Ethical Software Engineering By Design

Alexander Pretschner

Software&Systems Engineering@TUM – bidt – fortiss – CDTM

jww Julian Nida-Rümelin, Severin Kacianka, Niina Zuber, Jan Gogoll

ECSS – October 25th, 2022

bidt Ein Institut der Bayerischen
Akademie der Wissenschaften

Introduction

Software engineering at universities: impact of research; education focused on technical problems

Diffusion of responsibility: from research to practice

Researchers → Applied Scientists → Product Teams → Marketing, Comms, etc. → Customers& Public

[Taken from: Hanna Wallach, keynote at NeurIPS 2020, https://nbiair.com/#Recordings]

Software engineering in practice: every day business is impact of engineering

EU regulation: (high risk) AI applications

Teaching, development, certification

Ethics

Values and Impact: Deontological and consequentialist ethics

Reproach to ethicists: „Useless!" (and to software engineers: „Not informatics anymore!")

Indeed: >120 Codes of Conduct for AI/Software/Systems Engineering rather fruitless

Reason: Software context-specific; hence values and trade-offs context-specific
       Contexts: application domain, technology, users' culture, developers' culture, optimization goals, …

Examples: Corona app, face recognition, data integration, care robots, resume analyzers, etc. – but also software without AI/data: camera surveillance w/o FR, BitTorrent, Telegram, Bitcoin, website preferences, ...

Genericity of CoCs hence necessary. Only way out: schema that caters to context specificity.

"Rather of Deontological Usefulness": what about trade-offs?

https://link.springer.com/article/10.1007/s13347-021-00451-w

https://standards.ieee.org/industry-connections/ec/ead-v1.html
https://www.acm.org/code-of-ethics
https://gi.de/ueber-uns/organisation/unsere-ethischen-leitlinien/
http://www.ethics.org.au/on-ethics/blog/november-2018/with-great-power-comes-great-responsibility-%E2%80%93-but

Gogoll, Zuber, Kacianka, Greger, Pretschner, Nida-Rümelin:
Ethics in the Software Development Process: from Codes of Conduct to Ethical Deliberation. *Philos. Technol.* 34: 1085–1108, 2021

CoC necessarily generic - McNamara et al. (2018) find no evidence that CoCs influence behaviour

Development: Ethical Deliberation in Agile Processes

https://www.bidt.digital/wp-content/uploads/2021/04/Digital-Transformation-and-Ethics_Zuber-et-al_EN.pdf

No simple way out. Need to address concerns in a context-specific manner: think!
Can be done in a systematic way

Development driven by EDAP: "Ethical Deliberation in Agile Processes"

Key idea: start with values; continuously reflect on **mechanisms, not yes/no**

Characteristics of agility blend particularly well:
planning; incrementality; empowerment; learning

https://www.nature.com/articles/s41599-022-01206-4

Later today: How to teach?

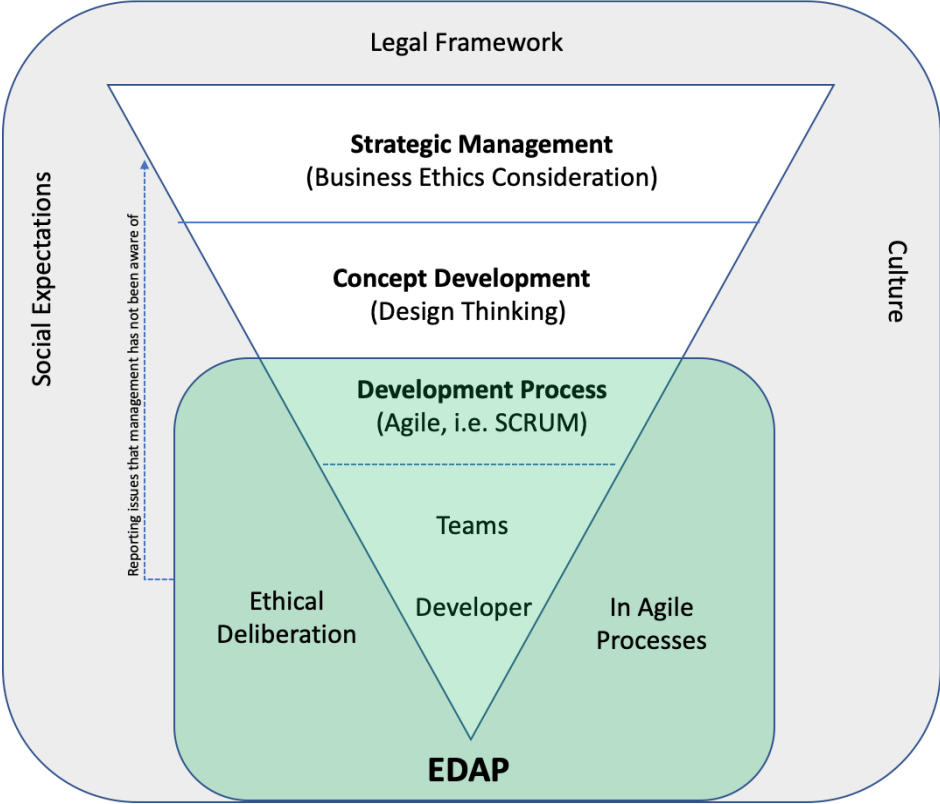bidt Ein Institut der Bayerischen
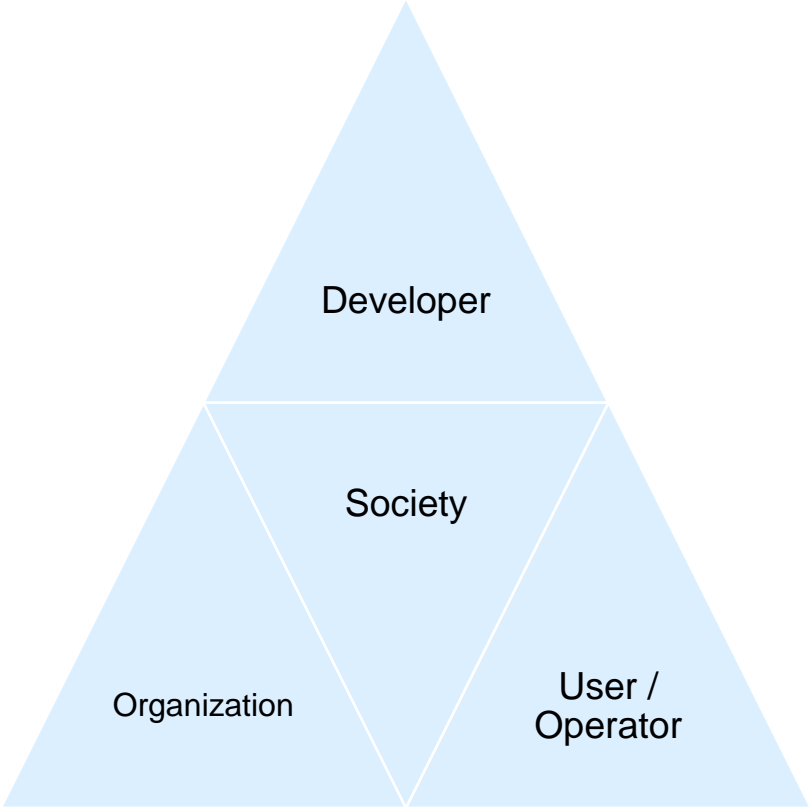Akademie der Wissenschaften

Software Engineering and AI

Ethical issues are not confined to AI – but this is suggested by the current debate!
A centralized Corona app? Palantir Foundry? Integration of registers?
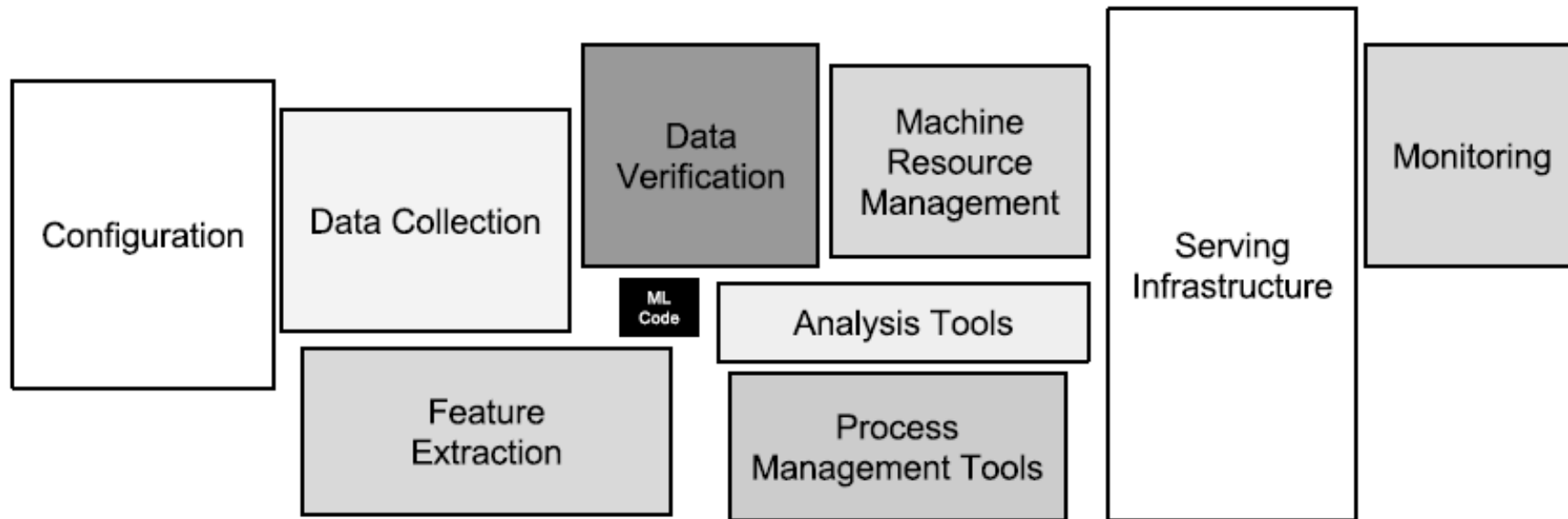
Sculley et al.: Hidden Technical Debt in ML systems, Proc. NIPS 2015: 2503–2511

Who is responsible?

Teaching: Later Today

My conclusion: Classes not sufficient. Hypotheticals fun for everybody. Learning by doing using the EDAP schema. Card games. Who teaches best?

General ethics classes; or ethics in CS: Harvard EthiCS project
[Gross et al., Embedded EthiCS: Integrating Ethics Across CS Education, CACM 62(8), 2019]

General overviews available, e.g.
[Fiesler et al., What Do We Teach When We Teach Tech Ethics? A Syllabi Analysis, Proc. SIGCSE, 2020]
[Mulhearn et al., Review of Instructional Approaches in Ethics Education, Science and Engineering Ethics volume 23: 883–912, 2017]

Impact?
- „Less meat after an ethics class"
  [Schwitzgebel et al., Do ethics classes influence student behavior? Case study: Teaching the ethics of eating meat, Cognition 203 (2020)]
- Students „liked it" [Gross et al., Embedded EthiCS: Integrating Ethics Across CS Education, CACM 62(8), 2019]
- Overall, moderate effects (but "depends")
  [Mulhearn et al., Review of Instructional Approaches in Ethics Education, Science and Engineering Ethics volume 23: 883–912, 2017]

Implemented in forefront industry: E.g., Palantir

bidt Ein Institut der Bayerischen Akademie der Wissenschaften

Certification

Product Individual Process Organization

bidt Bayerisches Forschungsinstitut
für Digitale Transformation
EIN INSTITUT DER BAYERISCHEN AKADEMIE DER WISSENSCHAFTEN

Final Comments

Ethics not reproducible to compliance: default is people seen as risk! Empowerment?

Power asymmetry: programmers can quit

Considerations apply to both UI and program logic

DevOps, specifically MLOps: ethical consideration doesn't stop at a specific moment in time. For AI, context continuously changes (!)

Others tell us that the ideas generalize beyond Agile

bidt Ein Institut der Bayerischen Akademie der Wissenschaften

Wrap-Up

Along with others, software engineers individually responsible for the technology they build

Ethics not only an AI concern. Software is the concern.

Education: hypotheticals; practical courses; practical courses with EDAP

Certification: individuals, organizations, and systems (or development processes for a system)

Engineering: Structured deliberation with EDAP – agility and ethics seem a perfect match.
Pick *your* values. Ethics not reduced to yes or no: choose a mechanism.

Let's not overdo it though: what is the first-order problem here?