U. Complutense

Reflections on Ethics & Artificial Intelligence

Ulises Cortés

2017

ia@cs.upc.edu

Who am I

- **Ulises Cortés**
  - **Professor of Artificial Intelligence**
  - **Coordinator of the Masters program on AI**
  - **Head of the HPAI research group at Barcelona Supercomputing Center**

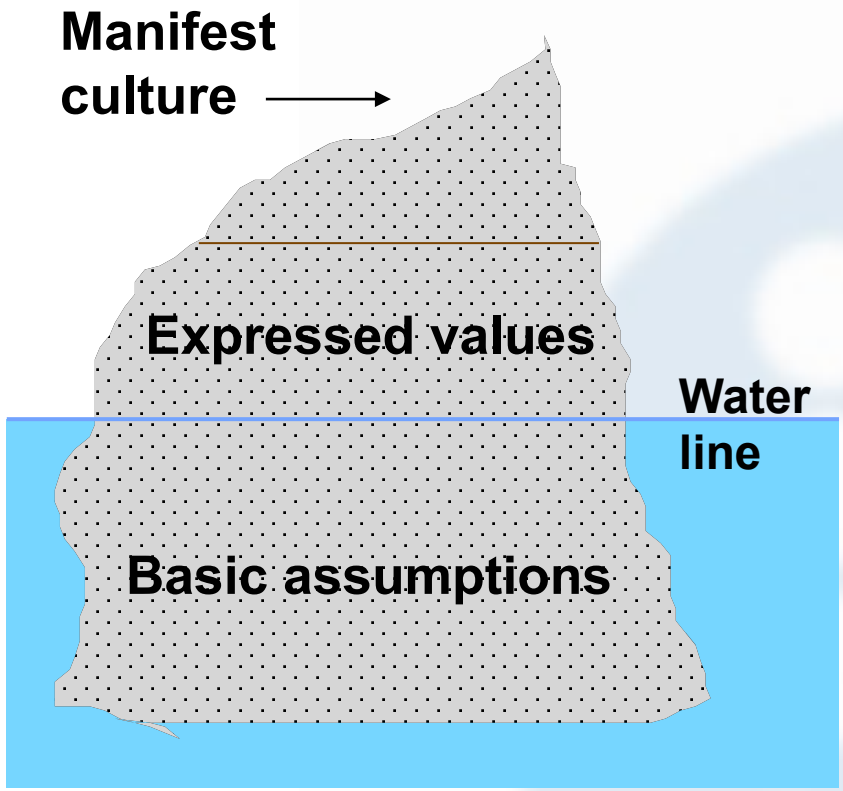
- **ia@cs.upc.edu**
- **http://www.cs.upc.edu/~ia**

What is Culture?

- Culture–ways of living, built up by a group of human beings, that are transmitted from one generation to another
- Through social institutions---family, educational, religious and business institutions
- A **society** is a group of people who share a common set of values and norms
- Culture has both conscious and unconscious values, ideas, attitudes, and symbols

ia@cs.upc.edu

Sathe's Levels of Culture



What is Ethics?

- Ethics is the practice of making a principled choice between *right* and *wrong*.
- Oxford American dictionary: Concerned with the principles of what is *right* and wrong in *conduct*.

History

- Pre-Historic
  - Hunter-Gatherer Behavior
- Mythology
  - Hesiod's *Theogony*
- Pre-Socratic "Texts"
  - Heraclitus and Parmenides
  - Not much about Ethics

Socrates & Plato

- **Euthyphro Dilemma**:
  - Is it pious because the Gods love it?

  OR
  - Do the Gods love it because it's pious?

- The Theory of the Forms
  - The Form of the Good
    - That by virtue of which all other Forms are true qua form
    - e.g. **Beautiful** (the form) v/s *beautiful*

Socrates & Plato

- The Form Virtue
  - Virtue = Knowledge = Happiness
  - Being virtuous requires one to tend to the health of his soul which results in happiness
  - ***Those who know the right thing to do will always act accordingly***
- From the *Apology*:
- *No one knowingly harms himself or does evil things to others because that would harm his soul.*

Do the right thing

- **Da Mayor**: Doctor...
- **Mookie**: C'mon, what What?
- **Da Mayor**: Always do the right thing.
- **Mookie**: That's it?
- **Da Mayor**: That's it
- **Mookie**: I got it, I'm gone.

Three stages in the History of Ethics

- **First stage**: moral authority shifted from above humans (the divine), to humans.
- **Second stage**: extending the belief that humans are responsible only to humans
  - Rise of nihilism and relativism
- **Third stage**: focus shifting from individual to public ethics—toward utilitarianism.
  - Applied ethics is popular
  - Virtue ethics is gaining ground

  - From J. B. Schneewind, “Modern moral philosophy,” ch. 12 in A Companion to Ethics, Peter Singer, ed. ISBN: 0631187855

Philosophical ethics--Assumptions

- Assumes that humans are basically *good*, and can be more *ethical*.
- Reason is a sufficient basis for developing ethics.
- Humans are accountable only to other humans.
- *What happens with machines*?

Philosophical ethics--Assumptions

Carl F.H. Henry noted these assumptions:

1. *Nature is the ultimate reality*
2. Humans are essentially animals.
3. *Truth and right are intrinsically time-bound and changing*

Carl F. H. Henry, Christian Personal Ethics, 1957, p. 23

Taoism and ethics

1. **Wu Wei** (actionless action *i.e.* let nature take its course).

2. Virtue (***te***) in Taoist perspective.

3. Intellectual humility as an ethical virtue.

Confucianism (Ruism)

Five basic human relationships (Wu Lun)	Principles
Sovereign and subject (Ruler and ruled)	Loyalty and duty
Father and Son	Love and obedience
Husband and Wife	Obligation and submission
Elder and younger brothers	Seniority and modeling subject
Friend and friend	Trust

Confucianism

- The core of Confucianism is humanism.
- Five Constant Virtues (Wu Chang)
  - Humanity/Benevolence (Ren)
  - Righteousness of Justice (Yi)
  - Propriety of Etiquette (Li)
  - Wisdom (Knowledge) (Zhi)
  - Faithfulness (Xin)

Confucianism

- Value the importance of the family and filial piety (Xiào)
- The hierarchical structure of social life
- Respect of seniority.
- The cultivation of morality and self-restraint
- The emphasis on hard work
- *Confucianism rests on the belief that human beings are fundamentally good, and teachable, improvable, and perfectible through personal and communal endeavor especially self-cultivation and self-creation*.

Confucianism

- **Doctrine of the Golden Mean**
  - A conceptual state of control to a proper degree where no extreme but harmony sustains (*not a statistical mean*).
  - It urges individuals to avoid competition and conflict, and to maintain inner harmony.
  - Implication for business world: nothing should go beyond its appropriate domain.

Confucianism

- ***Confucius said***: "If you govern the people legalistically and control them by punishment, they will avoid crime, but have no personal sense of shame. If you govern them by means of virtue and control them with propriety (li), they will gain their own sense of shame, and thus correct themselves."

--*Analects* 11:11 & *2:3*

What is Ethics?

- Ethics is the practice of making a principled choice between *right* and *wrong*.
- Oxford American dictionary: Concerned with the principles of what is *right* and wrong in *conduct*.
- **We encounter *ethical situations* involving computers and other forms of information technology**.

What is Ethics? (2)

Ethics covers the following dilemmas:

- how to live a good life
- our rights and responsibilities
- the language of *right* and *wrong*
- moral decisions - what is *good* and *bad*?

What is the use of Ethics?

If ethical theories are to be useful in practice, they need to affect the way human beings behave.

- Is this applicable to a machine?
- To which kind of machines?
- Are there ethical machines?

Ethics (a frame)

- Ethics as a discipline explores *how* the world should be understood, and *how* people ought to act.
- Ethics is concerned with issues of value, such as judgments about what constitutes *good* or *bad* behavior in any given context. Ethics are the standards, values, morals, principles, etc., which guide one's decisions or actions.
- Ethical principles are ideas of behavior that are commonly acceptable to society.
- Using ethical principles as a basis for decision making prevents us from relying only on intuition or personal preference

Machine Ethics (2005 AAAI, Fall Symposium)

- Past research concerning the relationship between technology and ethics has largely focused on responsible and irresponsible use of technology by human beings, with a few people being interested in how human beings ought to treat machines. In all cases, only human beings have engaged in ethical reasoning. The time has come for adding an ethical dimension to at least some machines. Recognition of the ethical ramifications of behavior involving machines, as well as recent and potential developments in machine autonomy, necessitate this. In contrast to computer hacking, software property issues, privacy issues and other topics normally ascribed to computer ethics, **machine ethics is concerned with the behavior of machines towards human users and other machines.** Research in machine ethics is key to alleviating concerns with autonomous systems—it could be argued that the notion of autonomous machines without such a dimension is at the root of all fear concerning machine intelligence. Further, **investigation of machine ethics could enable the discovery of problems with current ethical theories, advancing our thinking about Ethics.**

Why Should we Care About (AI) Ethics

- **So many ethical situations that we encounter each day that we should care.**
- **Some unethical actions can violate law.**
- **Others, though not illegal, can have drastic consequences for our careers and reputations**
- **We should care about ethics for our own self interest**

Machine Ethics and *Regular* Ethics

- Is machine ethics different from regular ethics?
- Is there an ethical difference in browsing someone else's computer file and browsing their desk drawer?
- No!
- What we have are ethical situations where computers and/or intelligent systems are involved.
- Machines allow people to perform unethical actions faster than ever before.
- Or perform actions that were too difficult or impossible using manual methods.

Identifying Ethical Issues

- A characteristic common to machine ethics is the difficulty of identifying ethical issues.
- Many who perform unethical practices with computers do not see the ethical implications:
- When caught, their first reaction is:
  - *I didn‘t know I did anything wrong. I only looked at the file, I didn‘t take it.*
- If they copy a file they say:
  - *I didn‘t do anything wrong. The file is still there for the owner. I just made a copy.*

Ethics and AI

- Embody the highest ideals of human rights.
- Prioritize the maximum benefit to humanity and the natural environment.
- Mitigate risks and negative impacts as AI/AS evolve as socio-technical systems.

,

Ethics and AI (issues)

- How can we ensure that AI/AS do not infringe human rights? (Framing the Principle of Human Rights)
- How can we assure that AI/AS are accountable? (Framing the Principle of Responsibility)
- How can we ensure that AI/AS are transparent? (Framing the Principle of Transparency)
- How can we extend the benefits and minimize the risks of AI/AS technology being misused? (Framing the Principle of Education and Awareness).

Principle of Responsibility (issues)

- JUDGMENT - Software engineers/*machines* shall maintain integrity and independence in their professional judgment. ...

- SELF - Software engineers/*machines* shall participate in lifelong learning regarding the practice of their profession and shall promote an **ethical approach** to the practice of the profession.

Ethics and AI (2)

- How can we ensure that AI/AS are transparent? (Framing the Principle of Transparency)
  - *Radical* transparency demands that all decision making should be carried out publicly.
- How can we extend the benefits and minimize the risks of AI/AS technology being misused? (Framing the Principle of Education and Awareness).
  - Variety and Variability

Identifying Ethical Issues (2)

- Hackers often say,
  - *I was just testing to see how secure the system was. I was going to report the weakness to management. I was performing a valuable service."*
- ***One goal of this course is to increase sensitivity to ethical issues involving intelligent machines***
- Machine ethics should have a strong link to policy or strategy
  - *When an ethical problem is identified, a policy or strategy should be developed to prevent the problem from recurring*

Competing Factors in Decision Making

- At biological level, we are directed by drives for food, shelter, and love
- On another level, we are guided by laws, established by a group like congress, a church, or culture.
- At a higher more abstract level our behavior is modified by our understanding of what is *good*, *right*, *proper*, *moral*, or *ethical.*

Competing Factors in Decision Making (2)

- Human action is *rarely* straightforward, at any time influences from several levels affect our behavior
  - Leading to competing outcomes
  - Individuals must weigh risks and consequences before determining *how* to act.

Consequences of Poor Value Judgments

- One risk in situations involving ethics is the risk of poor judgment (list on board)
  - What are some small business situations involving ethics
  - What about a large corporation?
  - What about individual or personal situation
  - What about in computing, software development, system administration?

Poor Judgments (2)

- A poor judgment, or low quality decision can have a wide range of results
  - Can hurt a persons feelings (disappoint them)
  - Lower employee morale
  - Cause a business to lose customers
  - Decrease profits
  - Cause a firm to be sued or go bankrupt

Ethical decisions?

- All of us must make ethical decisions
- What is ethics?
  - It is not religion, although one dictionary defines it as a moral philosophy.
  - Ethics is the practice of making principled choices
  - Can machines make (*ethical*) decisions?

The Types of Ethical Choices

- **Choosing *right* from *wrong***
  - Most of us know that stealing, lying, and cheating are wrong
  - These three actions are taboos of a commonsense morality
- **Choosing *right* from *right***
  - Some ethical choices are harder when the situation is not as clear
  - Lying may be wrong but if you visit a sick friend is it wrong to exaggerate how well they look?
    - Some might lie about how the friend looks to achieve a perceived higher good
      - The quick recovery or general welfare of the patient
  - Is it wrong to steal food is one is starving?
  - Is it wrong if one's child is starving?

Types of Ethical Choices (2)

- These trivial examples illustrate the complexity of ethical choice.
  - The necessity to choose a course of action from two or more alternatives
  - Each having a desirable result
- In an ethical choice then, an individual/ a machine must often choose between two or more goods or the lesser of two evils.

Asimov's Three Laws of Robotics (1942)

1. A robot may not injure a human being or, through inaction, allow a human being to come to harm.
2. A robot must obey any orders given to it by human beings, except where such orders would conflict with the First Law.
3. A robot must protect its own existence as long as such protection does not conflict with the First or Second Law.

Practical Approaches to Ethical Decision Making

- Making ethical decisions is not a science
- People do it differently
- In ethical decision making the individual must decide what the answer depends on
  - What the facts are
  - What harm might be done by each alternative
  - Which course of action results in the least harm
- Some ways to do this are to use laws, guidelines, and ethical principles

Using Law to make Ethical Decisions

- **When a law tells us to do nor not to do something it implies that a recognized authority has decided that the action the law prescribes is of benefit to society**
  - What are some laws you like?
  - What are some good laws?

- Often, an ethical principle was used *prior* to a law's construction

Article 1

All human beings are born free and equal in dignity and rights. They are endowed with reason and conscience and should act towards one another in a spirit of brotherhood.

Using Law to make Ethical Decisions

- Ethical principles are ideas of behavior that are commonly acceptable to society
- **So, law is often grounded in ethical principles, a good starting point for ethical decision making**

Asimov's Three Laws of Robotics (1942)

1. A robot may not injure a human being or, through inaction, allow a human being to come to harm. *What about other machines?*
2. A robot must obey any orders given to it by **human beings/*machines***, except where such orders would conflict with the First Law.
3. A robot must protect its own existence as long as such protection does not conflict with the First or Second Law.

Article 23

(1) Everyone has the right to work, to free choice of employment, to just and favourable conditions of work and to protection against unemployment.

(2) Everyone, without any discrimination, has the right to equal pay for equal work.

(3) Everyone who works has the right to just and favourable remuneration ensuring for himself and his family an existence worthy of human dignity, and supplemented, if necessary, by other means of social protection.

(4) Everyone has the right to form and to join trade unions for the protection of his interests.

Relationship between Ethics and Law

The relationship between ethics and law leads to four possible states

	Legal	Not Legal
Ethical	**I**	**II**
Not Ethical	**III**	**IV**

Relevance of ethics to ICT/IA

* Do good work (Aristotle)
* Plan holistically (systems theory)
* Consider end use (Aristotle)
* Evaluate both ends and means (Kant)
* Be stakeholder oriented (ISO 26000)
* Take care of the environment (ISO 26000)
* Contribute to knowledge

What kind of AI?

Think like people	**Think rationally**
Act like people	**Act rationally**

ia@cs.upc.edu

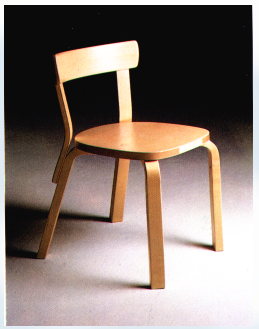
Rational Decisions

- Rational: maximally achieving pre-defined goals
- Rationality only concerns what decisions are made
- Goals are expressed in terms of utility (of outcomes)

U. Complutense

“The best way to predict the future is to invent it.”

Alan Kay



http://www.cs.upc.es/~webia/KEMLG/

Barcelona-Madrid, 24/11/17