Chapter 4
Addressing Ethical Challenges of Creating New Technology for Criminal Investigation: The VALCRI Project

P. Duquenoy, D. Gotterbarn, K. K. Kimppa, N. Patrignani, and B. L. William Wong

Introduction

For the past 3 years, we have been working on a European Union project involving multiple partnerships between software developers and several EU police departments to develop a decision support system based on ground-breaking technology. Three key problems identified and to be addressed by the project, were (DoW 2014, pp. 5–6): (a) Problems of interacting with a large dataset in Intelligence Analysis – processing large amounts of data into actionable intelligence; (b) Problems in Information Analysis facing the Analyst – making sense of large volumes of data, assembling relevant information in meaningful ways; (c) Problems arising from the Lack of Technical Skills – need user-interface that works with, and for analysts. To overcome these challenges the project aims to support intelligence analysis by providing "a system that facilitates human reasoning and analytic discourse, tightly coupled with semi-automated human-mediated semantic knowledge extraction." (http://valcri.org/). To avoid any potential ethical, social or privacy concerns this project was reviewed by an independent ethics board and guided by an internal Legal, Ethics and Privacy group (LEP).

P. Duquenoy (✉) · B. L. William Wong
Middlesex University, London, UK

D. Gotterbarn
Software Engineering Ethics Research Institute, Department of Computing,
East Tennessee State University, Johnson City, Tennessee, USA
e-mail: don@gotterbarn.com

K. K. Kimppa
University of Turku, Turku, Finland

N. Patrignani
Universita Cattolicà del Sacro Cuore, Milan, Italy

31

VALCRI Explained

The goals of the project are to: (a) "Assist police in the analysis of large and complex datasets in support of intelligence and investigative work; (b) Facilitate human reasoning and analytic discourse; (c) Augment rather than replace human decision-making; (d) Protect against human cognitive bias and abuses arising from accidental, inadvertent or deliberate violations of ethical, legal and privacy principles." (Shepherd 2015). The technical objectives of the project included developing: a system that addresses bias mitigation, social and legal factors into a single principled framework that developers can use to guide the system design. (DoW, pp. 9–10). The development team was skilled in solving the technical problems, but there were a host of other issues to address: legal constraints about ".privacy and security, and some significant ethical issues. Addressing the ethical and social constraints during development required an interactive process where the project could regularly be reviewed. To do this the project brought together technical designers, ethics, law and privacy experts, several end-user groups from EU as partners, and an independent ethics board. The main risks identified by the IEB were: Diverse social, professional and legal mores, privacy and data protection risks, risks of poor communication, data and reasoning provenance and the transparency of the entire analysis process for addressing significant ethical concerns. Issues related to cognitive bias were noted early in the project and mitigation measures in the form of design guidelines produced as part of the Human Issues Framework in WP3 (Haider et al. 2015).

Addressing Ethics in VALCRI

In the first stage ethics specialists were invited as members of an Independent Ethics Board (IEB) (composed of experienced ethics advisers) that would review and guide the project process from the ethics perspective, and ethics was specifically included in a work package (WP3) with the topics of privacy and law. During the project the legal, ethics, privacy group (LEP) was extended to include security (SEPL) to improve communications between the human issues side and the technical side, enabling the principles from LEP to be operationalised technically. To address both intentional and unintentional ethical lapses, the VALCRI project assigned the task of proactively identifying and developing ethical safeguards to the SEPL group and IEB. The concerns were with both process (alert developers to potential risk facilitated by their development of the software) and product (is the product ethically viable and does it mitigate known ethical problems). The teams identified modifications to the design and development which could be addressed by the system developers. Addressing these

identified elements required ethical design beyond mere technological solutions. The built-in ethical safeguards were needed to have the system both identify and either prevent or alert individuals to alter their (unintentional) decision strategies.

Security, Ethics, Privacy and Legal Group (SEPL)

SEPL's task of communicating security, ethics, privacy, legal issues required the 'translation' of human factors requirements into language more relevant to the technical teams. The insights about the human factors problems from both IEB and SEPL will contribute to other research projects beyond the area of crime visualization, and also contribute to resolving a wider range of concerns present in other sensitive projects, such as concerns regarding logging the users' actions and documenting the reasoning processes for auditing and evidence (meeting the "Transparency in analysis" requirement mentioned earlier).

Communications Problems in a Multi-Tiered Approach

Systems development communications was a problem between teams and with the IEB. The SEPL group works well enough but unfortunately not all developers have representatives in this group creating difficulties in ensuring that the proposals for securing ethical resolutions have been communicated to implementers. Security, being a well-defined subject, became the primary focus of the group which resulted in mistakenly equating security and ethics. An unintended consequence of this arrangement was that teams initially reduced ethical issues to privacy and data security. The word 'security', when applied to "secure the rights of those stakeholders impacted by the system", was planned to be addressed by simply limiting access to the system, that is, data security. Initially privacy rights or securing rights not to be falsely accused (due to confirmation bias) were not included in the system.

Addressing IEB, SEPL, VALCRI Team Interactions

An early underlying issue was that the complexity and inter-relatedness of multiple IEB issues were not recognised. In order to keep track of the concerns that were raised by the IEB, and to make sure there was follow-up with the relevant technical groups, a spreadsheet listing the concerns was developed by management. This tool which could have been used to help communication between teams was used

primarily as a task assignment device. For each concern on a row, a person or a group would be asked "How do you address this?" drawing responses like "It is not in the scope of our assignment" or "It is handled in the X module?" Since there was no management overview of the document, teams would have a different understanding of a concept or term depending on what function they were working on. What was meant by "track changes", "data reliability", and "transparency"? The way the concerns spreadsheet was used contributed to contradictory responses for issues about provenance. There was also a basic translation problem requiring technical people to interpret abstract issues raised by the IEB. What does "data transparency" actually mean to the different parties? These examples illustrate the significant limitations of document-based approaches to bridge the translation gap compared to face to face discussions. The benefits of the spreadsheet are recording and management of issues in one document (dates, notes added, a column for actions etc.) with the document providing a dynamic and ongoing record. The disadvantages are (i) sharing revisions with the IEB (ii) The nature of the document and the process (to contact relevant people in the consortium to ask if the issue is addressed) forces a 'silo' approach – that is a response to a specific issue and no other issues (iii) the issue can be phrased ambiguously (e.g. the 'transparency' example above) and responding to the concern as expressed in the document results in a query. Without face-to-face, or vocal communication in some way (that allows for queries, elaboration, explanations) the concerns expressed have to be interpreted by the responder who is mostly operating in unfamiliar territory – ethical/privacy/legal issues. If misinterpreted the responses are not likely to answer the concern. Both of these methods have benefits and both have been used, they are excellent methods for exploring the issues and sharing with developers the reasons why a particular item is an ethical concern. However, in the context of the 'face-to-face' option, unless someone has been allocated to record the detail, or some pre-prepared 'check sheet' is completed, records of the meeting rely on notes and remembering conversations.

Risks Identified

Some of the challenges faced and trade-offs required in developing and using a semi-automatic decision support system based on visually-aided thinking present different risks.

Standard risks There are different kinds of software risks; some are ethical risks discussed on the following pages, while others are common to most software projects. Initially, some risks such as changing legal constraints were not anticipated, including problems such as how to comply with the "right-to-an-explanation" feature of the new EU Data Protection (http://ec.europa.eu/justice/data-protection/). This requires "algorithm transparency" and "reasoning transparency". The systems development difficulty is that changes are required to provide both while also

protecting classified policing techniques. Transparency as a concept applied to the project has been useful in alleviating some of the issues, for example, showing the user how the system is working, as well as making transparent user actions in allowing users to record their analysis process and sources of information on a particular case (Islam et al. 2017). In addition, the security protocols for the project include transparency in the capability to record access to the system, changes to policies of access, and readable (auditable) logs, and 'super-logs' accessible only by independent authorities (as described by Schlehahn et al. 2017). The point of this is that some of the ethical concerns have been addressed by the application of different technologies that have been organised into different domain types with the project, such as 'security', provenance, etc.

Ethical Social and Legal Risks due to the nature of the project IEB identified issues include: accidental discrimination, the Mosaic effect, algorithmic opacity, data aggregation with mixed levels of reliability, data and reasoning provenance, and various biases, (IEB 2017). Several of these issues were then addressed by white papers. VALCRI also tracked and evaluated these concerns by constructing scenarios in which the concern might arise, comparing it to current police practices, using standard technical risk approaches of estimating the probability, the significance of impact and considering mitigation strategies. The results of this work were then incorporated into the requirements and design. The thawing and refreezing of requirements leads to its own set of problems, discussed below. Amongst the difficulties identified are two closely related issues whose relation may not immediately be apparent; first, the fact that decision support is automated leads to overconfidence in the "computerized" results (Chen and Koufas 2015) and second, the use of visual thinking techniques encouraging intuitive insightful analysis which impacts the investigators obligation to produce 'communicable knowledge' which can be used by other investigators and can be used in courts. Techniques used to address this overconfidence tendency include recording the competence level of the analysts and each step in the analysis. But this left open issues of information bias – a tendency to ignore disconfirming instances. A system needs to make clear when the evidence for suggested possible relations displayed by it is weak. Showing multiple relationships as if they are all based on the same level of evidence opens the door to information bias which is more likely when predictions are vague and outcome feedback is ambiguous. This leads to an erroneous level of analyst confidence in weak decisions. On the other hand, other research indicates that our optimism about using skill level indicators for the analyst would not resolve the overconfidence problem. Research has shown that when a decision maker's experience level is low that they base their decision on the empirical data (explicit information), whereas those with a high or medium experience level focus on their previous experiences (implicit information) and have a tendency to ignore disconfirming evidence. Using skill levels as criteria may encourage later analysts to suffer from overconfidence bias. (Millitello 2017). Investigators also have an obligation to produce 'communicable knowledge' useable by other investigators and useable in

courts. Resolving this problem helps to reduce overconfidence bias. To address this communicable knowledge concern requires the system to support transparently handling chains of reasoning – recording the reasons for decisions and connections made by the investigator, the articulation of the reasons by the investigator help develop communicable knowledge. Besides resolving the communicable knowledge concern, this approach addresses both the overconfidence in computerized decisions and the not looking at the empirical evidence overconfidence issue. Psychological studies have shown that when making a decision we tend to think we are more ethical than we really are (Tenbrunsel and Messick 2004). This combined with an over-confidence in one's analysis skills may lead to hasty judgment in relating evidence or giving more credibility to unreliable frail evidence. As we have indicated above, this overconfidence is added to by a tendency to give too much confidence to results coming out of a computer-decision support system designed to aid visual thinking. There are two ways to deal with this danger of high confidence in weak decisions. One is to introduce social sanctions on the analyst. A second method is to introduce provenance and the need to explicitly state the reason, i.e. empirical evidence. (Based 2010) The system cannot impose social sanctions so it must provide traceability for all decisions, i.e. decision provenance. The complexity of these systems is evident in that the attempt to resolve one form of bias may introduce another.

Thawing requirements and research projects In this research project several of the requirements were in flux and the various decisions on how to implement them could lead to very different ethical results. In this kind of project it is important not to underestimate the level of detail needed to make an ethical assessment; the level of detail needed to fill the 'gap' between technology development and soft human issues.

Steps Taken over the Course of the Project

The communication lines in the VALCRI ethics communications organised chart were too narrow. IEB would pass broad complex social concerns to management who would pass the concern on in summary form to developers who would try to interpret what was meant and then either try to implement their interpretation or might say 'it was not part of the project' in part because they did not want to spend a lot of time away from their "real" responsibilities. These difficulties were addressed by VALCRI in several ways. They made clear that mitigating the ethical issues was part of the "real work" and they widened the communications channels. IEB worked with SEPL and attended other working group meetings. VALCRI established additional sub-groups to co-ordinate work on linked themes. IEB was given access to all VALCRI documents and invited to prototype presentations and communications with user groups. The initial approach to addressing IEB concerns was one directional from IEB to developers, an approach which led to misunderstanding and

wasted effort. The addition of a feedback mechanism was critical in moving forward. The positive support of management in asserting the importance of addressing these issues and providing vehicles to address them resulted in a more effective development process and better system, and negotiating through disagreements improves understanding and reduces the gap between the different interests (ethics advisors and developers). A full time experienced ethics specialist inside the project with the ability to physically visit the various development sites when ethical issues were found would be extremely useful for a similar project.

Conclusions

In this paper, we have attempted to set out the challenges faced in embedding 'ethics in design' for research and technical development projects. The path to achieving that has highlighted difficulties in communicating requirements from the ethics side to the development teams. Some of the methods for communication and managing the process of addressing the issues identified have been more successful than others. The 'translation' from ethics issues to implementing solutions was problematic, particularly at the beginning of the project, but the steps taken to overcome this in the formation of cross-cutting groups and discussions between the SEPL group and the different technical teams resulted in a significant improvement in the process and understanding.

References

Based, A. (2010). Information handling in security solution decisions. In S. Tarek (Ed.), *Innovations and advance in computer science and engineering*. Dordrecht, Springer.

Chen, C., & Koufas, M. (2015). The impact of decision support system features on user overconfidence and risky behavior. *European Journal of Informatics Systems, 24*(6), 607.

DOW. (2014). Description of Work, VALCRI project. Internal document (contact B.L.W.Wong).

Haider, J., Pohl, M., Hillemann, E.-C., Nussbaumer, A., Attfield, S., Passmore, P., & Wong, B. L. W. (2015). Exploring the challenges of implementing guidelines for the design of visual analytics systems. *Proceedings of the Human Factors and Ergonomics Society 59th Annual Meeting, 59*(1), 259–263, Sage.

IEB. (2017). Independent ethics board second report on VALCRI project. The Midterm Report to European Commission January 2017. Second Mid-Term Project Review (2017) (MTR 2) D4.6 "D4.6UIDesign for Provenance and Uncertainty Final".

Islam, J., Anslow, C., Xu, K., & Wong, B. L. W. (2017). Analytical provenance for criminal intelligence analysis. VALCRI White Paper Series. Available from: http://valcri.org/our-content/uploads/2017/02/VALCRI-WP-2017-009-Provenance3.pdf. Accessed 30 Nov 2017.

Millitello, L. (2017). Applying NDM to change behavior and persuade" 13th international conference on naturalistic decision making 2017, Bath, UK.

Schlehahn, E., Duquenoy, P., Paudyal, P., George, C., Coudert, F., Delvaux, A., & Marquenie, T. (2017). The operationalisation of transparency in VALCRI. VALCRI white Paper Series, 2017. Available from: http://valcri.org/our-content/uploads/2017/02/VALCRI-WP-2017-005-Transparency.pdf. Accessed 30 Nov 2017.

Shepherd, I. (2015). Single Vision Statement of Goals. 2 June 2015 Krakow Consortium meeting.

Tenbrunsel, A., & Messick, D. (2004). Ethical fading: The role of self-deception in unethical behavior. *Social Justice Research June, 2004*, 223–236.