

Keynote: Privacy and Security in Real-World Data Annotation

Emma L. Tonkin

Email: e.l.tonkin@bristol.ac.uk

Digital Health, School of Electrical and Computer Engineering, The University of Bristol, Bristol, UK

KEYNOTE ABSTRACT

An accurate ground truth is key to both our ability to engineer systems able to represent and respond to human context and, perhaps even more importantly, our ability to develop interfaces capable of effectively representing datasets to the humans who use and work with the system. Data annotation, therefore, is of tremendous value in building and exploiting pervasive systems. However, it is also among the most expensive activities in many ubiquitous or pervasive system engineering processes, as well as presenting significant practical challenges to the implementer and the participant alike: methods of annotation are frequently either based in active user interaction – potentially disruptive to the user – or based in a method of surveillance such as video or audio data collection, enabling post-hoc manual annotation of the material.

In this talk, I briefly review legal and practical challenges to annotation through the lens of practical experience in a real-world digital health engineering context. I discuss strategies through which the task of annotation may be approached. The costs involved include both the financial cost of manual annotation and the toll that annotation strategies may take on participants in terms of time or exposure to surveillance technologies. The decisions taken in choosing annotation platforms, methods and proposed outcomes may have a significant impact on the quality of the annotations and on their applicability for particular engineering tasks, and, importantly, on the quality of the datasets resulting from real-world data collection processes. As participants volunteer their time and, often, access to their private homes, to support engineering research and development, we may consider what our responsibilities are from an ethical engineering perspective, collecting sensitive personal information in a manner that respects the requirements of the dataset, the goals of the engineering process and the participants' own goals and interests.

Finally, I identify challenges: do we always endeavour to build security, privacy assessment and seeking ongoing participant feedback and consent into our research? Do we actively involve engineers into the design and pilot stages of our annotation work, evaluating the quality of annotations against key competency questions prior to commencing the data collection process? In view of the many recent controversies surrounding collection of ground truth information, what more can the sector do to ensure that a broad variety of perspectives are represented in the engineering process and are clearly communicated to participants, system users and users of the datasets we build?

SPEAKER'S BIOGRAPHY

Dr Emma Tonkin is a researcher in Digital Health at the University of Bristol, UK, working on the SPHERE project. Prior to joining the SPHERE project, her research at King's College London focused on the mapping and characterisation of semantic shift and on the effective visualisation of timeseries data drawn from historical corpora. Her background is in library and information science, personal and professional information management practices. She holds a PhD in computer science from the University of Bristol, an MSc in Human-Computer Communication and undergraduate degrees in modern and ancient languages and physics. Her interests include user-focused annotation, analysis and visualisation of spatiotemporal data and privacy-first secure deployment of sensor systems. She is a member of the IEEE.