JYVÄSKYLÄN YLIOPISTO
UNIVERSITY OF JYVÄSKYLÄ

Implementing Ethics in AI

Professor Pekka Abrahamsson

15th International Conference on Emerging Technologies for a Smarter World (CEWIT 2019)
Nov-6th, 2019, Stony Brook, NY, USA

Agenda

- Motivation
- State-of-the-art tools and methods
- Empirical observations
- Conclusions
- References

Failing AI

- “San Francisco Bans Facial Recognition Technology”

Source: New York Times, 2019, bit.ly/2Qs9vKi

- Amazon scraps secret AI recruiting tool that showed bias against women

Source: Reuters, 2018, bit.ly/2Kw77hK

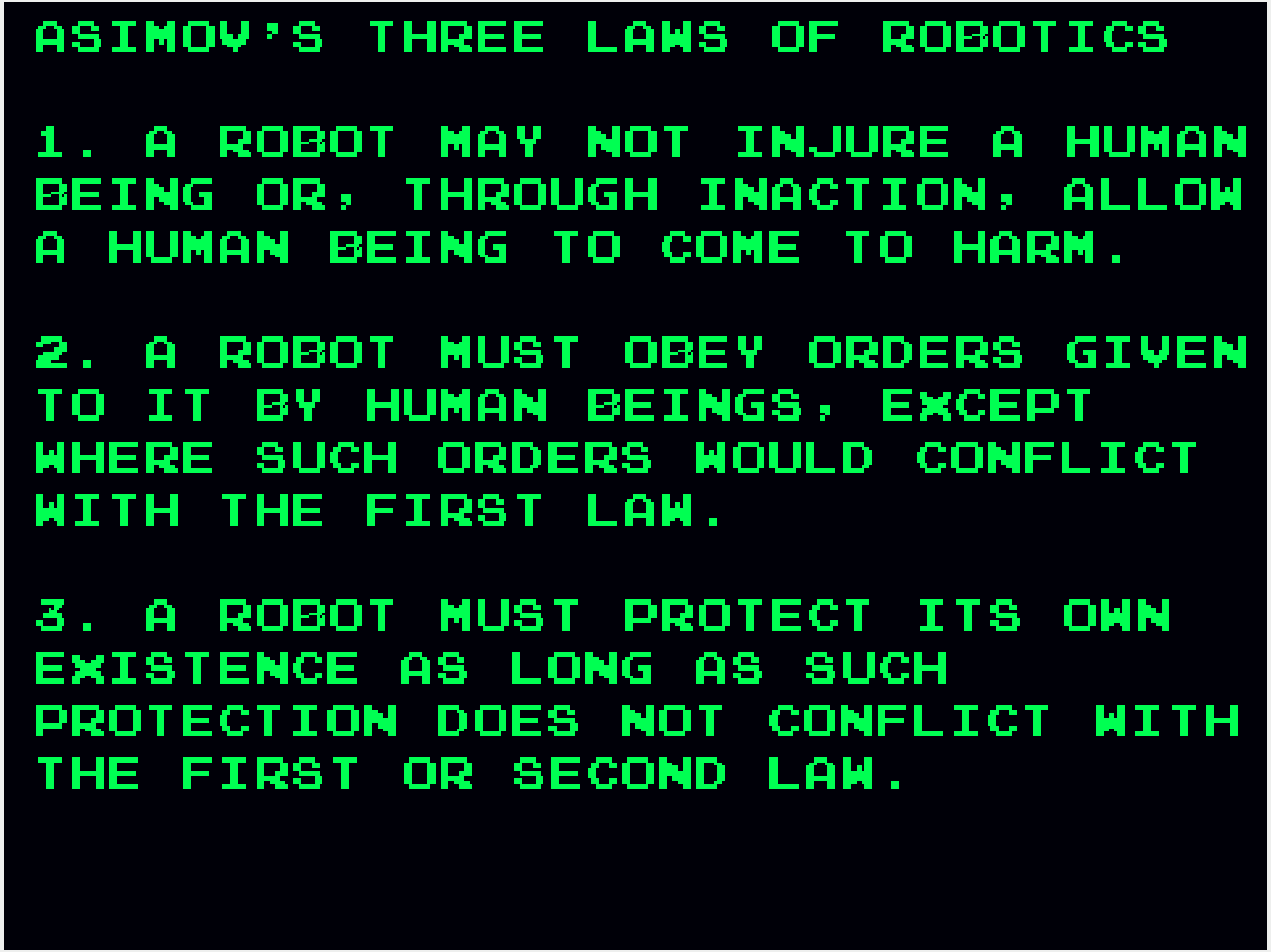
- A Popular Algorithm Is No Better at Predicting Crimes Than Random People

Source: The Atlantic, 2018, bit.ly/2OrULbA

Key ethical risks from the corporate viewpoint

- Bias and discrimination
- Erosion of Privacy
- Poor accountability
- Workforce displacement and transitions

Source: Schatsky, D., et al 2019. Can AI be ethical? Why enterprises shouldn't wait for AI regulation. Deloitte report



ASIMOV'S THREE LAWS OF ROBOTICS

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 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.

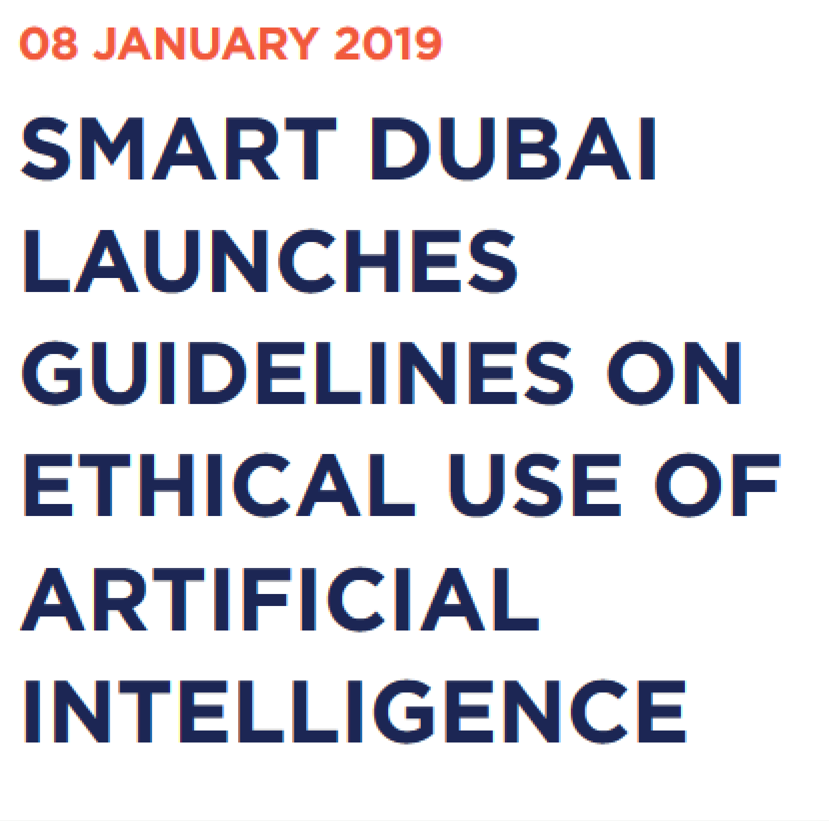
ASIMO TICS

1. A HUMAN
BEING ALLOW
A HUM RM.

2. A GIVEN
TO IT PT
WHERE FLICT
WITH

3. A OWN
EXIST
PROTE T WITH
THE F

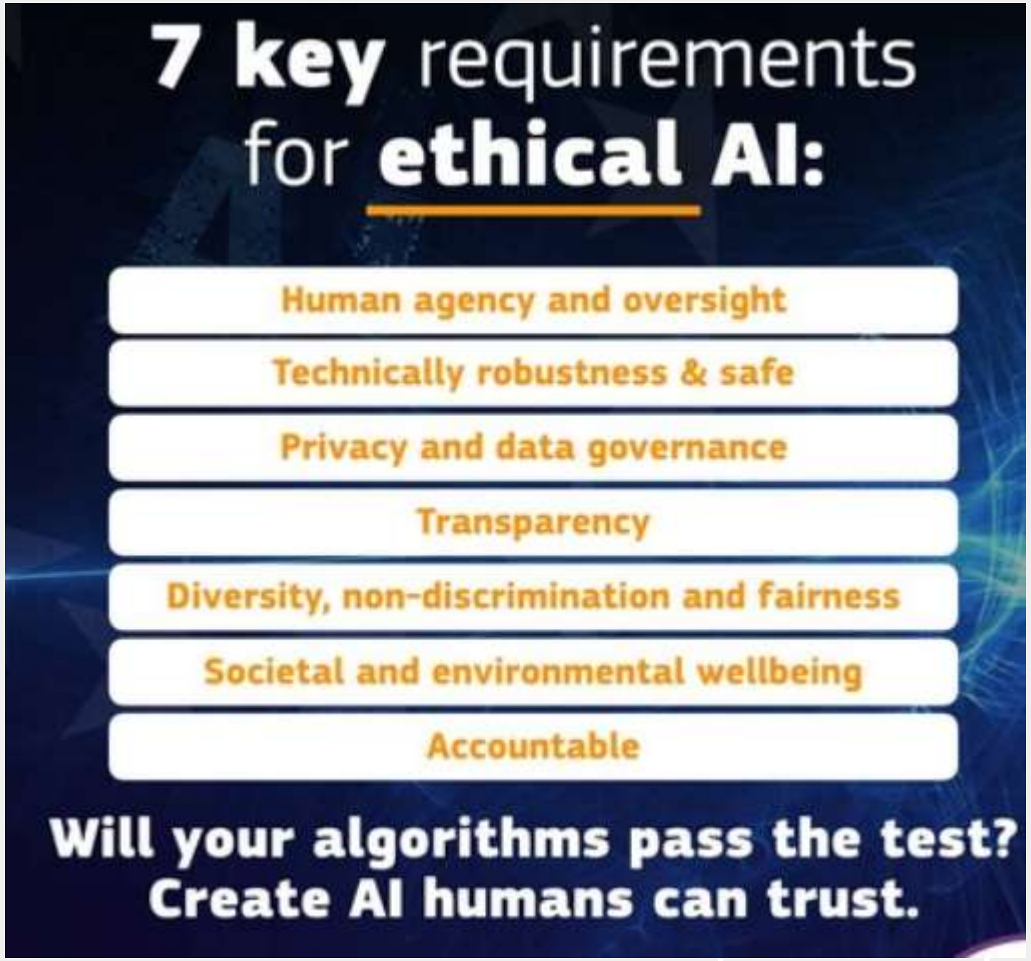
S1282
SIGNET
35¢
MAN-LIKE MACHINES RULE THE WORLD!
Fascinating Tales of a Strange Tomorrow
I, ROBOT
Isaac Asimov



08 JANUARY 2019

SMART DUBAI LAUNCHES GUIDELINES ON ETHICAL USE OF ARTIFICIAL INTELLIGENCE

ETHICS GUIDELINES FOR TRUSTWORTHY AI



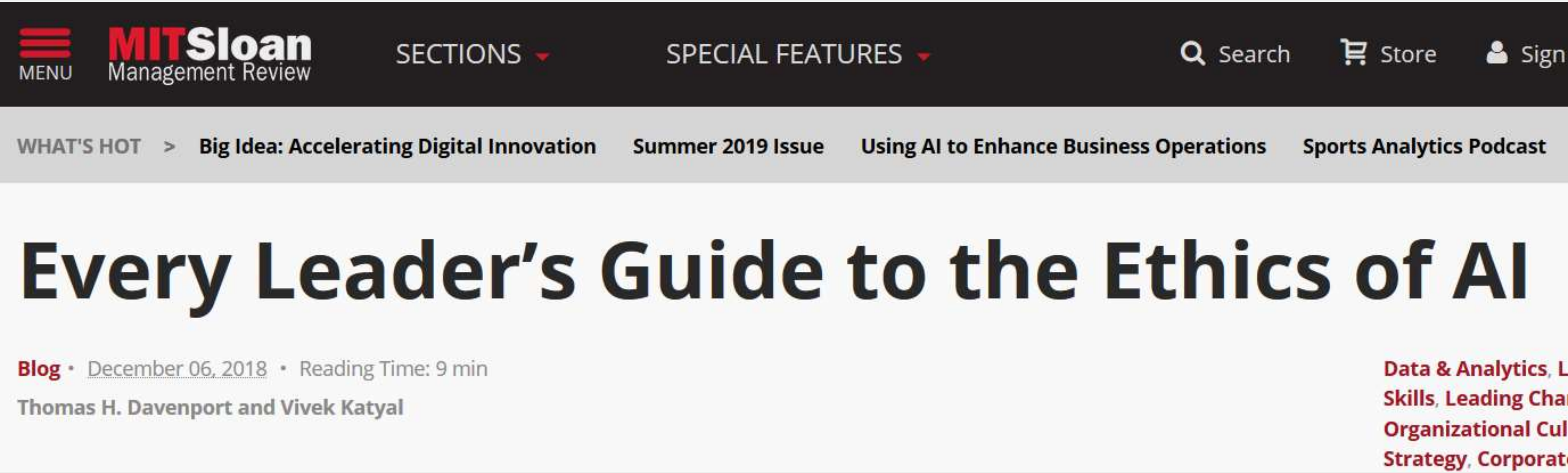
7 key requirements for ethical AI:

- Human agency and oversight
- Technically robustness & safe
- Privacy and data governance
- Transparency
- Diversity, non-discrimination and fairness
- Societal and environmental wellbeing
- Accountable

Will your algorithms pass the test?
Create AI humans can trust.

AI Ethics Guidelines Global Inventory

UPDATED: 23 MAY 2019


MENU
MIT Sloan Management Review
SECTIONS
SPECIAL FEATURES
Search
Store
Sign
WHAT'S HOT >
Big Idea: Accelerating Digital Innovation
Summer 2019 Issue
Using AI to Enhance Business Operations
Sports Analytics Podcast

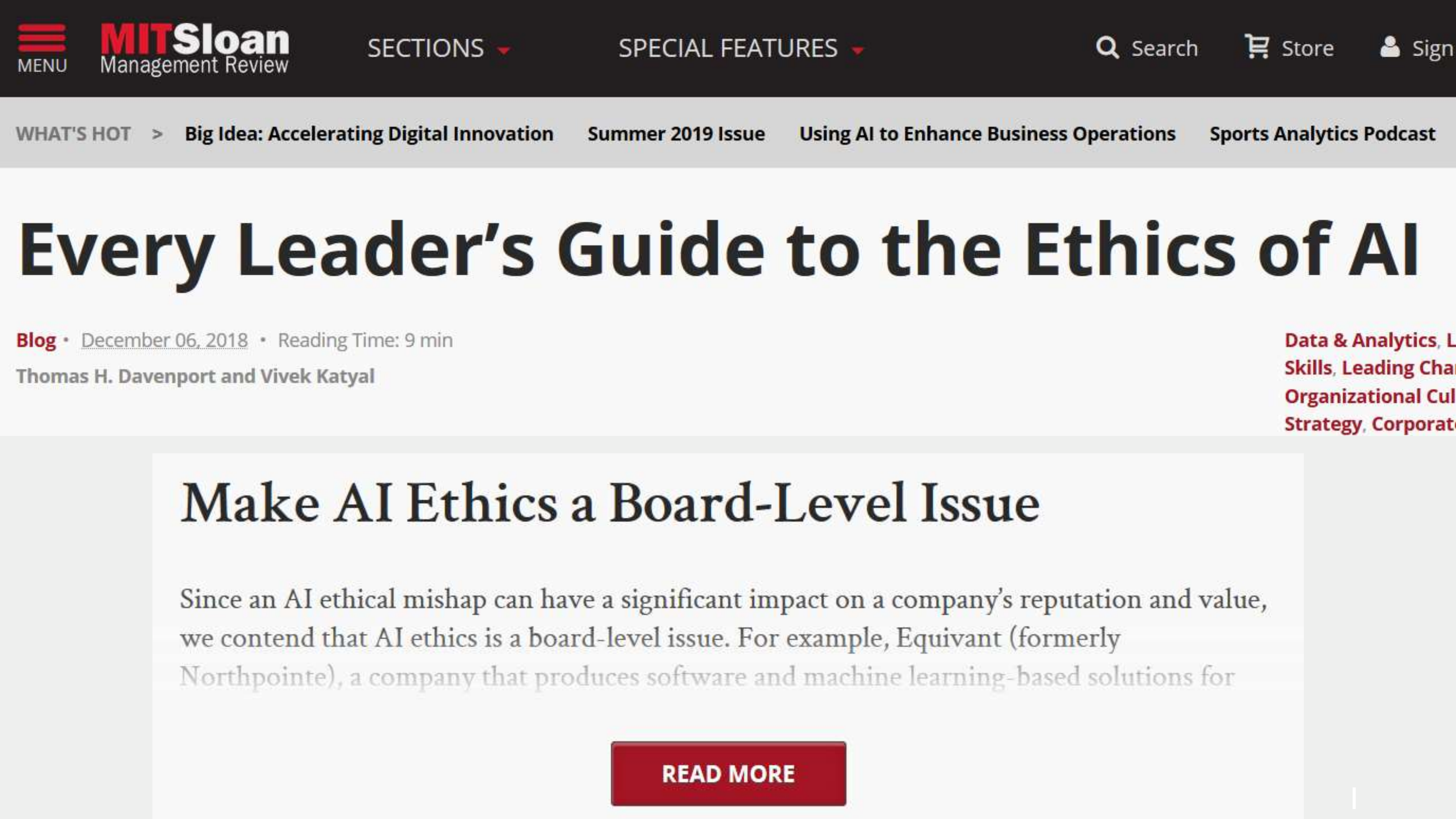
Every Leader's Guide to the Ethics of AI

Blog • December 06, 2018 • Reading Time: 9 min

Thomas H. Davenport and Vivek Katyal

Data & Analytics, L
Skills, Leading Cha
Organizational Cul
Strategy, Corporat

Source: https://sloanreview.mit.edu/article/every-leaders-guide-to-the-ethics-of-ai/



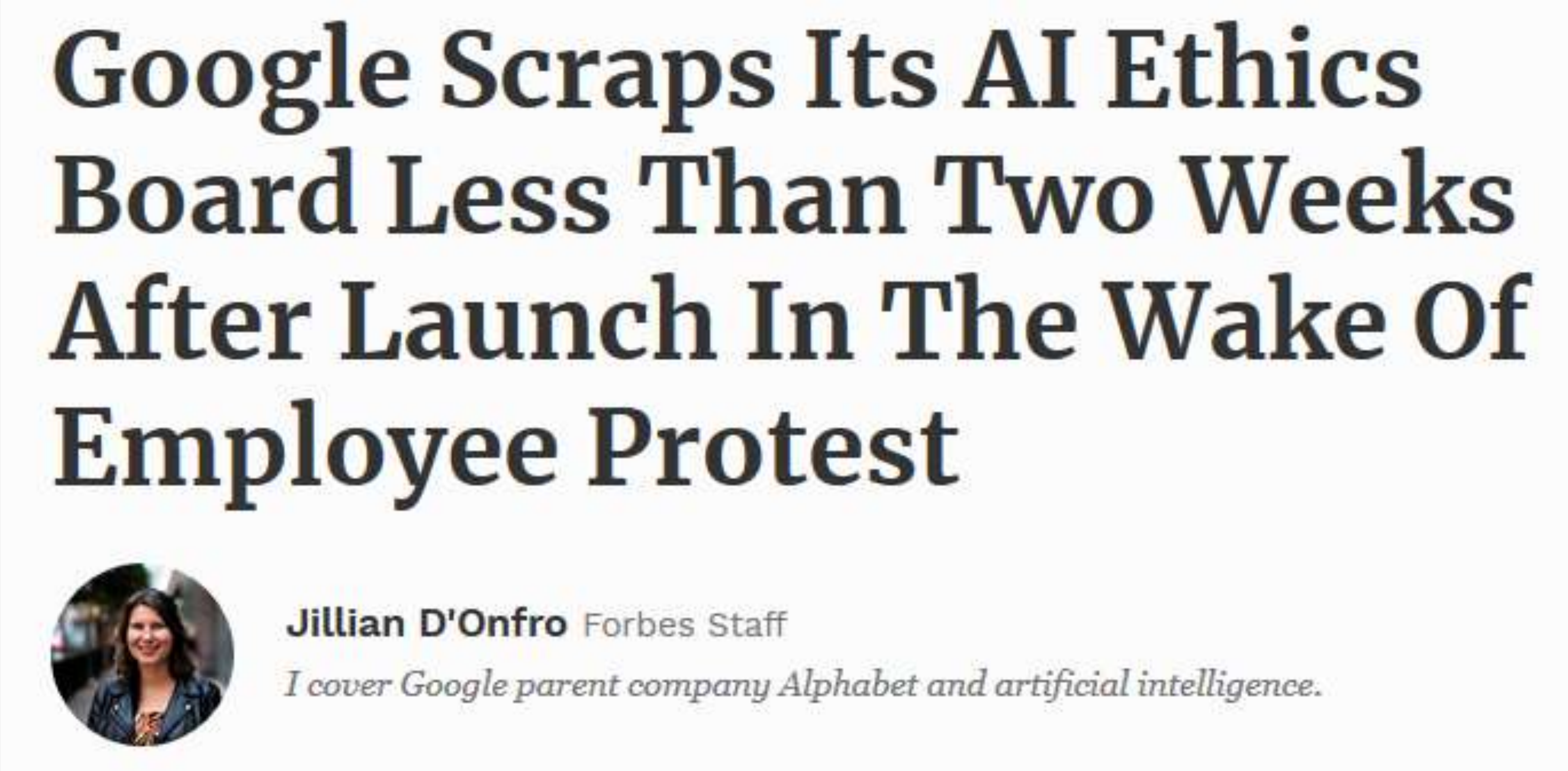
Every Leader's Guide to the Ethics of AI

Blog • December 06, 2018 • Reading Time: 9 min

Thomas H. Davenport and Vivek Katyal

Make AI Ethics a Board-Level Issue

Since an AI ethical mishap can have a significant impact on a company's reputation and value, we contend that AI ethics is a board-level issue. For example, Equivant (formerly Northpointe), a company that produces software and machine learning-based solutions for



Google Scraps Its AI Ethics Board Less Than Two Weeks After Launch In The Wake Of Employee Protest

Jillian D'Onfro Forbes Staff

I cover Google parent company Alphabet and artificial intelligence.

Source: https://www.forbes.com/sites/jilliandonfro/2019/04/04/google-cancels-its-ai-ethics
-board-less-than-two-weeks-after-launch-in-the-wake-of-employee-protest/#3e22a5b06e28

What should the leaders do?

- “Until regulations catch up, AI-oriented companies must establish their own ethical frameworks”

Source: Sloan Management Review, 2018, bit.ly/2NXF0Ky

FEATURE

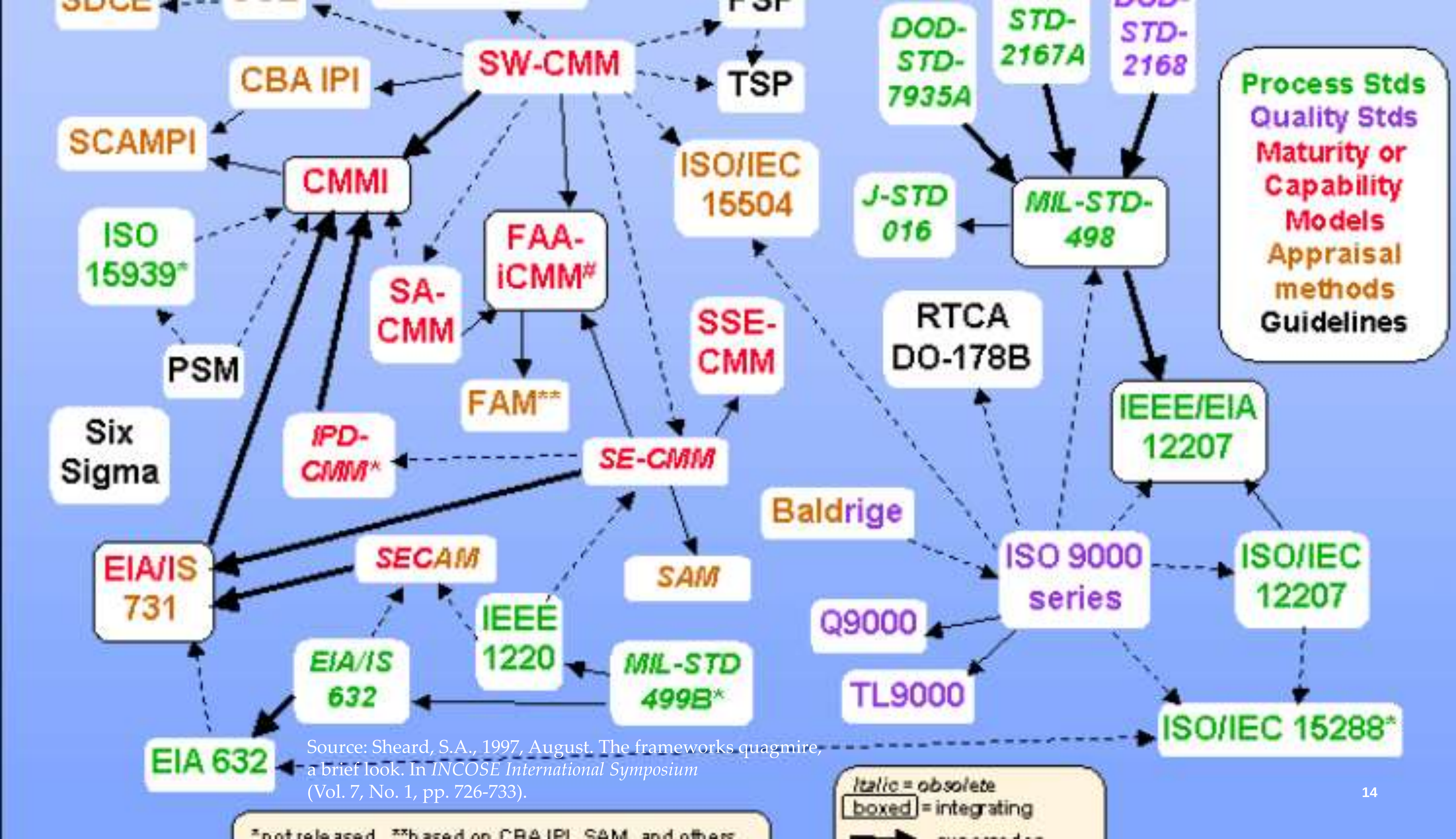
Can AI be ethical? Why enterprises shouldn’t wait for AI regulation

David Schatsky, Vivek Katyal, Satish Iyengar, and Rameeta Chauhan

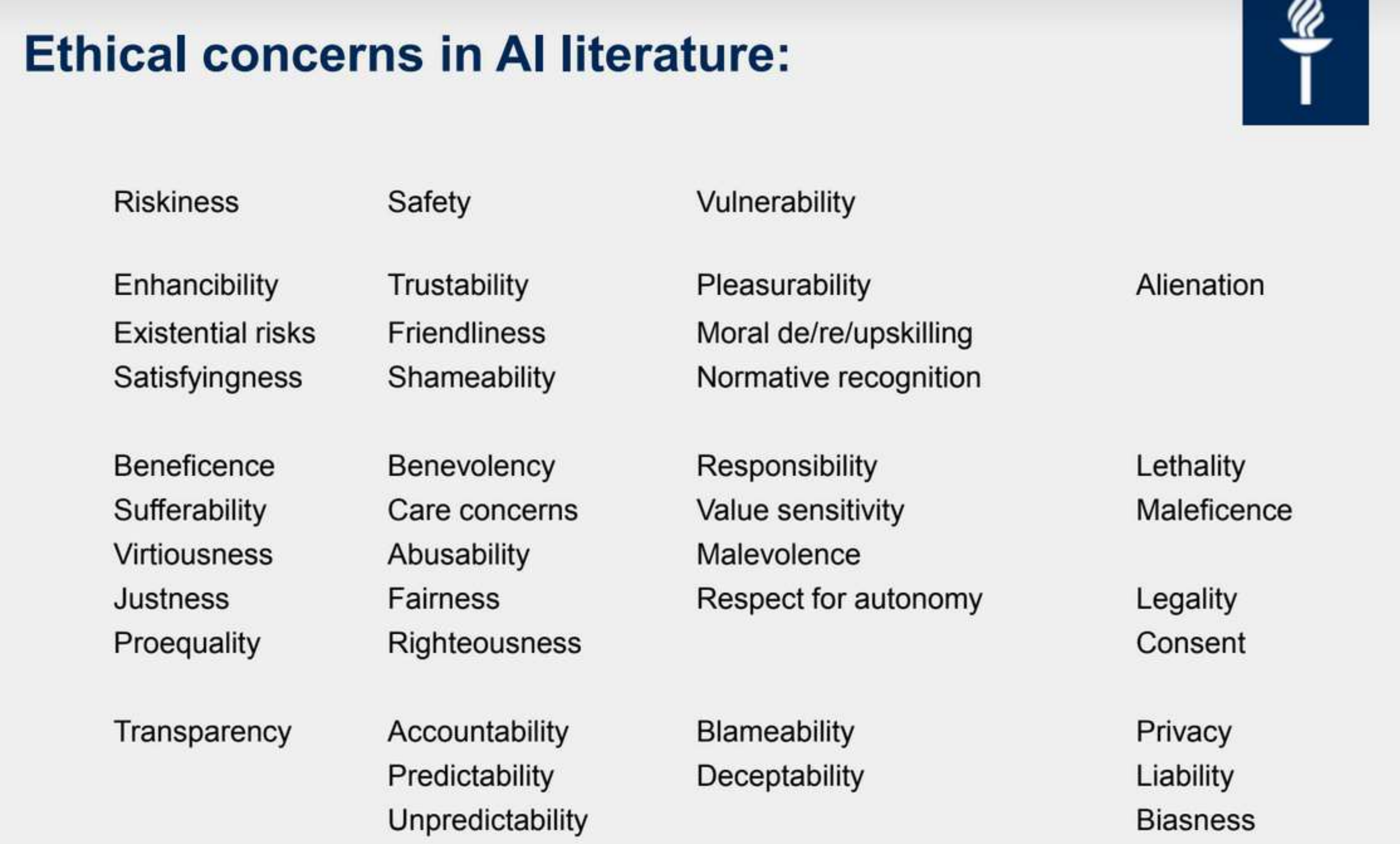
Source: Schatsky, D., et al 2019. Can AI be ethical? Why enterprises shouldn’t wait for AI regulation. Deloitte report

Our key research problem

How to empower developers to do Ethically Aligned Design in practice in software and systems development?



Source: Sheard, S.A., 1997, August. The frameworks quagmire, a brief look. In *INCOSE International Symposium* (Vol. 7, No. 1, pp. 726-733).

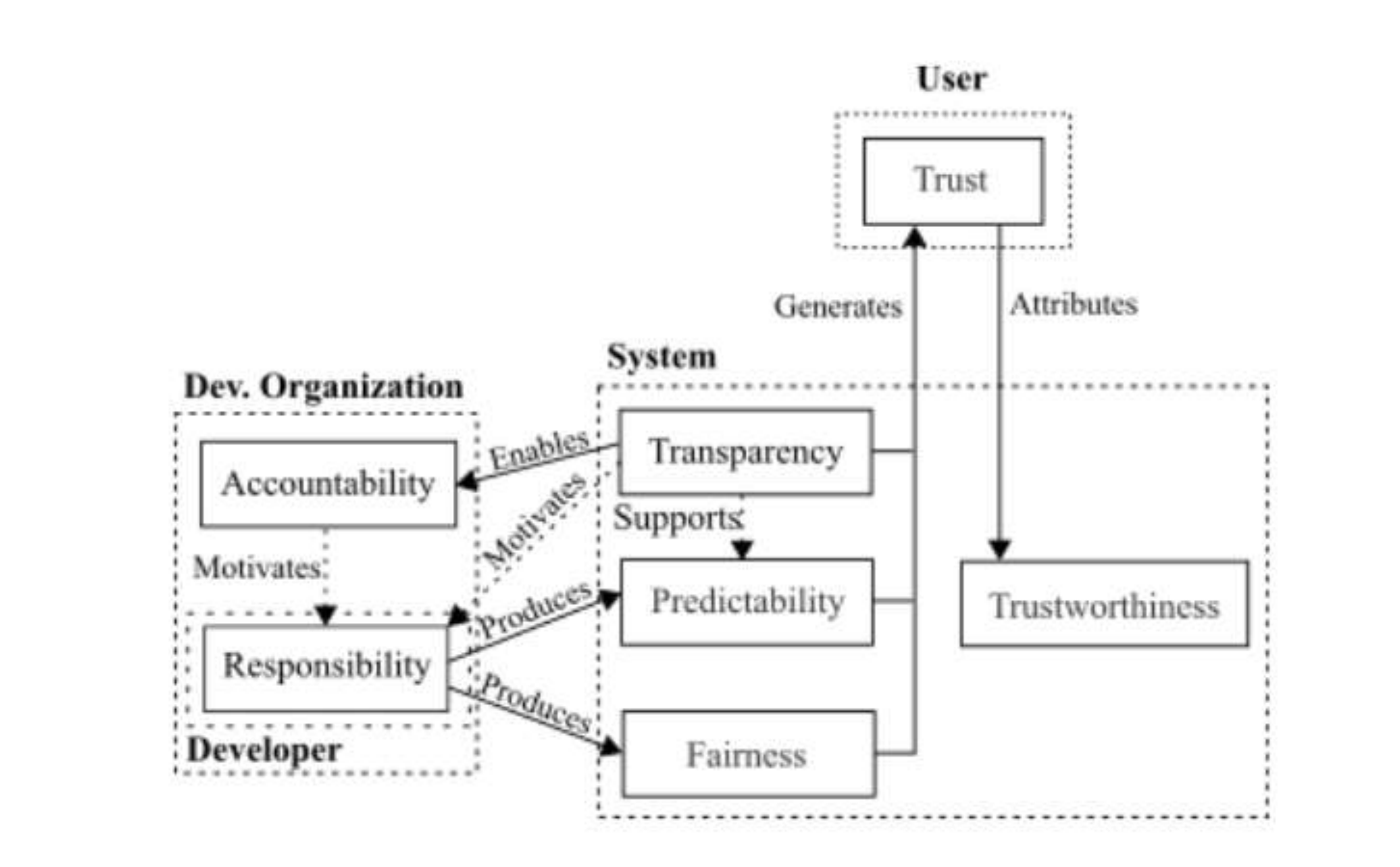


Ethical concerns in AI literature:

Riskiness	Safety	Vulnerability	
Enhancibility	Trustability	Pleasurability	Alienation
Existential risks	Friendliness	Moral de/re/upskilling	
Satisfyingness	Shameability	Normative recognition	
Beneficence	Benevolency	Responsibility	Lethality
Sufferability	Care concerns	Value sensitivity	Maleficence
Virtiousness	Abusability	Malevolence	
Justness	Fairness	Respect for autonomy	Legality
Proequality	Righteousness		Consent
Transparency	Accountability	Blameability	Privacy
	Predictability	Deceptability	Liability
	Unpredictability		Biasness

Concerns drawn from Vakkuri, V. and Abrahamsson, P., 2018, June. The Key Concepts of Ethics of Artificial Intelligence. In *2018 IEEE International Conference on Engineering, Technology and Innovation (ICE/ITMC)* (pp. 1-6). IEEE.
Author’s version available online at https://arxiv.org/abs/1809.07027

Conceptualization of the Relations Between Currently Discussed AI Ethics Construct

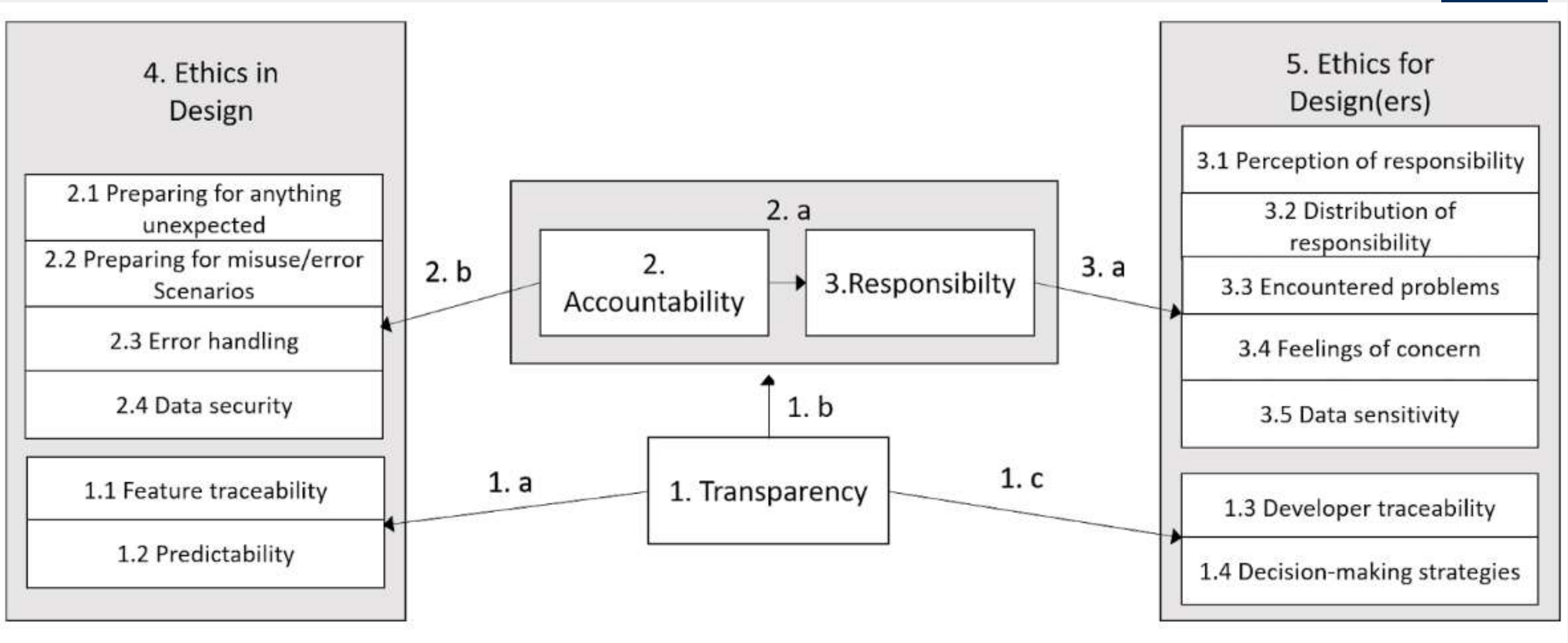


Vakkuri et al. 2020. Ethically Aligned Design of Autonomous Systems: Industry viewpoint and an empirical study. To appear in the proceedings of the Transport Research Arena (TRA 2020) *arXiv preprint https://arxiv.org/abs/1906.07946*.

AI Ethics discussion

- Long-standing area of study
  - Old scenarios slowly becoming reality
- Recent discourse has centered around a few central constructs
  - Transparency, Accountability
    - → Responsibility, Fairness… Trustworthiness
- Focus on high-level principles
  - Discussion largely conceptual
  - Little empirical data exists

Tangible Research Framework



Vakkuri, V., Kemell, K.K. and Abrahamsson, P., 2019.
AI Ethics in Industry: A Research Framework. To appear in the proceedings of the 3rd *Seminar on Technology Ethics, Turku, Finland, arXiv preprint arXiv:1910.12695*.

JYVÄSKYLÄN YLIOPISTO
UNIVERSITY OF JYVÄSKYLÄ

Empirical Study Results

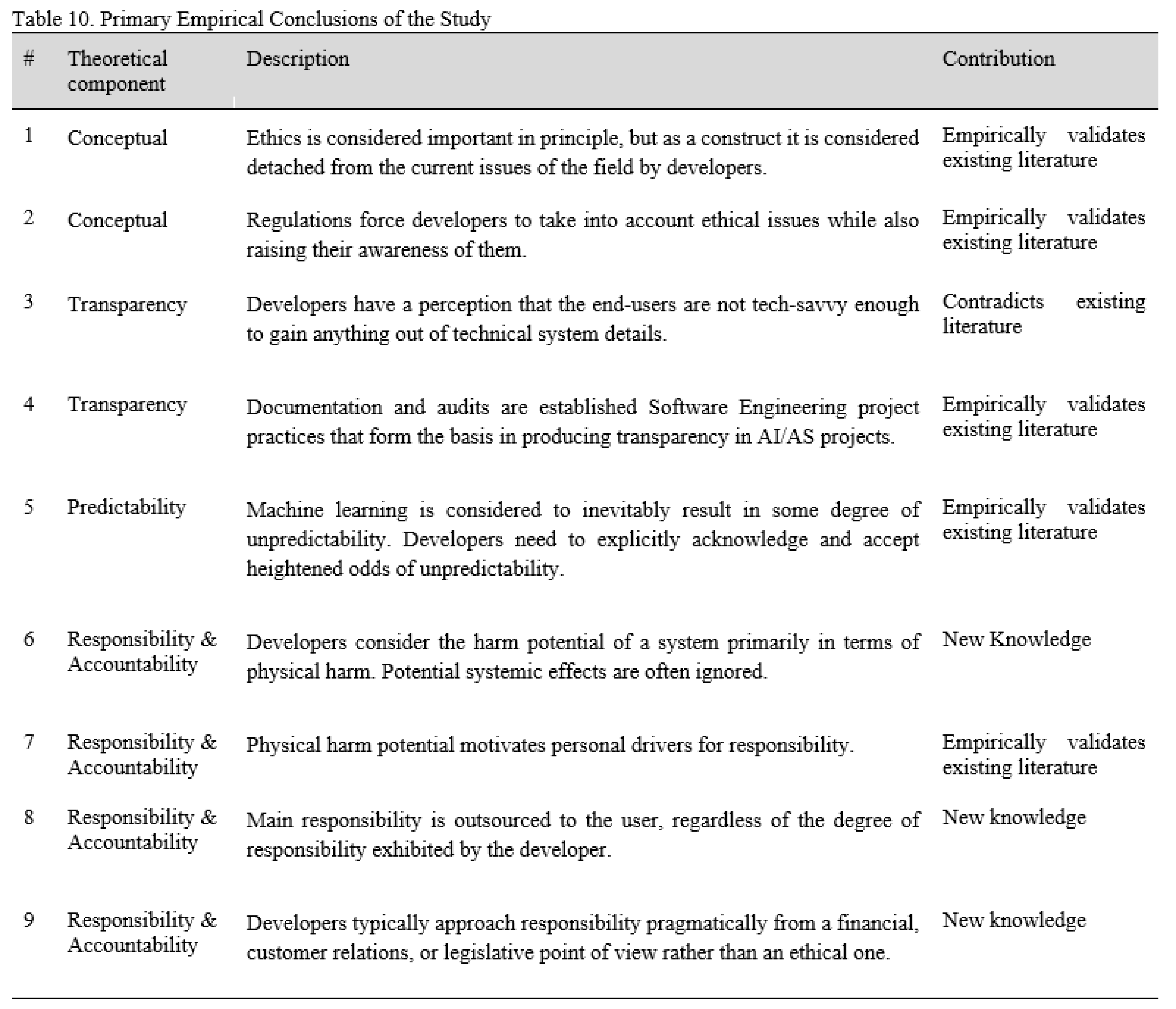


Table 10. Primary Empirical Conclusions of the Study

#	Theoretical component	Description	Contribution
1	Conceptual	Ethics is considered important in principle, but as a construct it is considered detached from the current issues of the field by developers.	Empirically validates existing literature
2	Conceptual	Regulations force developers to take into account ethical issues while also raising their awareness of them.	Empirically validates existing literature
3	Transparency	Developers have a perception that the end-users are not tech-savvy enough to gain anything out of technical system details.	Contradicts existing literature
4	Transparency	Documentation and audits are established Software Engineering project practices that form the basis in producing transparency in AI/AS projects.	Empirically validates existing literature
5	Predictability	Machine learning is considered to inevitably result in some degree of unpredictability. Developers need to explicitly acknowledge and accept heightened odds of unpredictability.	Empirically validates existing literature
6	Responsibility & Accountability	Developers consider the harm potential of a system primarily in terms of physical harm. Potential systemic effects are often ignored.	New Knowledge
7	Responsibility & Accountability	Physical harm potential motivates personal drivers for responsibility.	Empirically validates existing literature
8	Responsibility & Accountability	Main responsibility is outsourced to the user, regardless of the degree of responsibility exhibited by the developer.	New knowledge
9	Responsibility & Accountability	Developers typically approach responsibility pragmatically from a financial, customer relations, or legislative point of view rather than an ethical one.	New knowledge

Vakkuri et al. 2020. Ethically Aligned Design of Autonomous Systems: Industry viewpoint and an empirical study. To appear in the proceedings of the Transport Research Arena (TRA 2020) *arXiv https://arxiv.org/abs/1906.07946*.

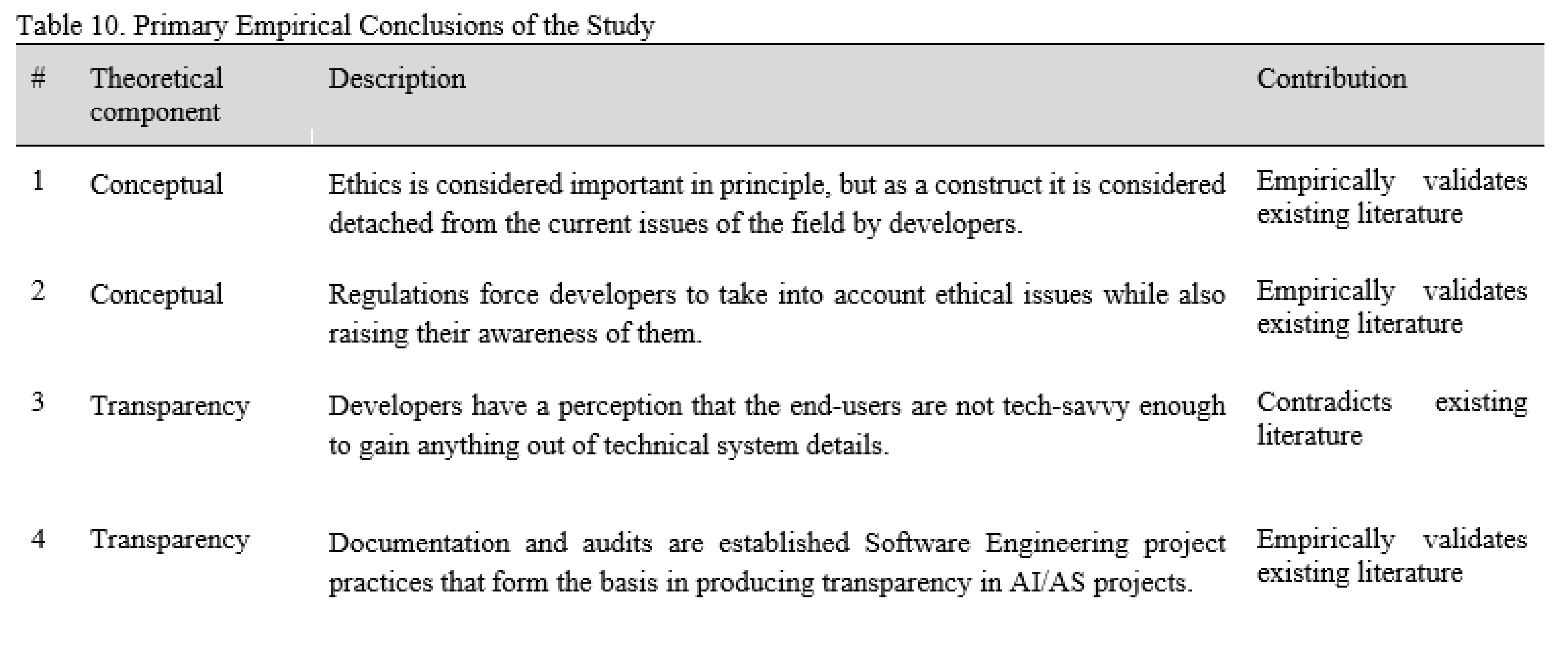
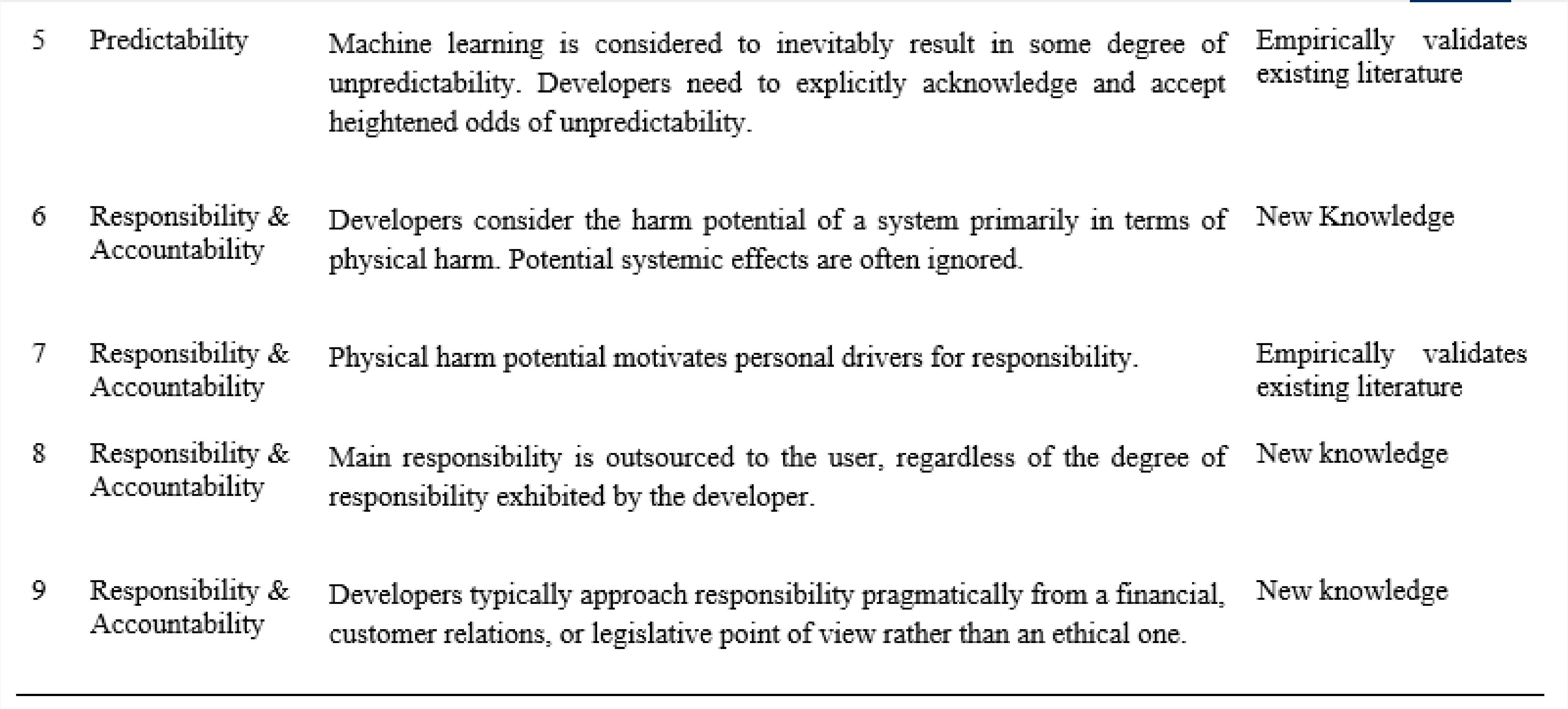


Table 10. Primary Empirical Conclusions of the Study

#	Theoretical component	Description	Contribution
1	Conceptual	Ethics is considered important in principle, but as a construct it is considered detached from the current issues of the field by developers.	Empirically validates existing literature
2	Conceptual	Regulations force developers to take into account ethical issues while also raising their awareness of them.	Empirically validates existing literature
3	Transparency	Developers have a perception that the end-users are not tech-savvy enough to gain anything out of technical system details.	Contradicts existing literature
4	Transparency	Documentation and audits are established Software Engineering project practices that form the basis in producing transparency in AI/AS projects.	Empirically validates existing literature

Vakkuri et al. 2020. Ethically Aligned Design of Autonomous Systems: Industry viewpoint and an empirical study. To appear in the proceedings of the Transport Research Arena (TRA 2020) *arXiv preprint https://arxiv.org/abs/1906.07946*.



5	Predictability	Machine learning is considered to inevitably result in some degree of unpredictability. Developers need to explicitly acknowledge and accept heightened odds of unpredictability.	Empirically validates existing literature
6	Responsibility & Accountability	Developers consider the harm potential of a system primarily in terms of physical harm. Potential systemic effects are often ignored.	New Knowledge
7	Responsibility & Accountability	Physical harm potential motivates personal drivers for responsibility.	Empirically validates existing literature
8	Responsibility & Accountability	Main responsibility is outsourced to the user, regardless of the degree of responsibility exhibited by the developer.	New knowledge
9	Responsibility & Accountability	Developers typically approach responsibility pragmatically from a financial, customer relations, or legislative point of view rather than an ethical one.	New knowledge

Vakkuri et al. 2020. Ethically Aligned Design of Autonomous Systems: Industry viewpoint and an empirical study. To appear in the proceedings of the Transport Research Arena (TRA 2020) *arXiv preprint https://arxiv.org/abs/1906.07946*.

Accountability

“It’s just a prototype”

“We have talked about the risks of decision-making support systems, but it doesn’t really affect what we do”

"It's really important how you handle any kind of data... that you preserve it correctly, among researchers, and don't hand it out to any government actors. [...] I personally can't see any way to harm anyone with the data we have though."

Source: Vakkuri V., Kemell KK., Abrahamsson P. (2019) Implementing Ethics in AI: Initial Results of an Industrial Multiple Case Study. In: Franch X., Männistö T., Martínez-Fernández S. (eds) Product-Focused Software Process Improvement. PROFES 2019. Lecture Notes in Computer Science, vol 11915. Springer, Cham, Author’s copy available at arxiv.org/abs/1906.12307

Responsibility

"Nobody wants to listen to ethics-related technical stuff. [...] It’s not relevant to the users”

“What could it affect... the distribution of funds in a region, or it could result in a school taking useless action... it does have its own risks, but no one is going to die because of it”

Source: Vakkuri V., Kemell KK., Abrahamsson P. (2019) Implementing Ethics in AI: Initial Results of an Industrial Multiple Case Study. In: Franch X., Männistö T., Martínez-Fernández S. (eds) Product-Focused Software Process Improvement. PROFES 2019. Lecture Notes in Computer Science, vol 11915. Springer, Cham, Author’s copy available at arxiv.org/abs/1906.12307

Responsibility

- The developers *could*, when asked, think of ways the system could negatively affect various stakeholders
- These were not addressed formally
- Developers did not have tools/methods to conduct formal ethical analyses in order to tackle such issues

Source: Vakkuri V., Kemell KK., Abrahamsson P. (2019) Implementing Ethics in AI: Initial Results of an Industrial Multiple Case Study. In: Franch X., Männistö T., Martínez-Fernández S. (eds) Product-Focused Software Process Improvement. PROFES 2019. Lecture Notes in Computer Science, vol 11915. Springer, Cham, Author’s copy available at arxiv.org/abs/1906.12307

Summary

- AI is spreading everywhere
- People voice out their concerns publicly
- AI Ethics is a growing concern for companies
- High level guidelines and principles exist but they fail to provide actionable advice to developers
- AI is developed by software engineers who need tangible methods to address the concerns
- Our studies provide empirical insight about the current state of practice
- First AI Ethics in Software Design methods are being published in 2020

Our AI Ethics research results are made public in Arxiv.org

The Key Concepts of Ethics of Artificial Intelligence	arxiv.org/abs/1809.07027
AI Ethics in Industry: A Research Framework	arxiv.org/abs/1910.12695
Implementing Ethics in AI: Initial results of an industrial multiple case study	arxiv.org/abs/1906.12307
Ethically Aligned Design of Autonomous Systems: Industry viewpoint	arxiv.org/abs/1906.07946
Ethically Aligned Design: An empirical evaluation of the RESOLVEDD-strategy	arxiv.org/abs/1905.06417

Thank you! Any Questions? Contact me at -> pekka.abrahamsson@jyu.fi