Department of Philosophy
Guest Lecture Series

Cultivating an AI Ethics Ecosystem

Oct. 17, 2024 ◆ 4:00 PM ◆ Hale Library Rm 581

It is commonplace by now to observe that the research, design, and deployment of recent AI systems are replete with ethical failings. Less obvious is that such shortcomings are not limited to those cases in which some agent or other (whether individual or group, government or private, for-profit or not) has simply failed to act conscientiously. Many well-intentioned actors, while devoting significant resources to behaving responsibly, still err. In this talk, I'll identify several recurring ethical failure modes and provide a partial diagnosis for their recurrence. Then I'll use that diagnosis to motivate a concentrated effort to cultivate an *AI Ethics Ecosystem*: a cross-sectoral network that orients its members toward a common set of foundational values, researches and trains its members with respect to the operationalization of those values in particular contexts, provides some form of governance and oversight, and more generally organizes and manages the necessary division of labor for adequately managing AI research, development, and deployment in a responsible way. Drawing primarily on the history and performance of health ethics ecosystems, I conclude with a modest attempt to show that such an ecosystem could in fact be developed for AI ethics, that the shortfalls of existing ethics ecosystems aren't as bad as they seem, and that failing to deliberately cultivate an AI ethics ecosystem will likely have quite bad consequences.

Jeff Behrends is a Senior Research Scholar & Associate Senior Lecturer in the Department of Philosophy at Harvard University, where he's also a Faculty Advisor to the Edmond and Lily Safra Center for Ethics and the Embedded EthiCS program.

Kansas State
University