



Applied Machine Learning Days 2022

Tuesday March 29, 2022 / 10:45 AM - 11:00 AM

“*AI Ethics Standards and Guidelines*”



Prof. Dr.
Dagmar
Monett



**Prof. Dr.
Dagmar Monett**



/monettdiaz



@dmonett



Hello!



Hochschule für
Wirtschaft und Recht Berlin
Berlin School of Economics and Law

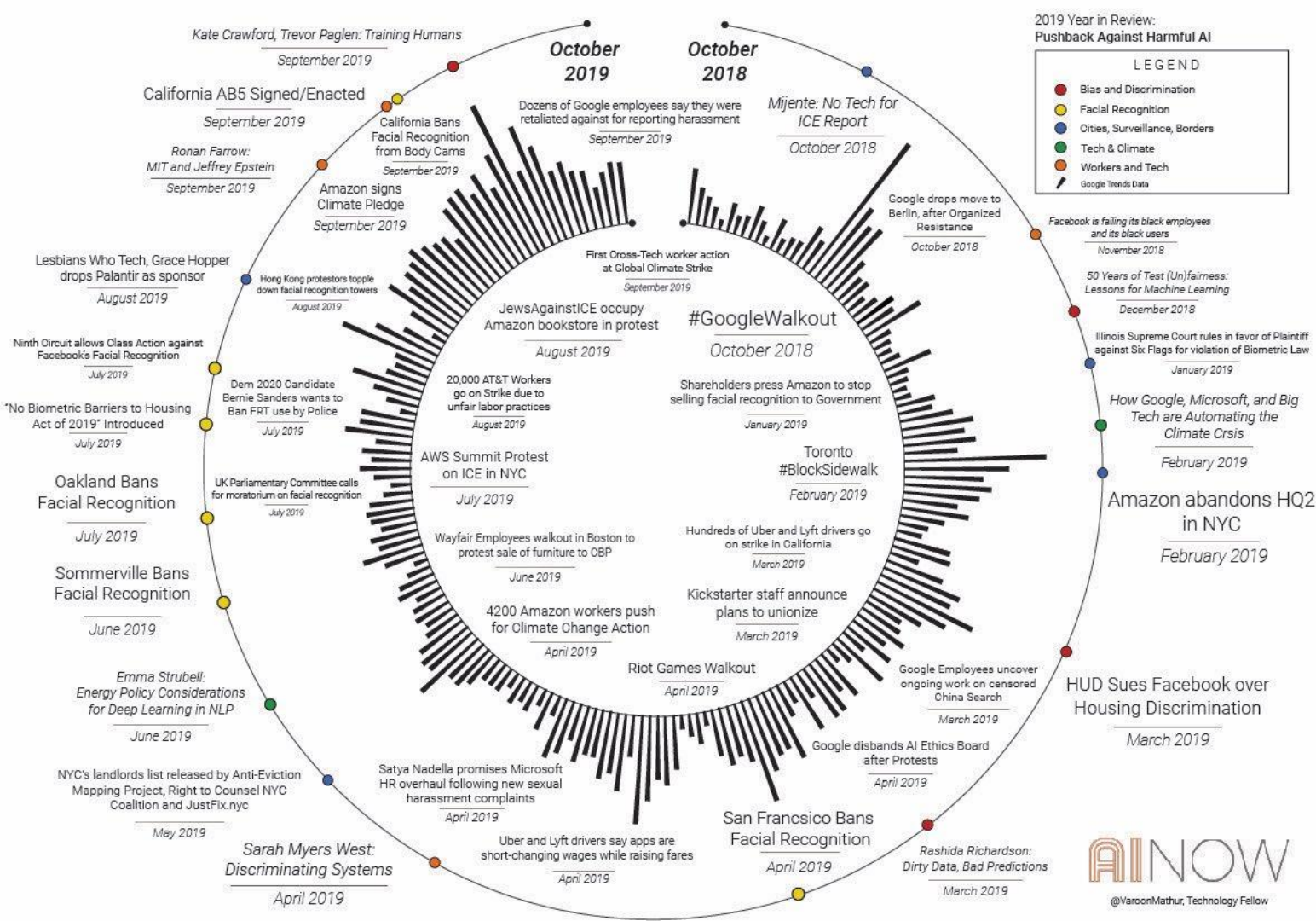
- ▶ Prof. Dr. Computer Science
Artificial Intelligence, Software Engineering
- ▶ Head of the Computer Science Dept. (Apr. 2022–)
Faculty of Cooperative Studies
Berlin School of Economics and Law (HWR Berlin)
- ▶ Co-Director M.Sc. Digital Transformation (since Dec. 2020)
Berlin Professional School
Berlin School of Economics and Law (HWR Berlin)



<http://monettdiaz.com>



dagmar.monett-diaz@hwr-berlin.de



Pushback Against Harmful AI

Courtesy of
@VaroonMathur,
Tech fellow
@AINowInstitute

“Ethical judgment corresponds with an ethical system.”

If the system is not ethical, we cannot have ethical outcomes.

It's actually much more than
only ethics...

**It is also standards, policies,
principles, law, values,
guidelines, recommendations.**

The
Alan Turing
Institute

Understanding artificial intelligence ethics and safety

A guide for the responsible
design and implementation of AI
systems in the public sector

Dr David Leslie
Public Policy Programme



(Leslie, 2019)

*“**AI ethics** is a set of values, principles, and techniques that employ widely accepted standards of right and wrong to guide moral conduct in the development and use of AI technologies.”*

Leslie, D. (2019). Understanding artificial intelligence ethics and safety: A guide for the responsible design and implementation of AI systems in the public sector. *The Alan Turing Institute*.

<https://doi.org/10.5281/zenodo.3240529>

Ethical Platform for the Responsible Delivery of an AI Project

SUM Values

that support, underwrite, and motivate
a responsible innovation ecosystem



Respect, Connect,
Care, Protect

Objectives: to provide an accessible framework for consideration of the moral scope of the social and ethical impacts of your project and to establish well-defined criteria to evaluate its ethical permissibility.

FAST Track Principles

that facilitate an actionable orientation to
the ethical design and use of AI systems



Fairness, Accountability,
Sustainability, Transparency

Objectives: to make sure that your project is bias-mitigating, non-discriminatory, and fair, and to safeguard public trust in your project's capacity to deliver safe and reliable AI innovation.

PBG Framework

that operationalises the values and principles
in an end-to-end workflow governance model



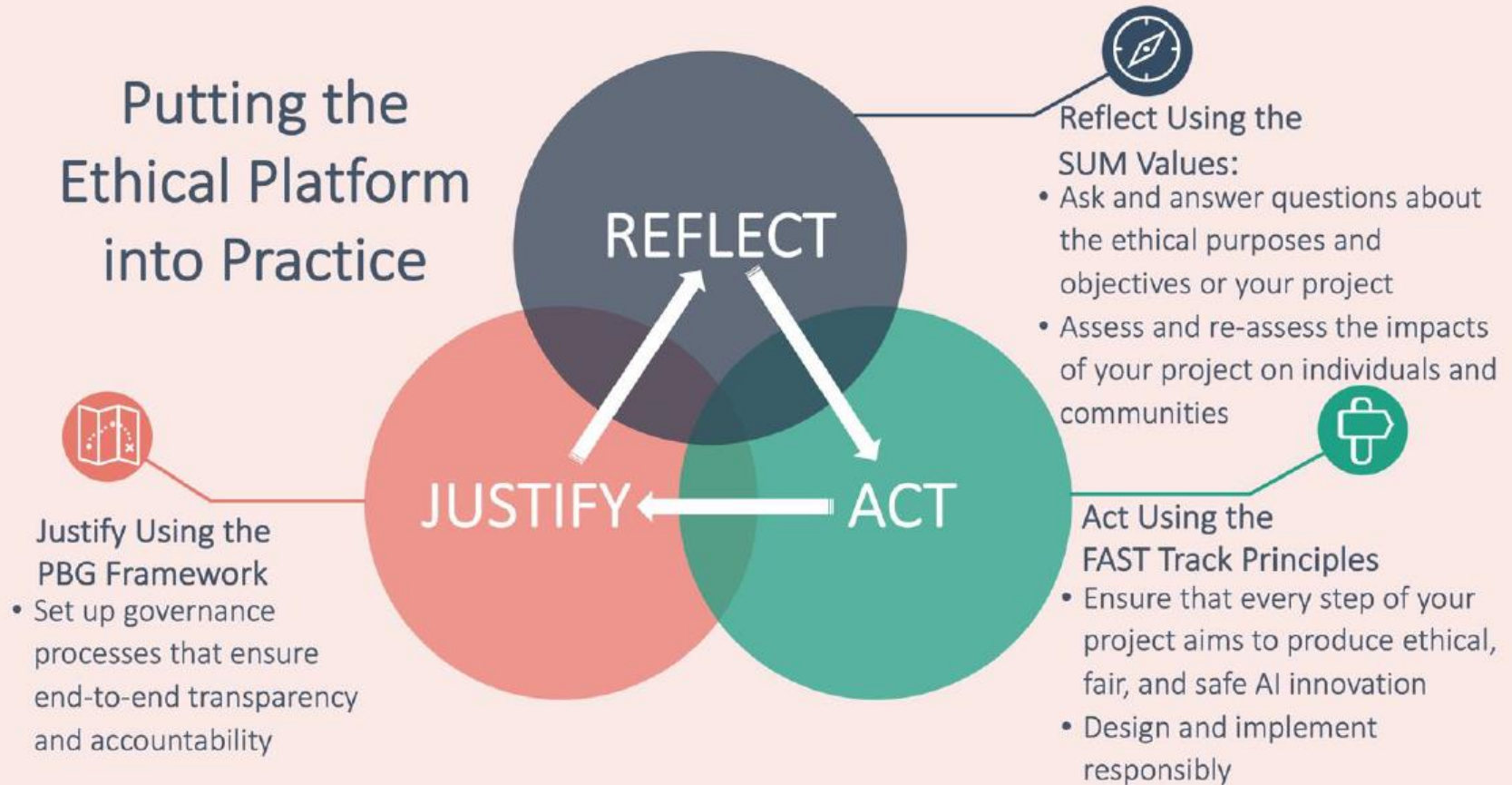
Process-Based
Governance Framework

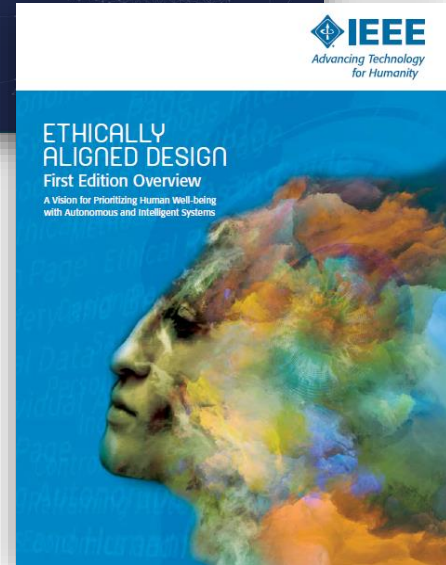
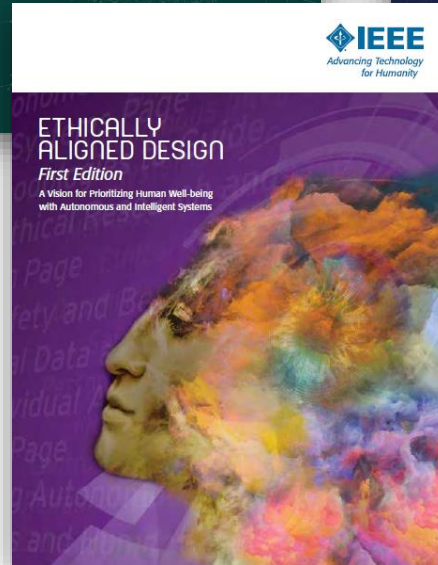
Objective: to set up transparent processes of design and implementation that safeguard the justifiability of both your AI project and its product as well as enable end-to-end accountability.

*“A framework of ethical values that **S**upport, **U**nderwrite, and **M**otivate
a responsible data design and use ecosystem.”*

(Leslie, 2022)

Putting the Ethical Platform into Practice







THE UNIVERSAL DECLARATION OF Human Rights

Article 1 Recognition of the inherent dignity and of the equal and inalienable rights of all members of the human family is the foundation of freedom, justice and peace in the world.

Article 2 Everyone has inherent rights and freedoms without distinction of race, sex, language, religion, political opinion, social status or birth.

Article 3 Everyone has the right to life, liberty and security of person.

Article 4 No one shall be subjected to torture or to cruel, inhuman or degrading treatment or punishment.

Article 5 No one shall be subjected to arbitrary arrest, detention or exile.

Article 6 Everyone has the right to recognition as a person before the law.

Article 7 Everyone has the right to equal treatment before the law.

Article 8 Everyone has the right to an effective remedy by the competent national authorities for the violations of his fundamental rights and freedoms recognized by the Constitution.

Article 9 No one shall be subjected to arbitrary arrest, detention or exile.

Article 10 Everyone has the right to a fair and public hearing by the competent national authorities for the determination of his rights and obligations.

Article 11 Everyone has the right to a fair and public hearing by the competent national authorities for the determination of his rights and obligations.

Article 12 No one shall be subjected to arbitrary interference with his privacy, family, home or correspondence, nor to attacks upon his honour and reputation.

Article 13 Everyone has the right to free movement and residence within the frontiers of each State.

Article 14 Everyone has the right to seek and to enjoy in other countries asylum from persecution.

Article 15 Everyone has the right to nationality.

Article 16 Everyone has the right to marry and to found a family.

Article 17 Everyone has the right to own property.

Article 18 Everyone has the right to freedom of thought, conscience and religion.

Article 19 Everyone has the right to freedom of opinion and expression.

Article 20 Everyone has the right to freedom of peaceful assembly and association.

Article 21 Everyone has the right to take part in the government of his country.

Article 22 Everyone has the right to social security.

Article 23 Everyone has the right to work.

Article 24 Everyone has the right to rest and leisure.

Article 25 Everyone has the right to a standard of living adequate for the health and well-being of himself and of his family.

Article 26 Everyone has the right to education.

Article 27 Everyone has the right to participate in the cultural life of the community.

Article 28 Everyone has the right to a social and international order in which his rights and freedoms can be fully realized.

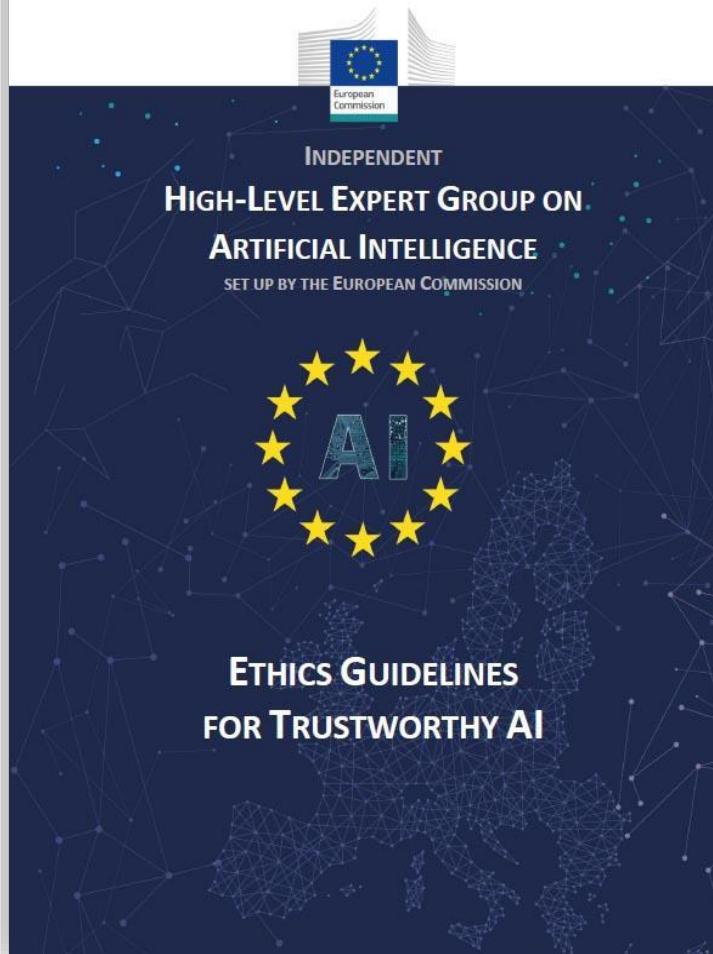
Article 29 Everyone has the right to the free development of his personality.

Article 30 Nothing shall be interpreted as authorizing any State, group or individual to engage in any activity aimed at destroying the rights and freedoms set forth in this Declaration.

“Everyone is entitled to all the rights and freedoms set forth in this Declaration, **without distinction of any kind**, such as race, colour, sex, language, religion, political or other opinion, national or social origin, property, birth or other status.”
(Article 2)

“No one shall be subjected to **arbitrary arrest, detention or exile**.”
(Article 9)

UNITED NATIONS

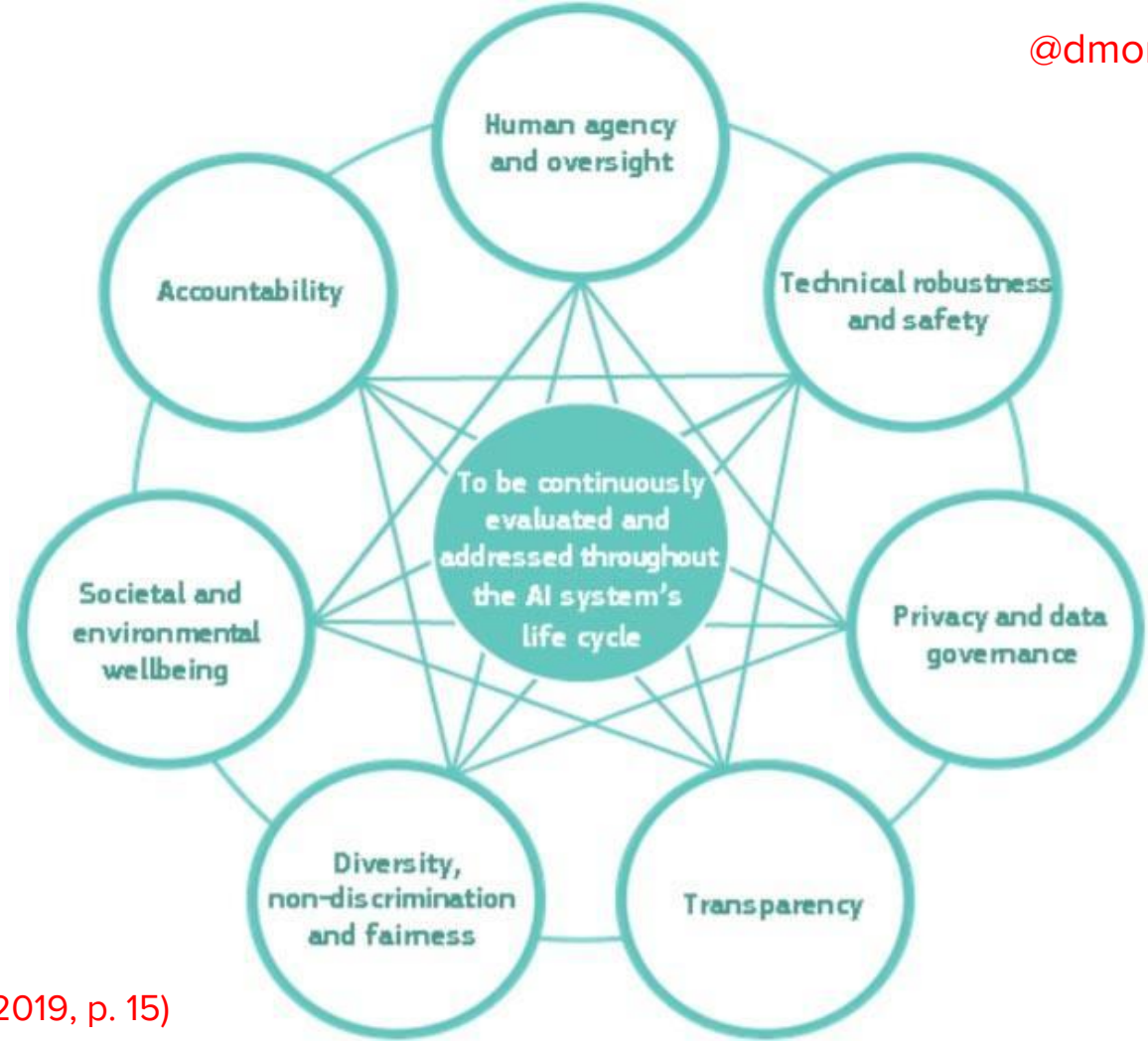


(EU, 2019a)

*“Ensure that the development, deployment and use of AI systems meets the **seven key requirements for Trustworthy AI**:*
(1) human agency and oversight,
(2) technical robustness and safety, (3) privacy and data governance, (4) transparency, (5) diversity, non-discrimination and fairness, (6) environmental and societal well-being and (7) accountability.”

<https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>

7 requirements for responsible AI



(Ethics Guidelines for Trustworthy AI, 2019, p. 15)

*“**Trustworthy AI** has three components, which should be met throughout the system's entire life cycle:*

- 1. it should be **lawful**, complying with all applicable laws and regulations;*
- 2. it should be **ethical**, ensuring adherence to ethical principles and values; and*
- 3. it should be **robust**, both from a technical and social perspective, since, even with good intentions, AI systems can cause unintentional harm.”*

The realization of Trustworthy AI

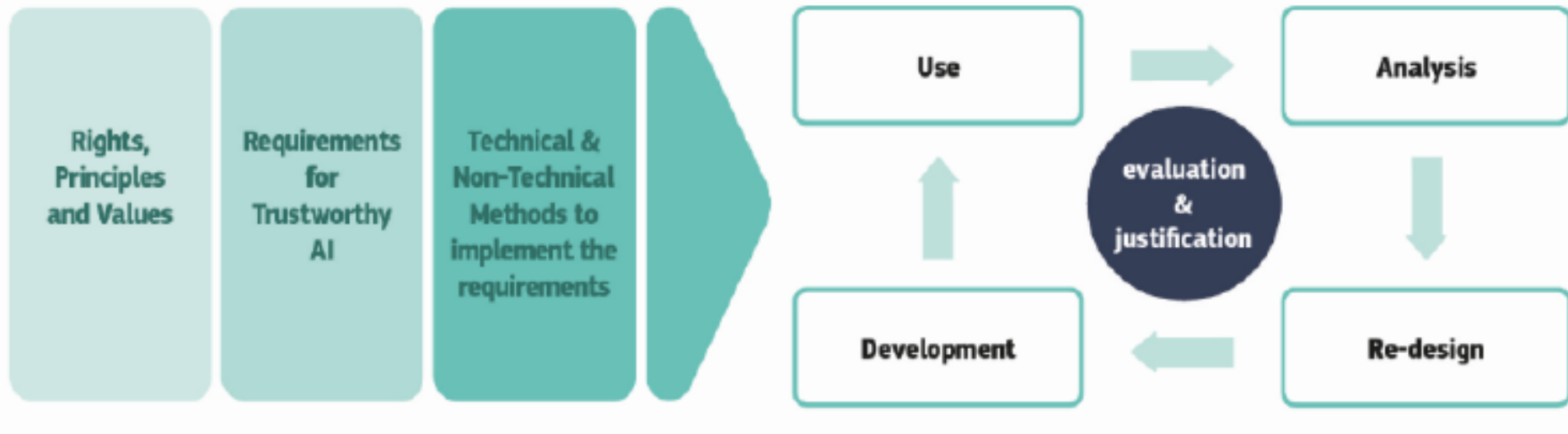


Figure 3: Realising Trustworthy AI throughout the system's entire life cycle

“An [AI system] cannot genuinely be deemed ‘ethical’ without accounting for the business it is involved in.”

“There can be no ethical AI without ethical businesses to build it.”

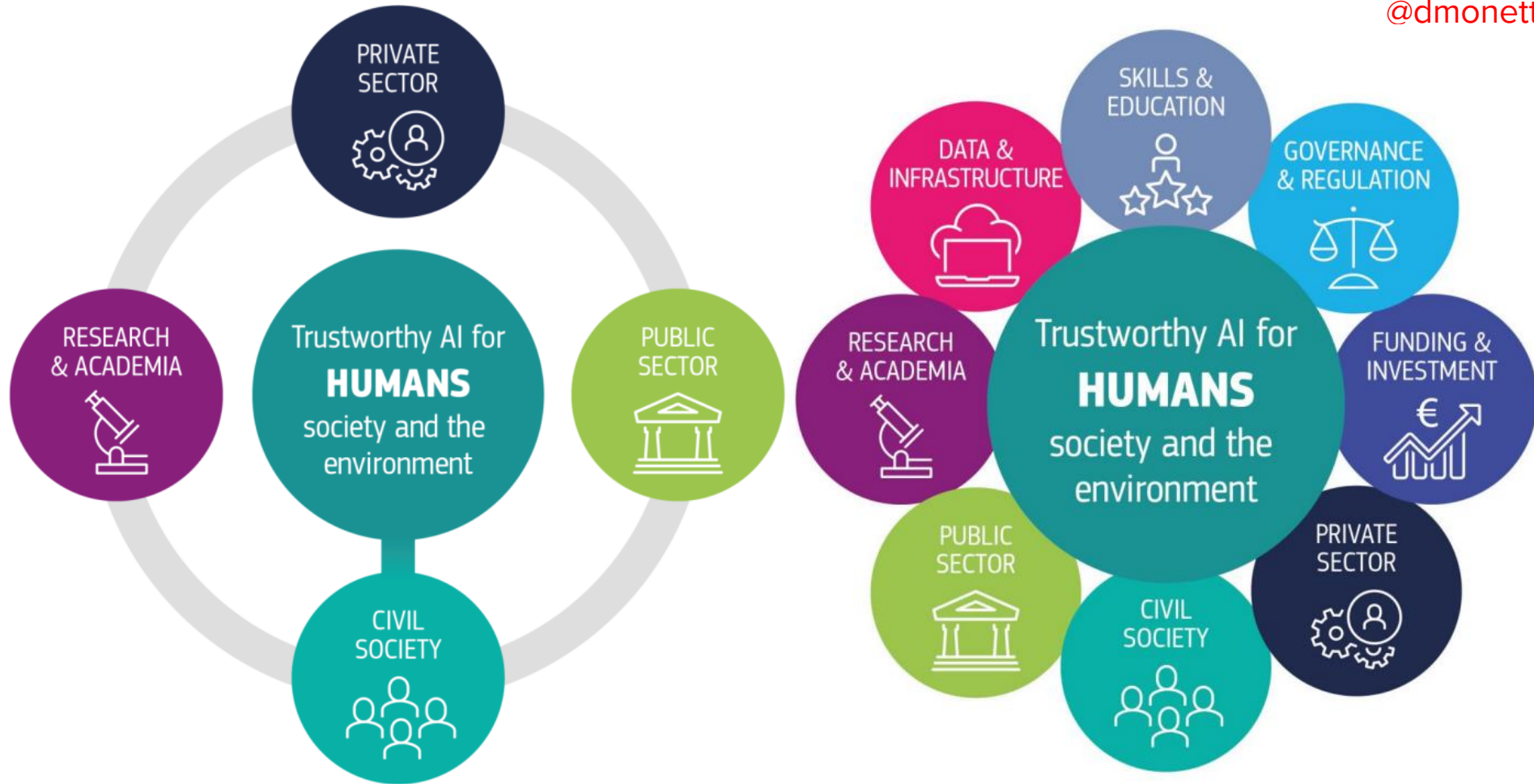
(Attard-Frost et al., 2022)



(EU, 2019b)

“[R]ecommendations that can guide Trustworthy AI towards sustainability, growth and competitiveness, as well as inclusion – while empowering, benefiting and protecting human beings.”

<https://digital-strategy.ec.europa.eu/en/library/policy-and-investment-recommendations-trustworthy-artificial-intelligence>



(Recommendations for Trustworthy AI, 2019, p. 8, 25)



(EU, 2020)

“[A] non-exhaustive Trustworthy AI assessment list [with 147 questions] to operationalise Trustworthy AI. It particularly applies to AI systems that directly interact with users, and is primarily addressed to developers and deployers of AI systems.”

<https://digital-strategy.ec.europa.eu/en/library/assessment-list-trustworthy-artificial-intelligence-altai-self-assessment>



What the company says
about its AI product



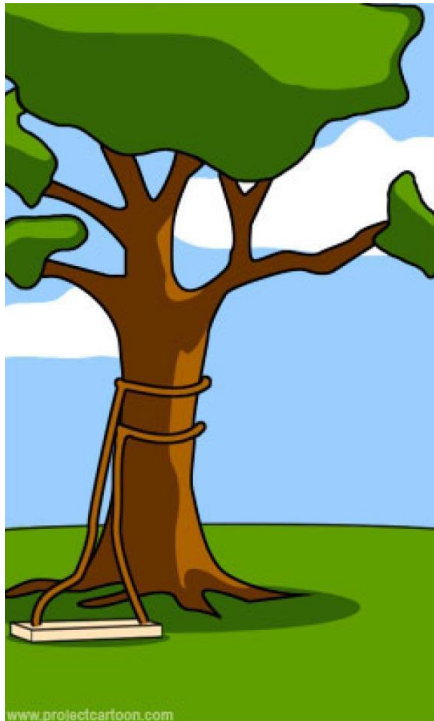
How robust the product is

*“Did you assess how your
system behaves in
unexpected situations and
environments?”*

(Trustworthy AI assessment list, 2019)



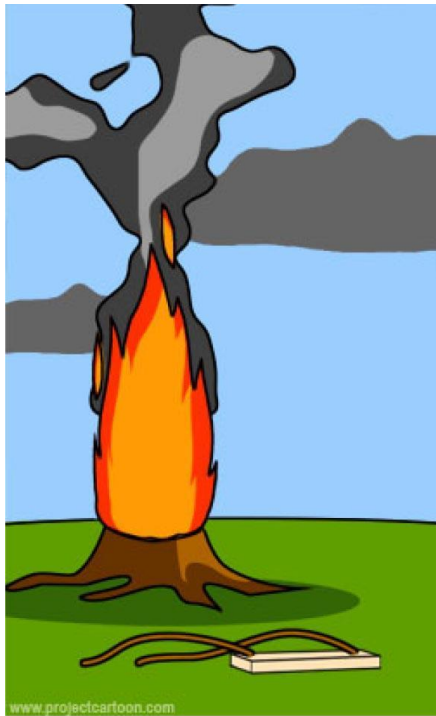
What the end user truly needs



How human-centered the product is

“In case the AI system features a chat bot or conversational system, are the human end users made aware of the fact that they are interacting with a non-human agent?”

(Trustworthy AI assessment list, 2019)



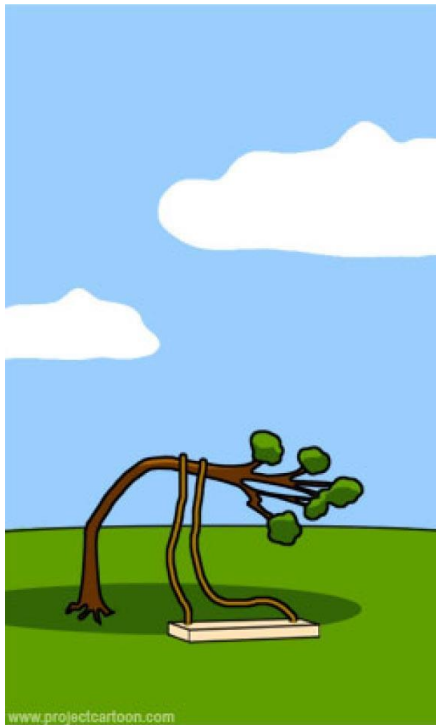
How the user data was processed



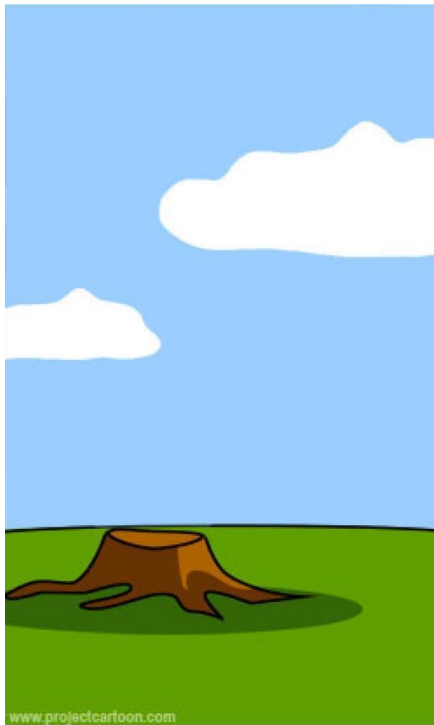
The product and the bias

“Did you consider ways to develop the AI system or train the model without or with minimal use of potentially sensitive or personal data?”

(Trustworthy AI assessment list, 2019)



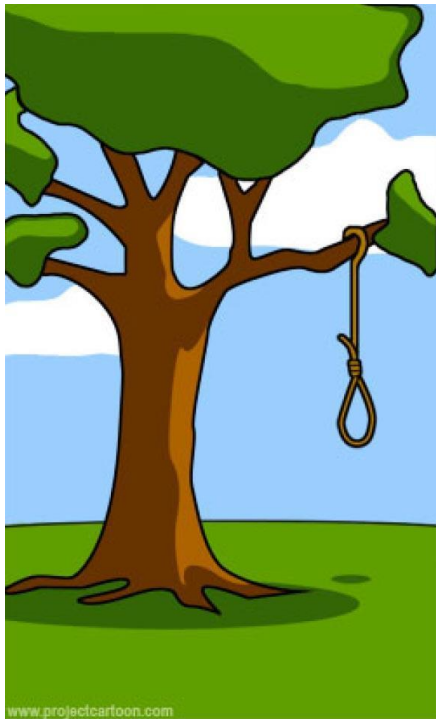
How fairness was considered



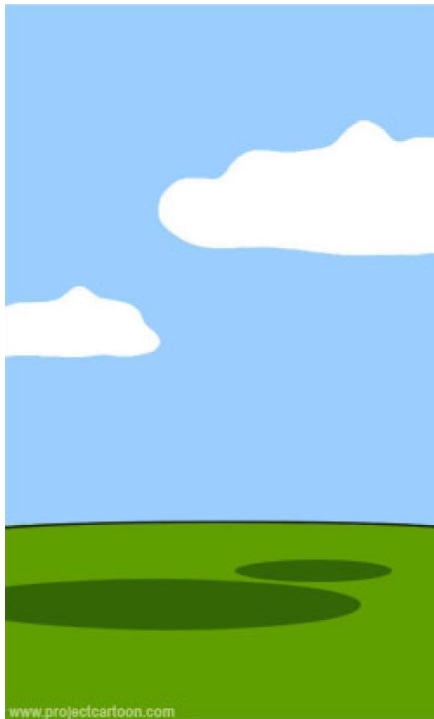
How ethical aspects were considered

“Did you consider diversity and representativeness of users in the data? Did you test for specific populations or problematic use cases?”

(Trustworthy AI assessment list, 2019)



Seeking for accountability

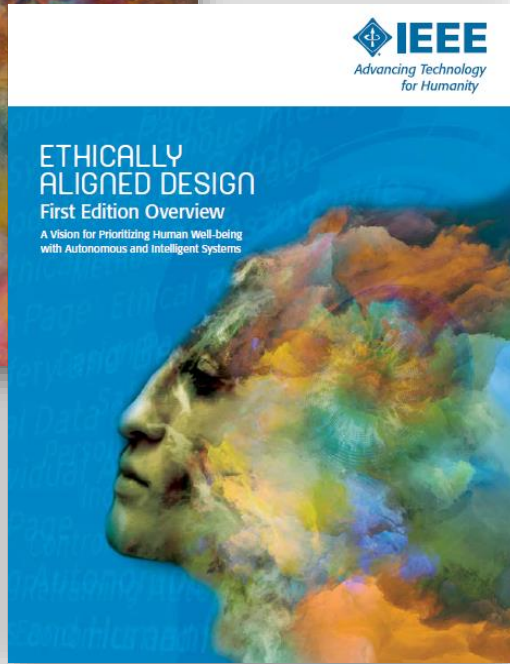
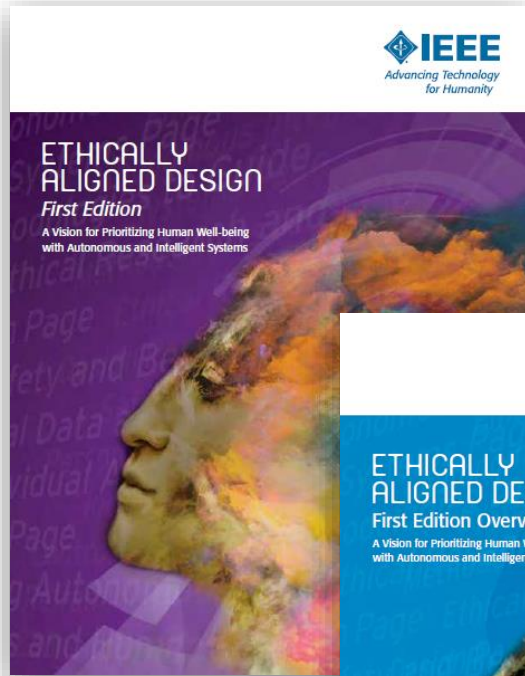


What the company thinks
about regulation

“Did you put in place mechanisms that facilitate the system’s auditability by internal and/or independent actors?”

(Trustworthy AI assessment list, 2019)

**What makes us human
makes us also
AI (ir)Responsible**



(IEEE, 2019)

*“[IEEE’s] commitment to engage global communities in creating **practical guidelines** that will **help ensure the development of ethically aligned autonomous and intelligent products, services, and systems.**”*

<https://ethicsinaction.ieee.org/#series>



(IEEE, 2019, p. 11)

- 1. Human Rights**—A/IS shall be created and operated to respect, promote, and protect internationally recognized human rights.
- 2. Well-being**—A/IS creators shall adopt increased human well-being as a primary success criterion for development.
- 3. Data Agency**—A/IS creators shall empower individuals with the ability to access and securely share their data, to maintain people’s capacity to have control over their identity.
- 4. Effectiveness**—A/IS creators and operators shall provide evidence of the effectiveness and fitness for purpose of A/IS.
- 5. Transparency**—The basis of a particular A/IS decision should always be discoverable.
- 6. Accountability**—A/IS shall be created and operated to provide an unambiguous rationale for all decisions made.
- 7. Awareness of Misuse**—A/IS creators shall guard against all potential misuses and risks of A/IS in operation.
- 8. Competence**—A/IS creators shall specify and operators shall adhere to the knowledge and skill required for safe and effective operation.

Be responsible.

Be responsible.

Perspective | [Published: 02 September 2019](#)

The global landscape of AI ethics guidelines

[Anna Jobin](#), [Marcello Ienca](#) & [Effy Vayena](#) 

[Nature Machine Intelligence](#) **1**, 389–399 (2019) | [Cite this article](#)

39k Accesses | **522** Citations | **740** Altmetric | [Metrics](#)

“84 documents containing ethical principles or guidelines for AI.”

“No single ethical principle appeared to be common to the entire corpus of documents, **although** there is an emerging convergence around the following principles: **transparency, justice and fairness, non-maleficence, responsibility, and privacy.**”

The Ethics of AI Business Practices: A Review of 47 AI Ethics Guidelines

AI and Ethics (Forthcoming)

55 Pages • Posted:

Blair Attard-Frost

University of Toronto, Faculty of Information

Andrés De los Ríos

Dublin, a Deloitte Business

Deneille R. Walters

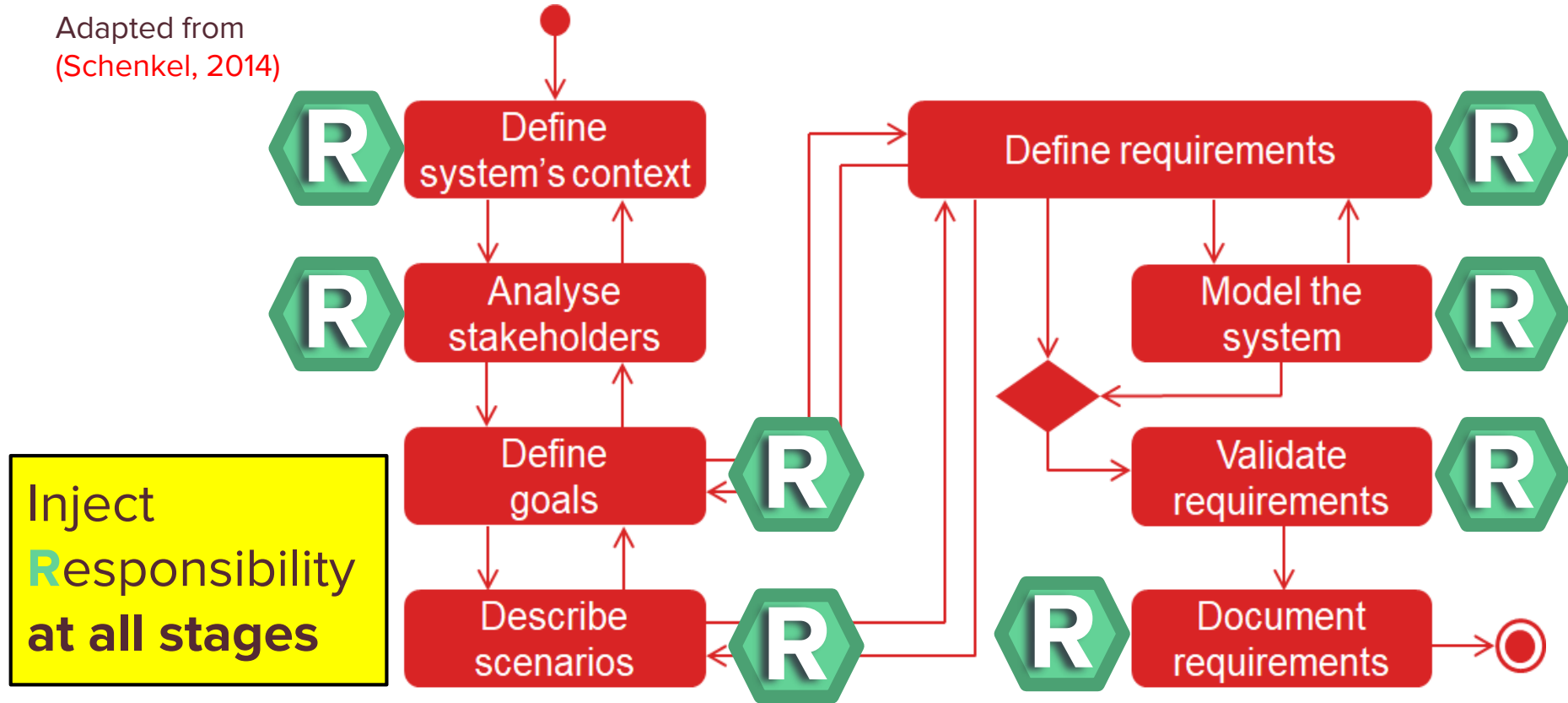
Independent Education Design & Technology Consultant

*“The **fairness** and **accountability** of AI business practices are the two most widely observed principles.”*

Develop responsible AI.

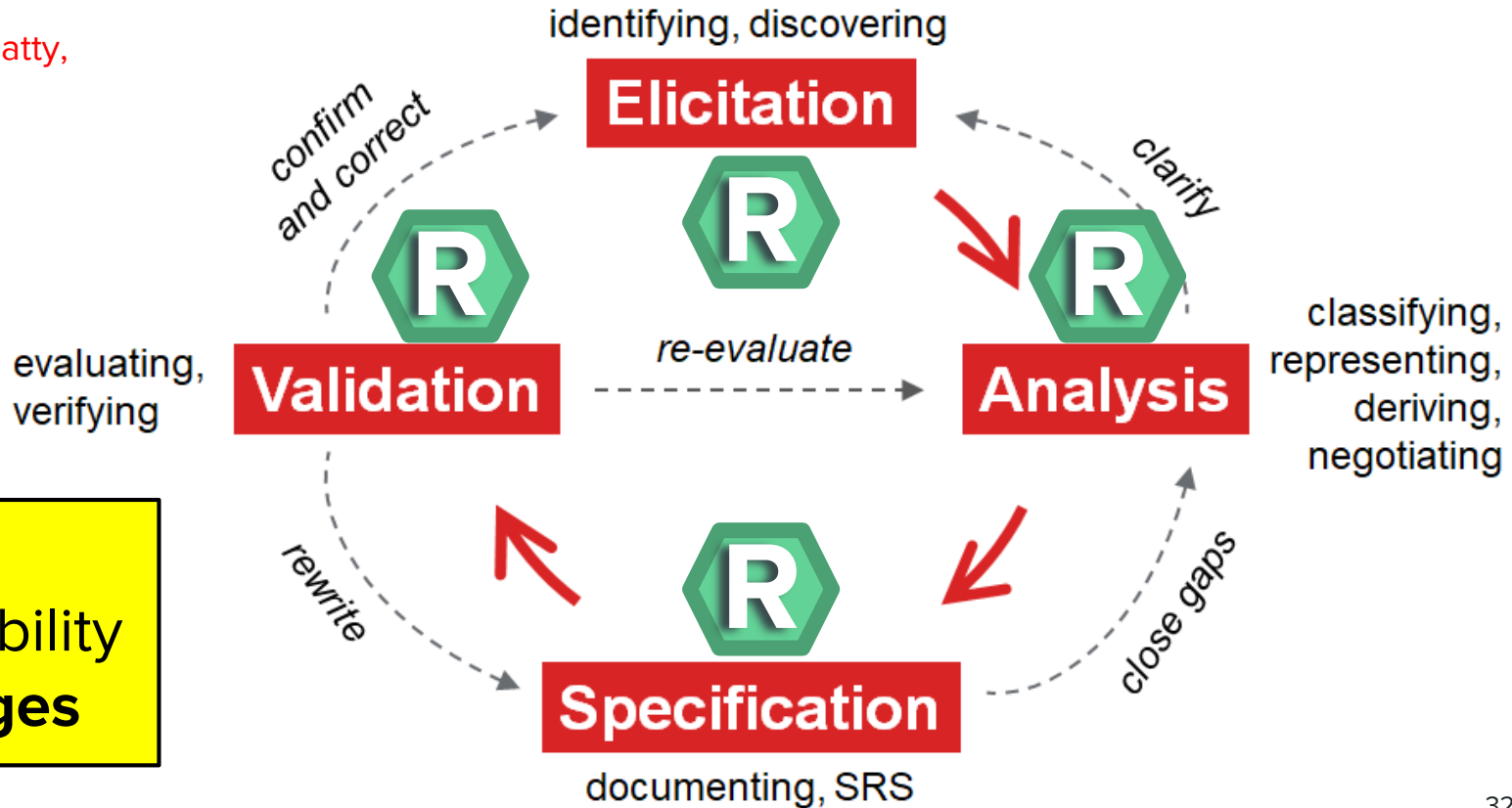
Requirements engineering as a process

Adapted from
(Schenkel, 2014)



Requirements development in detail

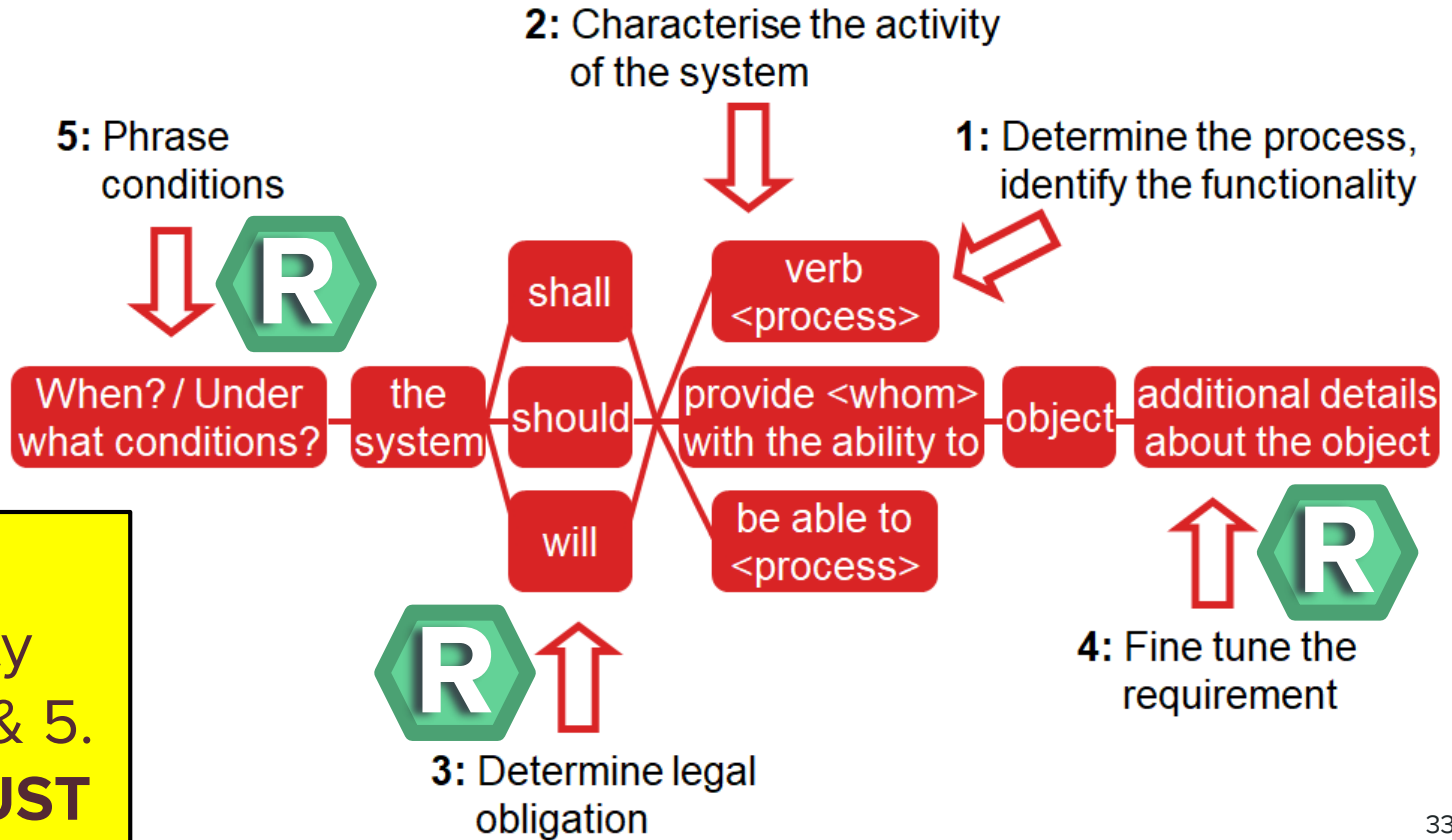
Adapted from
(Wiegers & Beatty,
2013)



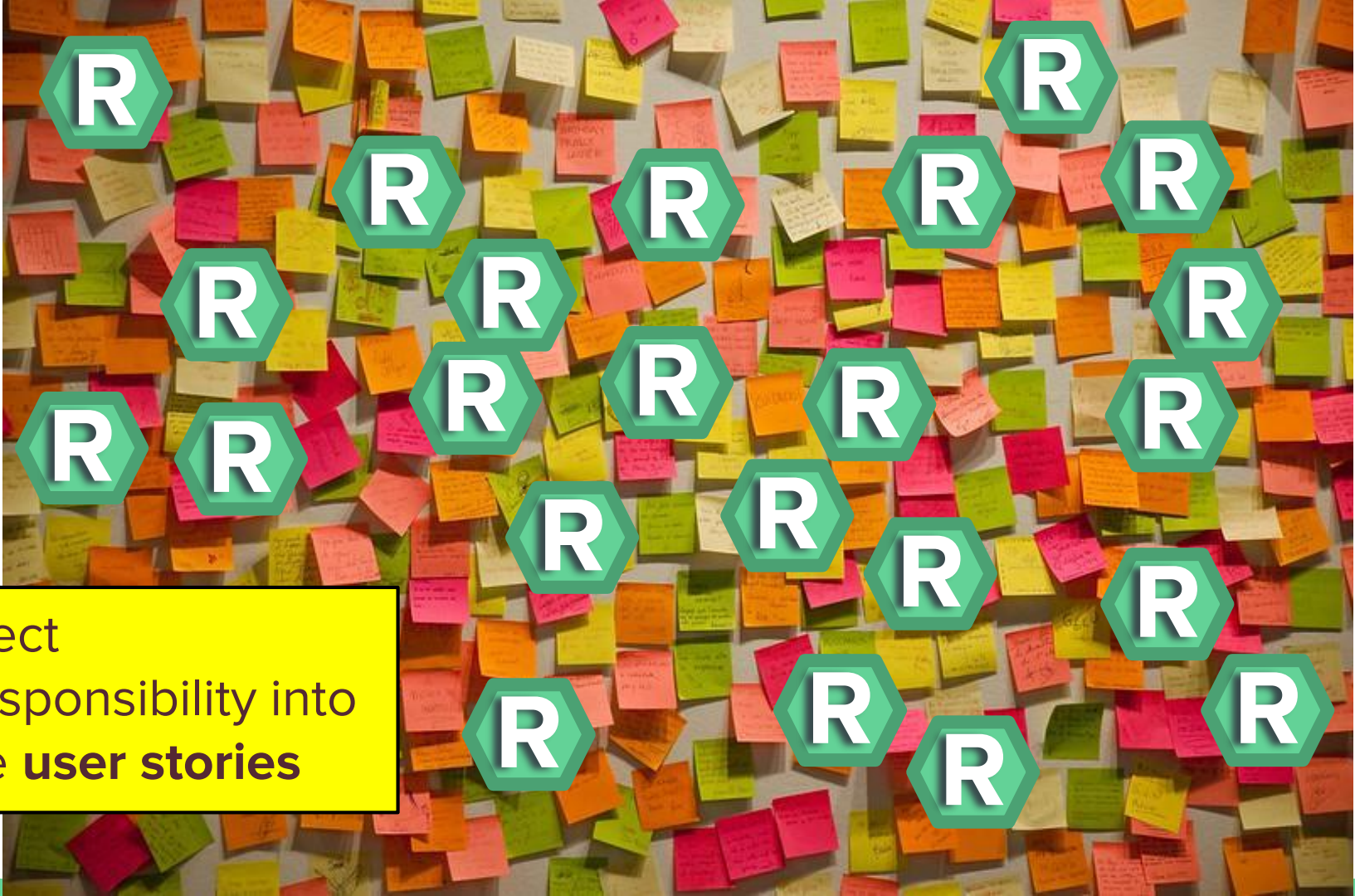
Inject
Responsibility
at all stages

High quality requirements

Adapted from
(Rupp et al., 2014)



Inject
Responsibility
into steps 4 & 5.
Make it a **MUST**



Inject
Responsibility into
the **user stories**

**Put it
on your backlog!**

That's how  gets done.

**Responsible AI needs
responsible humans.**

Be responsible.



Applied Machine Learning Days 2022

Tuesday March 29, 2022 / 10:45 AM - 11:00 AM

“*AI Ethics Standards and Guidelines*”



Prof. Dr.
Dagmar
Monett