Our research is concerned with furthering the understanding of ethics through computational means and incorporating ethical principles into machines. Since 2002, we have been instrumental in establishing *machine ethics* as a bona fide field of study, organizing and chairing the first symposium on the topic (*AAAI Fall 2005 Symposium on Machine Ethics*) and editing the *IEEE Intelligent Systems Special Issue on Machine Ethics* (August, 2006). We have developed a representation for ethical dilemmas based upon W. D. Ross' theory of *prima facie* duties and have used it in conjunction with inductive logic programming to discover a novel ethical principle from cases (see our article in the IEEE Intelligent Systems special issue as well as our featured article in AI Magazine, Winter 2007). Subsequently, we used this principle to drive an expert system (MEDETHEX) that provides guidance for a particular ethical dilemma in the domain of healthcare. MEDETHEX was chosen as an emerging application for the *2005 Conference on Innovative Applications of Artificial Intelligence* (http://uhaweb.hartford.edu/anderson/machineethics/medethex.html).

We have recently used this principle to guide the behavior of a robot charged with reminding a patient to take medication in an ethically sensitive manner ("Robot be Good", *Scientific American*, October 2010). For a short video featured in, among others, The Los Angeles Times, The Boston Globe, and on Discovery.com see http://www.youtube.com/watch?v=ZLdvCDFriTQ.

Success of large scale pervasive computing is likely to be determined largely by the willingness of society to adopt it. Therefore, given the magnitude of resources required to undertake such initiatives, it is imperative that, from their inception, they are sensitive to the concerns of those being asked to embrace them. These concerns are deep and wide ranging but paramount among them is that such technological undertakings proceed only with full foreknowledge of their ethical import and effective strategies for dealing with these issues. Simply put, our vision entails using computational means to 1) help gain this knowledge and 2) provide a way of dealing with it.

Using a generalized version of the machine learning technique developed to discover the principle used in our earlier research, we are currently developing a *General Ethical Dilemma Analyzer*, GENETH (to be submitted to *AAAI-11*), a tool to be used by ethicists that facilitates discovery of ethical features, *prima facie* duties (duties each of which could be overridden on occasion by one of the other duties) and principles from particular cases of ethical dilemmas. Although there may not be a universally accepted *general* theory of ethics at this time, there is wide agreement on what is ethically permissible, and what is not, in *particular* cases and much can be learned from those cases. As the technology in question is typically created to function in specific, limited domains, determining what is ethically acceptable, and what is not, is a less daunting task than trying to devise a general theory of ethical and unethical behavior. GENETH

generalizes from determinations about particular cases, testing those generalizations on further cases, and repeats this process towards the end of developing general principles that agree with the original determinations. These principles determine the correct action when *prima facie* duties give conflicting advice and can be used to inform the engineering of the technology or, in the case of more autonomous machines, guide the behavior of these systems and provide a formalism for verification of this behavior.

GENETH derives its power from the Kantian insight that, to be rational, like cases must be treated in the same fashion-- contradictions in ethics is unacceptable. With two ethically identical cases – i.e. cases with the same ethically relevant feature(s) to the same degree – an action cannot be right in one of the cases, while the comparable action in the other case is considered to be wrong. Formal representation of ethical dilemmas and their solutions make it possible for the system to determine contradictions that need to be resolved. If GENETH encounters two cases that appear to be identical ethically, but it is believed that they should be treated differently, then there must be an ethically relevant difference between them. If the judgments are correct, then there must either be a *qualitative* distinction between them that must be revealed, or else there must be a *quantitative* difference between them. This can be translated into either a difference in the ethically relevant features between the two cases, i.e. a feature which appears in the one but not in the other case, or else a wider range of satisfaction or violation of existing features must be considered which would reveal a difference between the cases, i.e. there is a greater satisfaction or violation of existing features in the one, but not the other, case. GENETH helps automates this complex process, managing the case base, maintaining consistency across cases, tracking the evolution of the principle, suggesting new cases that would further differentiate the principle, etc.

As evidence of the potential of the system, lifting the assumptions entailed by our earlier research GENETH was successful in recreating the principle derived from that research. We envision GENETH engaging in a dialogue with ethicists to determine the ethically relevant features, *prima facie* duties, and principles of the dilemmas that might arise in the wide-scale deployment of pervasive computing. These then can be used to inform the engineering of pervasive computing technology. As evidence that the principles discovered by the system can be used to guide the behavior of more autonomous systems, we offer our recent work successfully embodying our discovered principle in a robot that balances the *prima facie* duties of beneficence, nonmaleficence, and respect for autonomy as it decides when to remind patients when to take their medication and when noncompliance should be reported.

We recognize that there is not agreement, even by ethicists, as to what is acceptable behavior in some circumstances and, as a result, it would be undesirable to allow technology to function in these areas. Thus, we maintain that it is important that the development of pervasive computing not outpace general agreement as to what is considered to be correct ethical behavior. Seen in this light, work in machine ethics, whose goal is discovering and implementing generally accepted ethical principles, is central to the wide-scale deployment of pervasive computing.