



A Unified Framework of Five Principles for AI in Society

Titel	A Unified Framework of Five Principles for AI in Society
Publicatietype	Artikel
Publicatiejaar	2019
Auteurs	Floridi, L. [1], & Cowls J. [2]
Secondaire titel	Harvard Data Science Review
Volume	1
Publicatiedatum	6
RefMan	10345
Samenvatting	<p>Artificial Intelligence (AI) is already having a major impact on society. As a result, many organizations have launched a wide range of initiatives to establish ethical principles for the adoption of socially beneficial AI. Unfortunately, the sheer volume of proposed principles threatens to overwhelm and confuse. How might this problem of ‘principle proliferation’ be solved? In this paper, we report the results of a fine-grained analysis of several of the highest-profile sets of ethical principles for AI. We assess whether these principles converge upon a set of agreed-upon principles, or diverge, with significant disagreement over what constitutes ‘ethical AI.’ Our analysis finds a high degree of overlap among the sets of principles we analyze. We then identify an overarching framework consisting of five core principles for ethical AI. Four of them are core principles commonly used in bioethics: beneficence, non-maleficence, autonomy, and justice. On the basis of our comparative analysis, we argue that a new principle is needed in addition: explicability, understood as incorporating both the epistemological sense of intelligibility (as an answer to the question ‘how does it work?’) and in the ethical sense of accountability (as an answer to the question: ‘who is responsible for the way it works?’). In the ensuing discussion, we note the limitations and assess the implications of this ethical framework for future efforts to create laws, rules, technical standards, and best practices for ethical AI in a wide range of contexts.</p>
Aantekeningen	<p>Er staan een vijftal elementen centraal</p> <ul style="list-style-type: none">• Autonomy (autonomie, altijd de mogelijkheid om te besluiten en in te grijpen)• Beneficence (het goede doen)• Fairness (eerlijk)• Non-maleficence (niet schadelijk)• Explicability (Enabling the other principles of autonomy, beneficence, fairness, justice and non-maleficence through intelligibility and accountability.)
URL	https://hdsr.mitpress.mit.edu/pub/l0jsh9d1 [3]
DOI	10.1162/99608f92.8cd550d1 [4]
Citation Key	ref_10345

Levenscyclus:

- [1.1 Definitie \[5\]](#)
- [2.3 Bepalen van de grondslag en de betrouwbaarheid \[6\]](#)
- [3.2 Vaststellen verantwoordelijkheden \[7\]](#)
- [3.5 Documenteren \[8\]](#)
- [4.2 Bewaren context \[9\]](#)

Informatiemodel:

- [Context \[10\]](#)

Trefwoord:

- [accountability \[11\]](#)
- [artificial intelligence \[12\]](#)
- [autonomy \[13\]](#)
- [benificence \[14\]](#)
- [ethics \[15\]](#)
- [fairness \[16\]](#)
- [intelligibility \[17\]](#)
- [non-maleficence \[18\]](#)
- [explicability \[19\]](#)

Datum eerste publicatie:

zaterdag, 21 december 2019 - 9:06am



Bron-URL:<https://labyrinth.renkonker.nl/content/unified-framework-five-principles-ai-society>

Links

- [1] <https://labyrinth.renkonker.nl/biblio?f%5Bauthor%5D=2> [2]
- <https://labyrinth.renkonker.nl/biblio?f%5Bauthor%5D=757> [3] <https://hdsr.mitpress.mit.edu/pub/l0jsh9d1> [4]
- <http://dx.doi.org/10.1162/99608f92.8cd550d1> [5] <https://labyrinth.renkonker.nl/levenscyclus/11-definitie> [6]
- <https://labyrinth.renkonker.nl/levenscyclus/23-bepalen-van-de-grondslag-en-de-betrouwbaarheid> [7]
- <https://labyrinth.renkonker.nl/levenscyclus/32-vaststellen-verantwoordelijkheden> [8]
- <https://labyrinth.renkonker.nl/levenscyclus/35-documenteren> [9]
- <https://labyrinth.renkonker.nl/levenscyclus/42-bewaren-context> [10]
- <https://labyrinth.renkonker.nl/contextmodel/context> [11] <https://labyrinth.renkonker.nl/tags/accountability> [12]
- <https://labyrinth.renkonker.nl/tags/artificial-intelligence> [13] <https://labyrinth.renkonker.nl/tags/autonomy> [14]
- <https://labyrinth.renkonker.nl/tags/benificence> [15] <https://labyrinth.renkonker.nl/tags/ethics> [16]
- <https://labyrinth.renkonker.nl/tags/fairness> [17] <https://labyrinth.renkonker.nl/tags/intelligibility> [18]
- <https://labyrinth.renkonker.nl/tags/non-maleficence> [19] <https://labyrinth.renkonker.nl/tags/explicability> [20]
- <https://www.addtoany.com/share#url=https%3A%2F%2Flabyrinth.renkonker.nl%2Fcontent%2Funified-framework-five-principles-ai-society&title=A%20Unified%20Framework%20of%20Five%20Principles%20for%20AI%20in%20Society>

Information is a [conceptual labyrinth](#) and at the same time a [hyperobject](#)