Paper by: Ville Vakkuri(B) , Kai-Kristian Kemell , Marianna Jantunen , and Pekka Abrahamsson
(2020)

“This is Just a Prototype”: How Ethics Are Ignored in Software Startup-Like Environments

Nikita Sade

Introduction / Abstract

Artificial Intelligence (AI) is everywhere in software and it helps make important decisions!

- AI can oftentimes operate underneath the surface of different types of software.
- AI is so unique because it can exert a “society wide influence” and stakeholders either do not realize that AI is in use and/or do not have the option to not use them.

The authors of this paper have found that there are very few and limited research methods for implementing AI ethics in software development.

Background: The Current State of AI Ethics

- Most of the current Ethics In AI works has been theoretical.
    - Received little attention because...
        1) Prior research was philosophical
        2) The field has not sensed the need to address ethical issues
- Vallach & Allen argue that AI based systems produce new requirements to consider. The propose that designers implicitly embed values in the technology the create.
- AI ethics is not considered “mainstream” yet
- As of this paper...
    - Places like France, Germany, the EU
    - Organizations like ISO, IEEE, ACM
    - Large companies like Microsoft, Google

All have AI ethics guidelines

However, the existing literature shows that AI developers lack the professional norms and methods to translate principles and guidelines in a useful or successful way.

Research Question

“How are AI ethics taken into consideration in software engineering projects when they are not formally considered?”

Underlying Arguments:

AI ethics literature is all theoretical and our understanding of the state of practice in AI ethics is lacking.

When ethics are not taken into consideration when developing AI based software systems there are potential socio-ethical issues that will arise.

Bridging the gap between research and practice in AI ethics is an ongoing issue.

About the Study

	Example	Participants
Case A	Statistical tool for detecting social marginalization	Data Analyst [RI], Consultant [R2], Project Coordinator [R3]
Case B	Speech recognition and NLP based tool for diagnostics	Developer [R4], Developer [R5], Project Manager [R6]
Case C	NLP based tool for indoor navigation	Developer [R7], Developer [R8]

Descriptions of each case

What is the study?

Multiple case study of 3 projects focused on AI healthcare systems

- Health care chosen specifically because due to the nature of human interactions ethical considerations should be higher
- Startup like nature due to ...
    - Agile methods, notable time pressure, scarce resources, development of functional prototypes

About the Study

Data Collection

- Semi-structured interviews to allow for flexibility
- Interviews have audio and were transcribed
- Conducted in Finnish

Data Analysis

- Grounded Theory: code the transcripts quote by quote
  - Why? Due to the lack of existing studies on the current state of practice in the area
- Commitment Model of Abrahamsson
  - Focuses on actions and concerns
    - Concerns: what ethical issues were of interest to the developer?
    - Actions: studied to understand how these concerns were addressed or if they were addressed at all
  - Every action has a concern but not every concern has an action.
  - No action = lack of commitment towards tackling those concerns.
  - Concerns and actions of each respondent was compared across cases in search of RECURRING concerns and actions between cases and respondents, understand the motivation behind the actions

About the Study

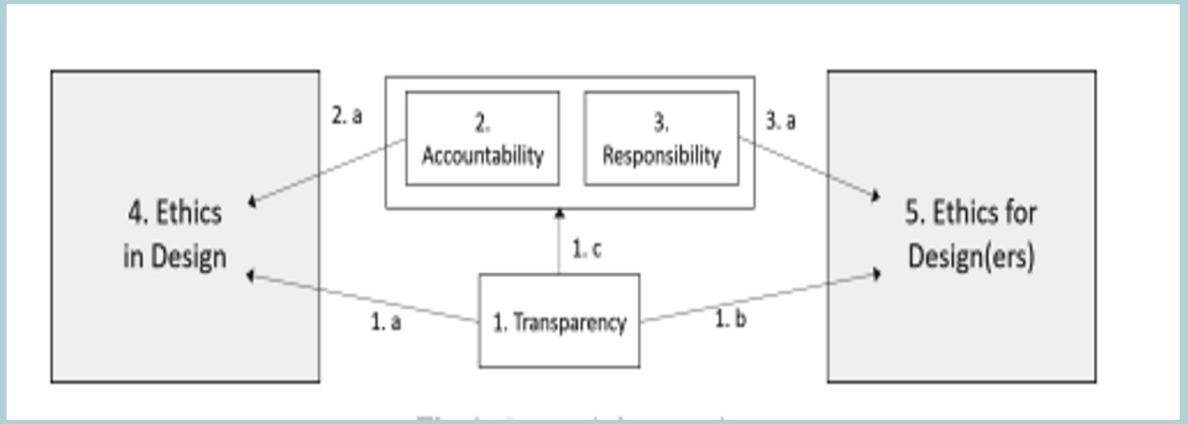
What is the goal of the study?

Help bridge the gap between theoretical and practical AI ethics

1) Help understand the current state of practice
2) Discover existing good practices that might help implement AI ethics

Research Model

ART Principles of Dignum



Ethics IN Design

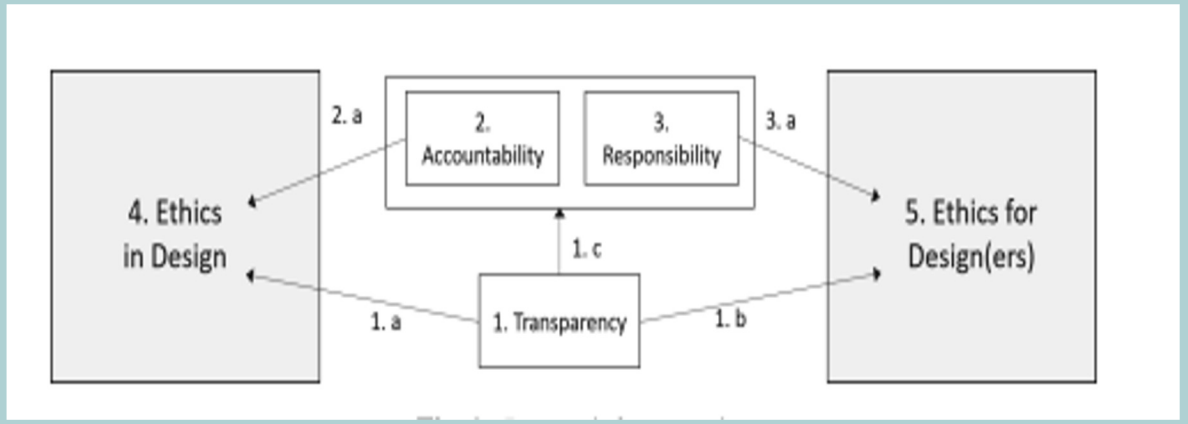
Regulatory and engineering methods for supporting ethical implications of AI systems.

Ethics FOR Design(ers)

Codes of conduct and standard for integrity and ethics

Research Model

ART Principles of Dignum



Transparency

Must understand **WHY** the system acts in a certain fashion and **WHO** made **WHAT** decisions

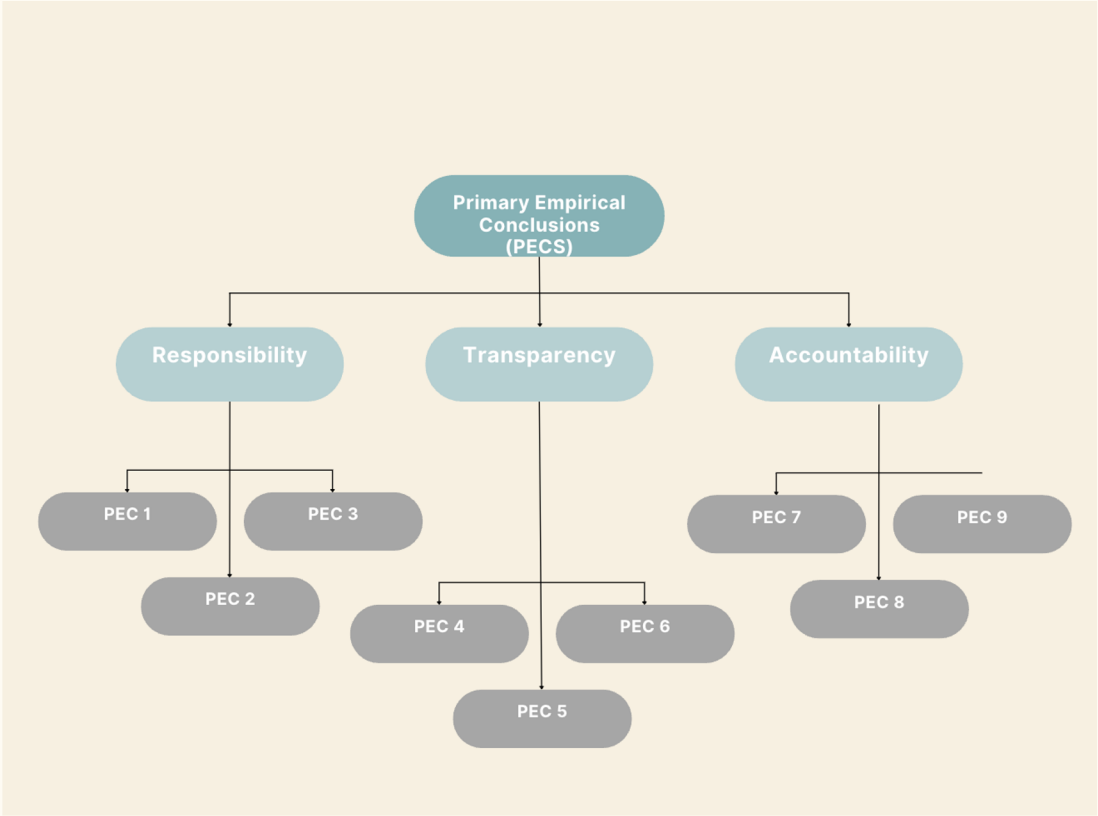
Accountability

WHO is accountable/ liable for the decisions and explanations of actions to stakeholders

Responsibility

A chain of responsibility that links actions of systems to all decisions.

Major Findings



Major Findings - Responsibility

#	Theoretical component	Description	Contribution
1	Responsibility	Developers feel most responsibility towards tackling problems related to software development, such as finding bugs, meeting project goals	Empirical validation
2	Responsibility	On a personal level, developers are concerned about the ethical aspects of product development. However, little is done to tackle these concerns	Novel
3	Responsibility	Responsibility of developers is unclear	Novel

- Developer concerns varied, but was not concerned specifically about ethics
- Concerns were detached from work
- Developers did not consider potential harm of system outside tangible harm

Major Findings - Transparency

4	Transparency	Black box systems are a typical issue in AI development	Empirical validation
5	Transparency	Developers recognize transparency as a goal, but it is not formally pursued	Contradicts existing literature
6	Transparency	Established SE practices, such as code documentation and code review, support transparency of systems development	Empirical validation

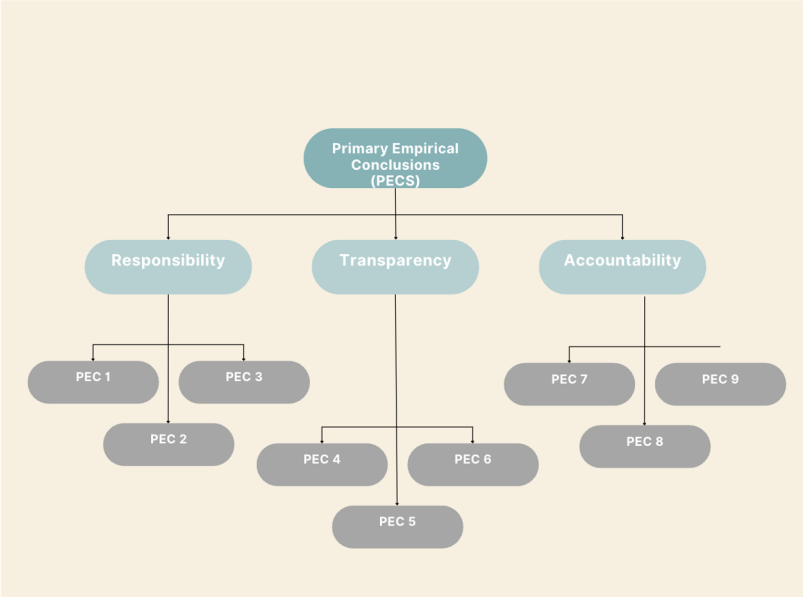
- Highlighted the importance of expertise in math
  - Cannot understand what you do not know
- Proper Documentation, version control a small team size all helped increase transparency

Major Findings - Accountability

7	Accountability	Developers feel accountable for error handling and have the means to deal with it	Empirical validation
8	Accountability	Product misuse and error scenarios are only considered during development. They are not considered in terms of the future operational life of the system out on the field	Contradicts existing literature
9	Accountability	Developers do not have plans to deal with unexpected behavior of the system resulting from e.g. machine learning or the future expansion of the use context of the system	Contradicts existing literature

- Accountability related to security was a high priority
- Basic error handling was covered however, potential threats was not an active concern

Major Findings - Summary



- Ethics were not implemented by a formal method or tool.
- Other issues led to the ethical problems being addressed.
- When prompted developers understood there was negative effects but had no direction on how to fix them.
    - Only noticed the issue when being prompted to look for them.

Strengths + Weakness of Study

Strengths

- Was able to find common theme among many answers while still providing evidence to support claims
- Strong framework to analyze results
- Organized findings well, was able to see what was new, different, or the same from past literature

Weaknesses

- Qualitative multi-case study - problems for generalizability of the data
- Cases based in Finland, other countries may have different approaches
- Language barrier - was translated from Finnish to English, some connotations of words may have changed

Discussion Points

- The study finds that when an ethical issue is tied to a practical issue it is more likely to be considered. Should AI ethics be a separate tool? Or would it be more effective to incorporate additional steps into existing methodologies?

- As AI becomes more and more incorporated into software, should there be a push for a uniform code of ethics? Or is the code of AI something that should be decided on a smaller scale (i.e. company, country..) ?

- If a similar study was conducted in a different industry (ie: law enforcement, education...) or in a different location (ie. United States, China ..) would there be similar or different results?

Personal Thoughts, Reflections, and Conclusion

- The study underlines a clear gap between the research and practice of ethics in AI. Further studies need to be done to help bridge this gap.
- Developers do not actively think about the ethical implications of the product they are making, there needs to be a tool or system in place to enforce that.
- Sometimes people get so caught up in making the *product* they forget the real world impact.
- There are consequences to having a faulty prototype, especially one that is used by the public.

Thank you!