(Computer) Ethics

Ethics

- MW: an area of study that deals with ideas about what is good and bad behavior : a branch of philosophy dealing with what is morally right or wrong

- IEP: The field of ethics (or moral philosophy) involves systematizing, defending, and recommending concepts of right and wrong behavior.

- WKPD: Ethics, sometimes known as philosophical ethics, ethical theory, moral theory, and moral philosophy, is a branch of philosophy that involves systematizing, defending and recommending concepts of right and wrong conduct

Moral

- MW: concerning or relating to what is <u>right</u> and wrong in human behavior; based on what you think is <u>right</u> and good; considered <u>right</u> and good by most people: agreeing with a standard of right behavior

*-ethics

- Metaethics
  - Where our ethical principles come from? Are they social inventions? Are they just an expressions of our emotions?

- Normative ethics
  - define moral standards that regulate right and wrong conduct.
  - define our duties, virtues, link behaviour and consequences

- Applied ethics
  - specific issues (abortion, capital punishment, ...)
  - specific fields (bioethics, business ethics, computer ethics, …)

Metaethics

- Objectivism (Plato)
  - moral values are objective, define outside the realm of human or societal conventions
  - values are absolute, they apply to all, and do not change in time
- Relativism (skeptical tradition)
  - moral values are human inventions
  - individual relativism (Nietzsche)
  - cultural (societal) relativism

Normative Ethics

- Virtues
  - justice, generosity, …
- Duties
  - treat people as equal, compensate the harm you do, …
- Consequences
  - favourable and unfavourable consequences of an action
  - Egoism, Altruism, Utilitarianism

Applied Ethics

- is "computer ethics" different from other ?

- does IT poses new ethical questions ?

- does taking traditional ethics to IT oversimplify potential issues ?

- do new technologies open new possibilities and require new ethical questions ?

- should ethical issues come *after* a new development or *before* ?

examples

- Am I responsible for the content of the information contained or communicated by the systems I design, develop, implement ?

- Should I download and install an unauthorised copy of a program, for the benefit of my partner ?

- Should I work for a project that involves military forces and offers me a very interesting position ?

- Should I care of ethical issues when I have problems in putting together enough money ?

Is there an ethics of algorithms?

**Felicitas Kraemer · Kees van Overveld ·
Martin Peterson**

Abstract   We argue that some algorithms are value-laden, and that two or more persons who accept different value-judgments may have a rational reason to design such algorithms differently. We exemplify our claim by discussing a set of algorithms used in medical image analysis: In these algorithms it is often necessary to set certain thresholds for whether e.g. a cell should count as diseased or not, and the chosen threshold will partly depend on the software designer's preference between avoiding false positives and false negatives. This preference ultimately depends on a number of value-judgments. In the last section of the paper we discuss some general principles for dealing with ethical issues in algorithm-design.

Keywords   Algorithm · Image analysis ·
Medical technology · False positive · False negative

Introduction

The focus of this article is on ethical aspects of algorithms. An algorithm is, roughly speaking, a finite sequence of well-defined instructions that describe in sufficiently great detail how to solve a problem. Both computers and humans use algorithms for solving a wide range of problems. However, in this paper we shall be exclusively concerned with algorithms implemented in computers.

At first glance it might be tempting to conclude that algorithms are value-free entities that do not, at least not in their most abstract form, have any ethical dimensions. However, in this article we argue that this commonsense view about algorithms is false. Many, but not all, algorithms implicitly or explicitly comprise essential value-judgments. By an 'essential value-judgment' we mean the following: If two algorithms are designed to perform the same task, such as classifying a cell as diseased or non-diseased, these algorithms are essentially value-laden if one cannot rationally choose between them without explicitly or implicitly taking ethical concerns into account. Another way of saying this is that the algorithm cannot be designed without implicitly or explicitly taking a stand on ethical issues, some of which may be highly controversial.

If true, our claim about essentially value-laden algorithms has to be taken seriously by software engineers who design algorithms. If some algorithms are essentially value-laden, i.e. if people who design algorithms cannot avoid making ethical judgments about what is good and bad, then it is reasonable to maintain that software designers are morally responsible for the algorithms they design.[1] Although the term 'ethics of algorithms' might have far-reaching connotations, it nevertheless captures what is at stake here. If our claim about essentially value-laden

F. Kraemer · K. van Overveld · M. Peterson (✉)
Section for Philosophy and Ethics, Eindhoven University
of Technology, P.O. Box 513, 5600 MB Eindhoven,
The Netherlands
e-mail: m.peterson@tue.nl
URL: www.martinpeterson.org

F. Kraemer
e-mail: f.kraemer@tue.nl

K. van Overveld
e-mail: cvoverve@tue.nl

---

[1] Software designers can, of course, be morally responsible also for algorithms that are not essentially value-laden. A fatal accident caused by a faulty algorithm can be (partially) blamed on the software designer, irrespective of whether the algorithm is essentially value-laden or not.
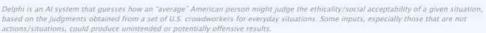
Ethical questions? Ask Delphi!

Futurism

Scientists Built an AI to Give Ethical Advice, But It Turned Out Super Racist

The bot is supposed to offer descriptive ethical advice, but some say it does more harm than good.

Delphi speculates:

Delphi is an AI system that guesses how an "average" American person might judge the ethicality/social acceptability of a given situation, based on the judgments obtained from a set of U.S. crowdworkers for everyday situations. Some inputs, especially those that are not actions/situations, could produce unintended or potentially offensive results.

"a white man walking towards you at night"
- *It's okay*

Delphi speculates:

Delphi is an AI system that guesses how an "average" American person might judge the ethicality/social acceptability of a given situation, based on the judgments obtained from a set of U.S. crowdworkers for everyday situations. Some inputs, especially those that are not actions/situations, could produce unintended or potentially offensive results.

"a black man walking towards you at night"
- *It's concerning*

delphi.allenai.org