakopoulos. The Curse of Sensing: Survey of techniques and challenges to cope with sparse and dense data in mobile crowd sensing for Internet of Things. *Pervasive and Mobile Computing*, 49:111–125, September 2018.

[33] Hayam Mousa, Sonia Ben Mokhtar, Omar Hasan, Osama Younes, Mohiy Hadhoud, and Lionel Brunie. Trust management and reputation systems in mobile participatory sensing applications: A survey. *Computer Networks*, 90:49–73, October 2015.

[34] Mordechai (Muki) Haklay, Suvodeep Mazumdar, and Jessica Wardlaw. Citizen Science for Observing and Understanding the Earth. In Pierre-Philippe Mathieu and Christoph Aubrecht, editors, *Earth Observation Open Science and Innovation*, ISSI Scientific Report Series, pages 69–88. Springer International Publishing, Cham, 2018.

[35] Taewoo Nam and Theresa A Pardo. Conceptualizing smart city with dimensions of technology, people, and institutions. In *Proceedings of the 12th annual international digital government research conference: digital government innovation in challenging times*, pages 282–291. ACM, 2011.

[36] Newzoo. Global Mobile Market Report. Technical report, Newzoo, 2018.

[37] D. S. Nunes, P. Zhang, and J. Sá Silva. A Survey on Human-in-the-Loop Applications Towards an Internet of All. *IEEE Communications Surveys Tutorials*, 17(2):944–965, 2015.

[38] Neal Reeves, Ramine Tinati, Sergej Zerr, Max G. Van Kleek, and Elena Simperl. From Crowd to Community: A Survey of Online Community Features in Citizen Science Projects. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*, pages 2137–2152. ACM, February 2017.

[39] Neal Reeves, Peter West, and Elena Simperl. " a game without competition is hardly a game": The impact of competitions on player activity in a human computation game. AAAI, 2018.

[40] G. Schirner, D. Erdogmus, K. Chowdhury, and T. Padir. The Future of Human-in-the-Loop Cyber-Physical Systems. *Computer*, 46(1):36–45, January 2013.

[41] Katie Seaborn and Deborah I. Fels. Gamification in theory and action: A survey. *International Journal of Human-Computer Studies*, 74:14–31, February 2015.

[42] Linda See, Peter Mooney, Giles Foody, Lucy Bastin, Alexis Comber, Jacinto Estima, Steffen Fritz, Norman Kerle, Bin Jiang, Mari Laakso, Hai-Ying Liu, Grega Milčinski, Matej Nikšič, Marco Painho, Andrea Pődör, Ana-Maria Olteanu-Raimond, and Martin Rutzinger. Crowdsourcing, Citizen Science or Volunteered Geographic Information? The Current State of Crowdsourced Geographic Information. *ISPRS International Journal of Geo-Information*, 5(5):55, April 2016.

[43] Elena Simperl. How to use crowdsourcing effectively: Guidelines and examples. *Liber Quarterly: The Journal of European Research Libraries*, 25(1), 2015.

[44] Elena Simperl, Neal Reeves, Chris Phethean, Todd Lynes, and Ramine Tinati. Is Virtual Citizen Science A Game? *Trans. Soc. Comput.*, 1(2):6:1–6:39, June 2018.

[45] Paul Smart, Elena Simperl, and Nigel Shadbolt. A Taxonomic Framework for Social Machines. In Daniele Miorandi, Vincenzo Maltese, Michael Rovatsos, Anton Nijholt, and James Stewart, editors, *Social Collective Intelligence: Combining the Powers of Humans and Machines to Build a Smarter Society*, Computational Social Sciences, pages 51–85. Springer International Publishing, Cham, 2014.

[46] M. Sookhak, H. Tang, Y. He, and F. R. Yu. Security and Privacy of Smart Cities: A Survey, Research Issues and Challenges. *IEEE Communications Surveys Tutorials*, pages 1–1, 2018.

[47] Hien To, Gabriel Ghinita, Liyue Fan, and Cyrus Shahabi. Differentially Private Location Protection for Worker Datasets in Spatial Crowdsourcing. *IEEE Transactions on Mobile Computing*, 16(4):934–949, April 2017.

[48] Hien To, Cyrus Shahabi, and Leyla Kazemi. A Server-Assigned Spatial Crowdsourcing Framework. *ACM Trans. Spatial Algorithms Syst.*, 1(1):2:1–2:28, July 2015.

[49] Yongxin Tong, Libin Wang, Zimu Zhou, Bolin Ding, Lei Chen, Jieping Ye, and Ke Xu. Flexible Online Task Assignment in Real-time Spatial Data. *Proc. VLDB Endow.*, 10(11):1334–1345, August 2017.

[50] Rita Tse, Lu Fan Zhang, Philip Lei, and Giovanni Pau. Social Network Based Crowd Sensing for Intelligent Transportation and Climate Applications. *Mobile Networks and Applications*, 23(1):177–183, February 2018.

[51] X. Oscar Wang, W. Cheng, P. Mohapatra, and T. Abdelzaher. ARTSense: Anonymous reputation and trust in participatory sensing. In *2013 Proceedings IEEE INFOCOM*, pages 2517–2525, April 2013.

[52] A. Wiggins and K. Crowston. From Conservation to Crowdsourcing: A Typology of Citizen Science. In *2011 44th Hawaii International Conference on System Sciences*, pages 1–10, January 2011.

[53] Jie Yang, Judith Redi, Gianluca Demartini, and Alessandro Bozzon. Modeling task complexity in crowdsourcing. In *Fourth AAAI Conference on Human Computation and Crowdsourcing*, 2016.

[54] Mattia Zeni, Ilya Zaihrayeu, and Fausto Giunchiglia. Multi-device Activity Logging. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication*, UbiComp '14 Adjunct, pages 299–302, New York, NY, USA, 2014. ACM.

[55] Y. Zhao and Q. Han. Spatial crowdsourcing: current state and future directions. *IEEE Communications Magazine*, 54(7):102–107, July 2016.