Accountability in practice:

Critical insights from a case-study in Scandinavia of the development of applied AI for clinical healthcare

By Anne Henriksen & Simon Enni
Clarifications...

Artificial intelligence (AI)

• ‘AI’ refers to machine learning techniques used for developing algorithms that are trained over time to provide the best prediction for a specific task (Russell & Norvig, 2018)

Developers

• In the research case, the developing team consists of both a range of different professionals. However, in this presentation, ‘developer’ mainly refers to data engineers.
Introduction

• Interest: Automated decision making & accountability in practice

• Healthcare: a ‘high stakes zone’ where an accountability system urgently needs to be in place!

• Mechanisms for accountability in an accountability system need to reflect the design and development practice characterizing AI
Research question

How is the encounter between, on the one hand, ethical principles, certification standards, and explanation methods, and, on the other hand, the practice context of AI design and development?

How do these three mechanisms function to implement and ensure accountability in practice?
Theoretical conception

Ensuring accountability is about ensuring that those bearing a certain responsibility are acting accountable in accordance with that responsibility, and can account for their actions (demonstrate accountability) by providing accounts. To whom accounts are provided depends on the purpose of the account.

(Bryson & Theodorou, 2019; Bryson, 2018; Bryson & Winfield, 2019; Bryson, Diamantis & Grant, 2017; Bovens, 2007; Garfinkel, 1967, 1984)
Research design, methods & empirical data

Case
Scandinavian AI company designing and developing applied AI for healthcare and accounting

Research approach
• Ethnographic, explorative

Method
• Participant observation
• Research interviews
• Document analysis

Empirical data
• Participation in project meetings over a period of 14 months
• 6 months of day-to-day participant observation in AI company
• 36 recorded interviews and meetings
• 15 recordings fully transcribed
• Pictures from design workshops
• Documents (project proposal, powerpoints, etc.)
“Even though it might sound inhumane, there is not much different from a data point of view between moving a patient through the healthcare system and an invoice through the accounting system”

(Interview with CEO of AI company in online article, Aug. 2020)
Empirical insights & findings
1. Ethical principles I

Developers do not regard principles as a lever for accountability but are highly critical of them due to frustration at the extent to which and the way in which ethics are being discussed and outlined. Why?

1. The discussion does not reflect the developers’ understanding of AI but quite the contrary
2. Ethical principles are discussed on a very general level while distinctions between types of decisions are important in a development context

Hence, principles are considered by developers to base on misunderstandings of what AI is and what kind of decisions AI should operate.
1. Ethical principles II

The general misunderstanding of what AI is has really surprised me. Really, AI is just statistics on speed and nothing more than that. I don’t understand why people question what AI is but don’t question, for example, what an MRI is, because, in my opinion, MRI is just as unstable as an algorithm may be... It’s not that I am against legislation but I just think the general discussion is too generalizing and stereotypical, and is missing the point. In fact, I think it is damaging to the work that we’re actually doing.

(Engineer, Feb. 2020)

Consequences:

Principles do not help developers manage issues on the ground and are regarded by them as irrelevant and almost as a hindrance to their work.
2. Certifications & auditing I

Developers are highly motivated for applying for certifications as a means of verifying the quality of their products and business. Still, this is not a fully well-functioning mechanism for accountability. Why?

1. Certifications require a huge effort by developers while guidelines for how to comply to standards are lacking

2. The auditors conducting certifications in the case do not have sufficient knowledge of AI and ML

Hence, the developers spend a lot of energy on preparation but are met by a certification body that generally lacks resources.
2. Certifications & auditing II

We are about to apply for the certification in ISO-13485 but there is absolutely *nothing* for us to follow in order to implement the standard. There are some FDA guidelines from the US; we're not even ready in the EU yet! Seriously, wake up, please!

(Director, Feb. 2020)

I think it's extremely frustrating that the national and European authorities release ethical guidelines and make all kinds of statements instead of just rolling up their sleeves! Seriously, we cannot allow ourselves to just do nothing! Really, we do *everything* that we can because we know that our business *cannot* survive if is compromised by any of this.

(Director, Feb. 2020)

**Consequences:**

Certifications become a matter of pro forma/rubber stamping and may undermine the motivation the ideal of accountability.
3. Explanation methods

Developers rely on xAI as a means of producing accounts in the clinical setting, however this changes. Why?

1. xAI provides *functional* explanations of the interpretation of predictive models at an *average* level

2. Explanations may counteract the effect of implementing AI in a clinical setting

Hence, there may be a pure fit between the logics of explanation and xAI, and the need for explanation in a clinical setting.
3. Explanation methods

“In principle, this is an explanation of the model’s interpretation of a patient suffering from sepsis, however this is not necessarily an interpretation that reflects reality”.

“It’s not an explanation of the complexity of the model. Rather, it’s an explanation of the reality of the model”.

“There has to be an explanation in all cases – I think that is a general misunderstanding of what we are developing”.

(Engineer, Feb. 2020)

Consequences:

We cannot rely on the current version of xAI as a stable mechanism for accountability in the usage situation and may rethink ‘explanation’.
Some considerations...

• Ethical principles need to be discussed on the ground if developers are to adopt them (more case based, less principled)

• The development of further certification standards will not help?

• If certifications are to play a role in the demonstration of ethical conduct, it is crucial that resources are prioritized.

• We should not expect that the technical method of xAI will solve problems of accountability in practice.

• Are accountability operationalized and implemented in such a way that we risk eroding this value and ideal which we urgently need developers to aspire for?
Accountability in practice:

Critical insights from a case-study in Scandinavia of the development of applied AI for clinical healthcare

By Anne Henriksen & Simon Enni
Clarifications...

Artificial intelligence (AI)

• ‘AI’ refers to machine learning techniques used for developing algorithms that are trained over time to provide the best prediction for a specific task (Russell & Norvig, 2018)

Developers

• In the research case, the developing team consists of both a range of different professionals. However, in this presentation, ‘developer’ mainly refers to data engineers.
Introduction

• Interest: Automated decision making & accountability in practice

• Healthcare: a ‘high stakes zone’ where an accountability system urgently needs to be in place!

• Mechanisms for accountability in an accountability system need to reflect the design and development practice characterizing AI
Research question

How is the encounter between, on the one hand, ethical principles, certification standards, and explanation methods, and, on the other hand, the practice context of AI design and development?

How do these three mechanisms function to implement and ensure accountability in practice?

1. Ethical principles
2. Certifications & auditing
3. Explanation methods
Theoretical conception

Ensuring accountability is about ensuring that those bearing a certain responsibility are acting accountable in accordance with that responsibility, and can account for their actions (demonstrate accountability) by providing accounts. To whom accounts are provided depends on the purpose of the account.

(Bryson & Theodorou, 2019; Bryson, 2018; Bryson & Winfield, 2019; Bryson, Diamantis & Grant, 2017; Bovens, 2007; Garfinkel, 1967, 1984)
Research design, methods & empirical data

Case
Scandinavian AI company designing and developing applied AI for healthcare and accounting

Research approach
• Ethnographic, explorative

Method
• Participant observation
• Research interviews
• Document analysis

Empirical data
• Participation in project meetings over a period of a 14 months
• 6 months of day-to-day participant observation in AI company
• 36 recorded interviews and meetings
• 15 recordings fully transcribed
• Pictures from design workshops
• Documents (project proposal, powerpoints, etc.)
“Even though it might sound inhumane, there is not much different from a data point of view between moving a patient through the healthcare system and an invoice through the accounting system”

(Interview with CEO of AI company in online article, Aug. 2020)
Empirical insights & findings
1. Ethical principles I

Developers do not regard principles as a lever for accountability but are highly critical of them due to frustration at the extent to which and the way in which ethics are being discussed and outlined. Why?

1. The discussion do not reflect the developers’ understanding of AI but quite the contrary
2. Ethical principles are discussed on a very general level while distinctions between types of decisions are important in a development context

Hence, principles are considered by developers to base on misunderstandings of what AI is and what kind of decisions AI should operate.
The general misunderstanding of what AI is has really surprised me. Really, AI is just statistics on speed and nothing more than that. I don’t understand why people question what AI is but don’t question, for example, what an MRI is, because, in my opinion, MRI is just as unstable as an algorithm may be... It’s not that I am against legislation but I just think the general discussion is too generalizing and stereotypical, and is missing the point. In fact, I think it is damaging to the work that we’re actually doing.

(Engineer, Feb. 2020)

Consequences:

Principles do not help developers manage issues on the ground and are regarded by them as irrelevant and almost as a hindrance to their work.
2. Certifications & auditing I

Developers are highly motivated for applying for certifications as a means of verifying the quality of their products and business. Still, this is not a fully well-functioning mechanism for accountability. Why?

1. Certifications require a *huge effort* by developers while guidelines for how to comply to standards are lacking

2. The auditors conducting certifications in the case do not have sufficient knowledge of AI and ML

Hence, the developers spend a lot of energy on preparation but are met by a certification body that generally lacks resources.
2. Certifications & auditing II

Consequences:

Certifications become a matter of pro forma/rubber stamping and may undermine the motivation the ideal of accountability.

We are about to apply for the certification in ISO-13485 but there is absolutely nothing for us to follow in order to implement the standard. There are some FDA guidelines from the US; we're not even ready in the EU yet! Seriously, wake up, please!

(Director, Feb. 2020)

I think it’s extremely frustrating that the national and European authorities release ethical guidelines and make all kinds of statements instead of just rolling up their sleeves! Seriously, we cannot allow ourselves to just do nothing! Really, we do everything that we can because we know that our business cannot survive if is compromised by any of this.

(Director, Feb. 2020)
3. Explanation methods

Developers rely on xAI as a means of producing accounts in the clinical setting, however this changes. Why?

1. xAI provides *functional* explanations of the interpretation of predictive models at an *average* level

2. Explanations may counteract the effect of implementing AI in a clinical setting

Hence, there may be a pure fit between the logics of explanation and xAI, and the need for explanation in a clinical setting.
3. Explanation methods

“In principle, this is an explanation of the model’s *interpretation* of a patient suffering from sepsis, however this is *not* necessarily an interpretation that reflects *reality*.”

“It’s not an explanation of the complexity of the model. Rather, it’s an explanation of the *reality* of the model”.

“There has to be an explanation in all cases – I think that is a general misunderstanding of what we are developing”.

(Engineer, Feb. 2020)

**Consequences:**

We cannot rely on the current version of xAI as a stable mechanism for accountability in the usage situation and may rethink ‘explanation’.
Some considerations...

• Ethical principles need to be discussed on the ground if developers are to adopt them (more case based, less principled)

• The development of further certification standards will not help?

• If certifications are to play a role in the demonstration of ethical conduct, it is crucial that resources are prioritized.

• We should not expect that the technical method of xAI will solve problems of accountability in practice.

• Are accountability operationalized and implemented in such a way that we risk eroding this value and ideal which we urgently need developers to aspire for?
thank you

Twitter @AnneHenriksenDK
LinkedIn Anne Henriksen
Mail annehenriksen@cc.au.dk