What Do We Teach When We Teach Embedded Ethics?

**Avigail Ferdman,** Department of Humanities and Arts, Technion-Israel Institute of Technology

**Emanuele Ratti,** Institute for Philosophy and Theory of Science,Johannes Kepler University Linz

Amid the increasing interest in the role that ethics should play in computer science education, and AI in particular, ‘embedded ethics’ is emerging as a distinct type of distributed pedagogy. By embedding philosophers directly into computer science courses, it teaches students how to think through the ethical and social implications of their work. It modifies existing courses in computer science, integrating ethics modules that enable students to identify ethical implications of technology while they are developing and implementing algorithms, learning to code and designing interactive systems. To both describe current trends in embedded ethics coursework and to provide guidance for further ethics inclusion in computing, we present an in-depth analysis of 65 embedded ethics course modules. We identify explicit and implicit values inherent in the module syllabus. We then distinguish between three types of normative perspectives: the ethical, the moral and the political. The ethical perspective asks: what are the virtues or ethical skills that computer scientists should cultivate? The moral perspective asks: what do computer scientists owe others? The political perspective asks: what are the political and social institutions and structures within which computer scientists work? Most embedded ethics modules address the moral perspective, focusing on algorithmic biases and discriminatory ML. However, our analysis shows that to a large extent these modules implicitly contain resources that, if made explicit, can help the students navigate their social responsibility within the larger political and social structures (the political perspective), and to cultivate virtues necessary to become not only a skillful AI developer but a good human being (the ethical perspective). Exposing these implicit perspectives is important in two ways: first, it makes explicit the latent resources that exist in embedded ethics curricula, thus providing analytic clarity. Second, some universities are looking to design embedded ethics programs in which computer scientists, rather than philosophers, are the ones who teach embedded ethics. The main reason for this shift is that it is thought that only computer scientists/AI specialists with the relevant professional experience may be able to spot ethical and political challenges in the complex bundle of engineering practices. However, at least a first generation of computer scientists/AI trainers may lack the relevant ethical and political sensitivity to actually spot those challenges. Our work can provide a blueprint to overcome this problem. In particular, we propose to train future trainers on the basis of our tripartite framework for analyzing modules.