Managing Emergent Ethical Concerns for Software Engineering in Society

Awais Rashid, Karenza Moore, Corinne May-Chahal
Security Lancaster Research Centre
Lancaster University, Lancaster LA1 4WA, UK
Email: a.rashid, karenza.moore, c.may-chahal @lancaster.ac.uk

Ruzanna Chitchyan
Department of Computer Science
University of Leicester, UK
Email: rc256@leicester.ac.uk

Abstract—This paper presents an initial framework for managing emergent ethical concerns during software engineering in society projects. We argue that such emergent considerations can neither be framed as absolute rules about how to act in relation to fixed and measurable conditions. Nor can they be addressed by simply framing them as non-functional requirements to be *satisficed*. Instead, a continuous process is needed that accepts the 'messiness' of social life and social research, seeks to understand complexity (rather than seek clarity), demands collective (not just individual) responsibility and focuses on dialogue over solutions. The framework has been derived based on retrospective analysis of ethical considerations in four software engineering in society projects in three different domains.

I. Introduction

Contemporary software systems exist in a complex and open hyper-connected eco-system. Consequently, software engineers are modelling and developing systems that are increasingly more and more open, handle large amounts of personal or other sensitive data and are intricately linked with the daily lives of individuals and communities. These software or software-based systems also play a key role in tackling major challenges pertaining to crime and public safety, sustainable living, energy demand, empowerment of otherwise marginalised groups and bridging the digital divide. As such, software engineers come into contact with and must address ethical concerns that arise from such *social embedding* of software. Traditionally, such concerns have been tackled by treating them as non-functional requirements and by evaluating and establishing trade-offs that weigh the potential benefits against the potential harms [3]. However, such approaches are representative of modernist, "rule-based understandings" of ethics, that is, pre-conceived notions of how to act in relation to fixed and measurable conditions. As a result, they do not cater for the 'messiness' of social life and social research and its continuous impact on design choices.

The Software Engineering Code of Ethics [11] has been proposed as a means to ethical decision-making [4]. However, there is insufficient understanding on how to operationalise such codes in complex societal settings especially those relating to emergent ethical considerations where the discipline has no 'agreed' standpoint. This opens up space for a critical, situated and reflexive mode of software engineering ethics which fits with the fluidity of contemporary life, not least that associated with the digital world [12].

In this paper we argue that, with the changing nature of software, software engineering ethics are necessarily situated and relational with respect to the particular contexts in which the software system would be deployed. This requires treating ethics as a fluid process of relations between researchers, the researched and other stakeholders such as gatekeepers and funders. This way of thinking about research ethics, although not unique to software engineering research, has been influential in the responses of researchers to the challenges of online research practice – with a rejection of 'one size fits all' approaches to research ethics in favour of ethical decision-making that is tailored to diverse digital spaces [2] and different groups of users [8]. Our proposed framework thus aims to break down the siloed nature of research ethics and software engineering ethics. We contend that there are important lessons for research ethics to be informed by situated and contextualised software engineering ethics (especially relevant in the modern digital world [13]). This is highly pertinent given the growth in non-specialists working on interdisciplinary projects involving software engineering.

Next, we provide a brief overview of the projects from which we derive our proposed framework. This is followed by a description of our framework and a discussion of the treatment of emergent ethical considerations in these projects.

II. Projects Studied

A. Policing tools for online child protection

We reflect on experiences in the development of two investigative toolkits. Project Isis [10] developed a new digital persona analysis approach to enable investigators to analyse the deception and masquerading tactics used by child sex offenders in online social networks. The toolkit operationalising the approach required use of modular software engineering approaches to enable easy adaptation and updates in response to new or changing offender tactics. Project iCOP [9], on the other hand, developed a novel approach that combines sophisticated filename and media analysis techniques to automatically flag new/previously unseen child abuse media on peer-to-peer (P2P) networks. This enables investigators to distinguish originators of such material in the large volume of (legitimate or illegitimate) activity on P2P networks. Service-oriented software engineering was used to integrate the various

analysis techniques and support integration with workflows and investigative software used by law enforcement.

B. Social media tools for community empowerment

UDesignIt [5] is an innovative social media platform offering a suite of tools that facilitate civic participation in the design of systems. Critically these tools target engagement of otherwise marginalised social groups such as young people, the disabled, the elderly or those digitally excluded due to socio-economic or criminogenic factors. Various interaction modes, e.g., playful interactive screens in public spaces, smartphone applications and second screens provide a means for capturing in-situ perceptions of citizens about particular issues. The system uses a combination of natural language processing and feature modelling to identify key themes being discussed and groups these themes according to their similarity to form a feature model-like structure. An "image-cloud" overlay (an extension of word clouds to image-based representations) provides a highly visual representation of the themes to make the resulting feature model accessible and comprehensible. We reflect on experience of deploying UDesignIt in two settings involving young people: one involving pupils from an 'outstanding' secondary school in a semi-rural town and the other engaging young people in an inner-city London borough through a UK charity that works with young ex-offenders.

C. Future utility infrastructures

The final project in our corpus, All-in-One, is a multi-disciplinary study of future utility infrastructures for sustainable communities in a 100 years timeframe [6]. This involved exploring whether a single utility product can supply all the services that the end-users need, for instance, can the end-user have devices that 'transform' electricity to water via local sewage and rainwater recycling? The project explored the scientific and technological gaps with regards to such an "all-in-one" utility vision. This included requirements analysis, scientific and technological feasibility studies and economic and risk analysis. Two alternatives of the all-in-one vision were studied: (i) given a set of required utility services and input(s), a possible technological chain for utility provision was suggested, and where chain links were missing, new technological development opportunities were identified; and (ii) the provision of all utilities through a single physical infrastructure was considered, such as delivery of electricity and water through a single water pipeline.

III. PROPOSED FRAMEWORK

Based on a retrospective analysis of our experiences in managing emergent ethical concerns in the above projects, we propose an iterative framework that treats ethics as a continuous process. The framework, shown in Fig. 1, is inspired by Boehm's Spiral Model of software development [1] and accepts the 'messiness' of social life and social research. At the heart of the framework is a participatory approach involving various stakeholders and users in a continuous, iterative, process that seeks to understand the complexity

of the emergent ethical issues, demands collective (not just individual) responsibility and focuses on continuous dialogue and re-evaluation of design choices.
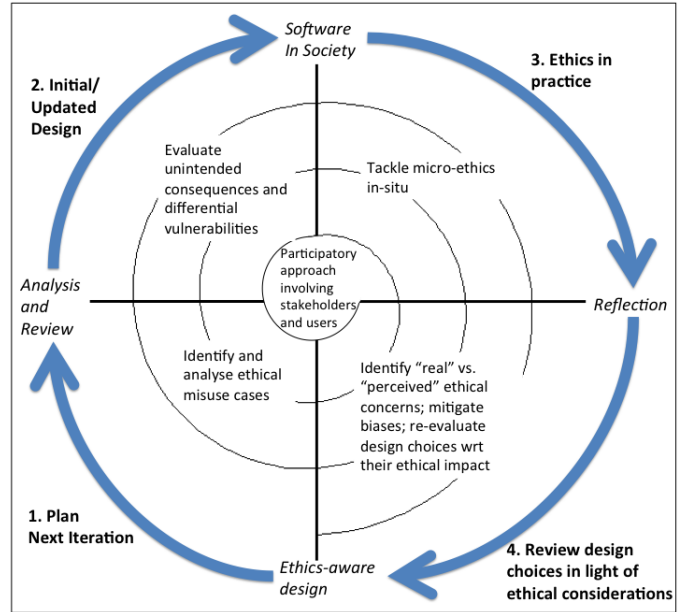

Fig. 1. Proposed framework for handling emergent ethical considerations

The framework highlights which step may be most suitable to tackling particular categories of emergent concerns. We identify four such categories:

1) Ethical misuse cases.
2) Unintended consequences.
3) Micro-ethics of emergent content, contact and conduct risks.
4) Differential vulnerability across user groups.

However, these categories are not orthogonal and an ethical issue may belong to more than one category. Therefore, the depiction in Fig. 1 represents "good practice guidelines" identified from our experience rather than a strict set of universally applicable rules. We next discuss how ethical concerns in each category manifested in the projects in Section II and reflect on our experience of managing those concerns.

A. Ethical misuse cases

This is perhaps the most basic and obvious category of ethical issues. Throughout our corpus of projects, we encountered recurring concerns with respect to rogue users who may have legitimate access to the system and may misuse it to extract personal information of victims or users. Such issues are well-understood due to legislation such as the Data Protection Act and hence do not necessarily represent emergent ethical concerns. However, emergent concerns arose pertaining to particular contexts inhabited by the various systems. For instance, in the case of policing tools, such concerns pertained to protecting anonymity of offenders before sentencing (given the sensitive nature of crimes against children, the consequences

of revealing the identity of an ultimately innocent suspect can be seriously damaging), manipulation of the analysis models or results to gain a favourable outcome with regards to a suspect and the use of the toolkits to engage in illegal activity (for instance, by an offender who may be a police officer).

As another example, when the technological chain from the All-in-One project was presented to a group of stakeholders, large utility providers hailed the research as a tool to be used for planning pre-emptive takeovers of smaller, potentially competitive technology providers. Thus, instead of helping to identify and promote promising sustainable technologies, the result could lead to systematic shutdown of such technology providers in their infancy. Another example of an emergent misuse across all four projects was the potential for misusing the data captured by the various systems for political gain – by utilising it as a basis for positive or negative outlooks with regards to crime, community priorities (such as wind energy providers manipulating a community to build local wind turbines) or sustainability.

B. Unintended consequences

The unintended consequences relate to higher-order ethical impacts of particular design choices. Typically, software requirements only consider first-order impacts arising due to particular features or non-functional properties. Second and third order impacts do not become apparent until software comes into contact with real users or real operational settings. Such unintended consequences, if elicited effectively, can point to key ethical considerations. For instance, in the case of both policing projects, an unintended consequence that emerged following initial user trials was the issue of *too much intelligence*. While the tools could generate much more data and flag lots of potential cases compared to the prevalent manual efforts, this would lead to overloading and stretching already stressed investigative resources. While overloading could be mitigated by prioritising cases, the principle of *duty of care* meant that law enforcement officers would still need to investigate every case that was brought to their attention. This, in turn, required a reconsideration of the toolkits' role (e.g., intelligence provision vs. investigation and forensics) as well as their positioning in investigative and legislative frameworks governing online policing activities. Another example involved the investigators becoming *over-reliant* on the toolkits. Though the toolkits had a high degree of accuracy, given the underlying artificial intelligence and machine learning methods, they could not be 100% accurate. However, investigators could, over time, become over-reliant on them especially as they led to more effective intelligence and apprehension of criminals. This led to reconsideration of the user-interface and reporting features to ensure that suitable confidence levels were attached to any results generated through the toolkits and workflow steps "designed-in" to enforce cross-checking and examination of any decisions resulting from the analyses in the toolkits.

A similar consideration of "designing-in" checks was required when generating infrastructure design outlooks in the All-in-One project. Here it soon became apparent that pro-

viding all utilities via a single physical infrastructure, while financially attractive, would also lead to increased vulnerability of this infrastructure to threats/attacks. Another example of unintended consequences involved the provision of smartphones to users engaging with UDesignIt. While one of the trials focused on enabling young people in a deprived area of London to capture spaces where they felt safe or unsafe, giving young adults expensive desirable consumer goods would exacerbate the vulnerability of the users to theft from their person (this was experienced by at least one participant).

C. Micro-ethics of contact, content and conduct risks

Micro-ethics represent the minutiae of ethical decision-making which require software engineers to negotiate the fluidity of the context in which the software is to be situated. In our experience, these mainly pertained to contact, content and conduct risks [7] (albeit this may be due to three out of four of our projects relating to young people/young victims). An example of such micro-ethics included the need to refine the performance and accuracy of the Isis and iCOP toolkits. Given the nature of the material, there was a significant risk of harm to researchers being exposed to such material not to mention the illegality and immorality of observing such material. Protocols needed to be defined to ensure that any initial training of the tools was undertaken on non-sensitive data or features extracted from sensitive data in a way that they could not be used to reconstruct the original data. Any user testing was undertaken by trained law enforcement officers in closed room settings with accuracy and performance data captured through built-in "debugger-like" components. This was followed by debriefing with the law enforcement officers and the cycle repeated. Another example involved the use of smartphones by young people in the UDesignIt trials. Here the concern pertained to exchange of "sexualised" images amongst participants (i.e. sexting), a trend deemed prevalent at the time of the trials. On the one hand, we needed to abstain from over-regulating our participants as prior research has shown that discourses of vulnerability and risk compete with discourses of children and young people's participation and involvement [7]. Yet the 'risk' demanded a response relevant to the local research moment and to our research participants thought to be vulnerable as a result of their age, gender and social standing. We tackled this through the development of an Acceptable Use Policy (AUP). We explained the AUP to our participants, but with a keen awareness of the 'delicacy' of explaining without appearing to labour under the assumption that participants would undertake the undesirable behaviours we were setting out. This was essential to avoid both colluding with the negative stereotypes about the young people we were engaged with and appearing to 'lecture' the participants hence discouraging them from participating[1].

Another consideration relates to marginalising certain communities when providing a software solution. We faced this is-

---

[1] These are examples of higher order impacts resulting from decision-making with regards to emergent ethical concerns hence highlighting the non-orthogonality of these categories.

sue in the All-in-One project, where a set of key infrastructure designs centred on use of high capacity underground tunnels, which are often already present and excellently positioned in dense urban spaces. Since the future is increasingly urban, this seems to be an effective solution. This, however, is not cost-effective for rural areas. Should the thrust of the future utilities provision move in this direction, the rural areas are likely to be left with the use of outdated services. Such situations are not unusual in software engineering projects, where the interests of the weakly influential stakeholders are marginalised. To ensure that the voice of rural inhabitants was preserved, we provided a range of complementary outlooks which would also accommodate rural areas.

D. Differential vulnerability across user groups

During the various projects we were differentially reticent about various user groups' resourcefulness in terms of managing particular risks. Our ethical concerns were shaped by the groups we were engaging with at a particular time. For instance, in the policing applications we expected law enforcement officers to be supported by training and counselling to manage the risks arising from exposure to sensitive media. However, we were much more pro-active with regards to protecting the software engineers from exposure to such media. On the one hand, this highlights the need to assess vulnerability of each user group separately. On the other hand, this reflects pre-conceived notions of risk and harm. These pre-conceptions were only too obvious in the UDesignIt trials. In the school trial, we worked with pupils who were 'trusted', 'responsible' and purposefully selected members of the youth forum. They had received parental consent to take part in the trial and had their participation supervised by the deputy head teacher. In contrast, the purposive sample of the charity trial participants, 'hand-picked' by key workers at the charity did not provide the same reassurances. Our concerns were grounded in key worker fears articulated from the very start of the study with regards to potential content, contact and conduct risks highlighted above. This led to two variants of the Acceptable Use Policy, one for the school trial and one for the charity trial, with the latter having a much more prescriptive policy with regards to abuse of the platform to harass, inconvenience or cause offence to any other person, create or send/receive any offensive, obscene or indecent material or to create, send/receive or install material that infringed on the copyright of another person. We have an example of one trial (the charity trial) where we maximised control over the process which was largely grounded in key worker fears about potential ethical issues that may arise, and another trial (the school trial) where we were less apprehensive about the participant-produced data and censorship/moderation was kept to a minimum. In the event, no reported ethical issues arose in either trial. This reflects the need to regularly question researcher and stakeholder bias with regards to perceived risks and whether the risks are indeed differential across different user groups or founded in preconception about the resourcefulness of those groups in managing those risks.

IV. Discussion and conclusion

Our experience shows that dealing with ethical concerns during software engineering in society goes well beyond managing ethical use of information and communication technologies and more familiar considerations such as data protection. In fact, questioning fixed, embodied and essentialistic notions of ethics opens up a space for ethics-aware design that is critical, situated and reflexive and, hence, sits well with contemporary digital life. Yet as we reflect on our 'ethical journey', we also note that, despite our best efforts, ultimately if unwittingly we were also, at times, involved in reproducing the risk culture that can act to curtail participation of marginalised user groups. We were only able to mitigate this due to a strong awareness and consideration of higher-order ethical impacts and unintended consequences (direct and indirect) of our design choices as well as "designing-in" ethical considerations as they came to the fore. We contend that without full and frank consideration as to their presumptive base, fears and cultural assumptions are too readily translated into ethical concerns of the sort that institutionalist ethics procedures demand software engineers to identify. This in turn may inhibit effective software engineering for positive ends in society, in itself running contrary to the principal aims and the welcome demand to 'do no harm, do good'. What is required instead is a reflexive, fluid ethics engagement that foregrounds consideration of the emergent consequences, intended and unintended, arising from social embedding of software and pays careful attention to in-situ microethics.

References

[1] B. Boehm, *A Spiral Model of Software Development and Enhancement*, ACM SIGSOFT Software Engineering Notes, ACM, 11(4): 14-24, 1986.

[2] R. Eynon, J. Fry, R. Schroeder, *The Ethics of Internet Research*, Handbook of Online Research Methods, SAGE, pp. 23-41, 2008.

[3] D. H. Gleason, *A software development solution*, Proc. Ethicomp: Systems of the Information Society, Poland, 2001.

[4] D. Gotterbarn, K. W. Miller, *The Public is the Priority: Making Decisions Using the Software Engineering Code of Ethics*, IEEE Computer 42(6): 66-73, 2009.

[5] P. Greenwood, A. Rashid, J. Walkerdine, *UDesignIt: Towards social media for community-driven design*, Proc. ICSE, pp. 1321-1324, 2012.

[6] F. Karaca, P Raven, J. Machell, L. Varga, F. Camci, R. Chitchyan, J. Boxall, B. Ulanicki, P. Skworcow, A. Strzelecka, L. Ozawa-Meida, T. Janus, *Single infrastructure utility provision to households: Technological feasibility study*, Futures Journal 46, Elsevier, pp. 35-48, 2013.

[7] S. Livingstone, *Children and the Internet*, Polity Press, 2009.

[8] C. May-Chahal, C. Mason, A. Rashid, J. Walkerdine, P. Rayson, P. Greenwood, *Safeguarding Cyborg Childhoods: Incorporating the On/Offline Behaviour of Children into Everyday Social Work Practices*, British Journal of Social Work, pp. 596-614, 2012.

[9] C. Peersman, C. Schulze, A. Rashid, M. Brennan, C. Fischer, *iCOP: Automatically Identifying New Child Abuse Media in P2P Networks*, IEEE Symposium on Security and Privacy Workshops, pp. 124-131, 2014.

[10] A. Rashid, A. Baron, P. Rayson, C. May-Chahal, P. Greenwood, J. Walkerdine, *Who Am I? Analyzing Digital Personas in Cybercrime Investigations*, IEEE Computer 46(4): 54-61, 2013.

[11] *Software Engineering Code of Ethics and Professional Practice*, ACM/IEEE, http://www.acm.org/about/se-code.

[12] N. Whitehead, *Control and Contingency: Maintaining Ethical Stances in Research*, International Journal of Internet Research Ethics, 3(1): 6-22, 2010.

[13] M. Zimmer, *But the Data is Already Public: On the Ethics of Research in Facebook*, Ethics and Information Technology, 12(4): 313-325, 2010.