Utilising the Assessment List for Trustworthy AI: Three Areas of Improvement

Adrian Gavornik¹, Juraj Podroužek¹, Matus Mesarcik¹, Sara Solarova¹, Stefan Oresko¹ and Maria Bielikova^{1,2}

¹Kempelen Institute of Intelligent Technologies, Mlynske nivy 5, Bratislava, 81109, Slovakia
²slovak.AI, Mlynske nivy 5, Bratislava, 81109, Slovakia

Abstract

This position paper contributes to the discussion on the utilisation of Assessment list for Trustworthy AI (ALTAI). Building on our own empirical experience stemming from cooperation with various AI teams and organisations, we suggest that ALTAI is suitable to be used as the core of the ethics-based assessment process. However, there are several areas in which ALTAI needs to be supplemented by additional tools and approaches. We introduce three extensions of ALTAI: (1) stakeholder analysis and segmentation; (2) data flows mapping; and (3) management of risks and countermeasures. We suggest that such enrichment will help ALTAI to narrow the gap between abstract principles and requirements of trustworthy AI and to increase the chances of implementing more ethical AI systems. Moreover, such an approach to ALTAI is also compatible with requirements expressed by existing and proposed EU legislation.

Keywords

AI ethics, ethics-based assessments, trustworthy AI, AI regulation, Assessment List for Trustworthy AI

1. Introduction and Background

The rapid development of AI systems in various areas of our lives together with the rising awareness of their unwanted harmful consequences have resulted in calls for more effective assessment and mitigation of societal and ethical risks. The field of AI ethics is now saturated with a plethora of guidelines and tools [1] that promise to deliver what is often called ethical or trustworthy AI based on universal human principles and values such as privacy, transparency, or fairness.

In a broader perspective, such frameworks aim to sensitize AI teams and organisations to become more aware of the role of moral deliberation as a process of actively identifying and weighing relevant ethical principles in the development, deployment, and use of AI systems. However, there seems to be a considerable gap between the rather abstract principles for trustworthy and ethical AI contained within such guidelines and the actual applicable practices that end up being implemented [2]. Some scholars go as far as calling AI Ethics "toothless

☆ adrian.gavornik@kinit.sk (A. Gavornik);

juraj.podrouzek@kinit.sk (J. Podroužek); matus.mesarcik@kinit.sk (M. Mesarcik); sara.solarova@kinit.sk (S. Solarova); stefan.oresko@kinit.sk (S. Oresko); maria.bielikova@kinit.sk

(M. Bielikova)

0000-0003-2800-3572 (S. Solarova); 0000-0002-5641-9279 (S. Oresko); 0000-0003-4105-3494 (M. Bielikova)



 2022 Coryright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).
 CEUR Workshop Proceedings (CEUR-WS.org) and trapped in an ethical principles approach" [3].

This paper contributes to the discussion on utilisation of Assessment list for Trustworthy AI (ALTAI) [4] which is a well-known AI ethics framework in Europe [5]. Based on our own empirical experience and the analysis of other similar ethical frameworks we suggest that AL-TAI is suitable to be used as the core of ethics-based [6] assessment of AI systems in organisations. However, we have identified three specific areas in which ALTAI falls short and needs to be supplemented with additional tools and approaches. Such extensions of ALTAI are validated by our own experience when facilitating numerous ethics-based assessments with AI teams and organisations, ranging from assessment of remote facial recognition systems [7] or social media sentiment analysis tools to research projects which employed AI techniques to audit recommendation algorithms of online platforms [8].

The main contribution of this paper lies in the opinion that the extension of ALTAI with (1) direct and indirect stakeholder analysis and their segmentation, (2) evaluation of data sources and mapping of data flows and (3) methodology for management of ethical risks and countermeasures will narrow the gap between abstract principles and practice and at the same time increase the chances of implementing more ethical and trustworthy choices in the development, deployment, and use of AI systems.

Our second contribution is to propose that our approach to ALTAI is compatible with requirements expressed by existing and proposed legislation such as the General Data Protection Regulation (GDPR) [9] or the currently prepared Artificial Intelligence Act (AIA) [10].

AIofAI '22: 2nd Workshop on Adverse Impacts and Collateral Effects of Artificial Intelligence Technologies, Vienna, Austria

 ^{0000-0001-5678-0083 (}A. Gavornik); 0000-0002-9691-0310
 (J. Podroužek); 0000-0003-3311-5333 (M. Mesarcik);

2. A Critical Reflection of ALTAI

One of the most well-known guidelines in the European context which established a benchmark for what may qualify as trustworthy AI [11] is the Ethics Guidelines for Trustworthy AI (EGTAI), developed by the High-Level Expert Group on Artificial Intelligence set up by the European Commission in 2019 [12]. Trustworthy AI introduced in EGTAI should meet three criteria: to be lawful, ethical, and robust; while conforming to several specific requirements and ethical principles.

The third chapter of EGTAI, later corroborated in a separate document - ALTAI, is an example of an ethicsbased framework that intends to guide the organisations from early development phases of their AI systems [4]. It proposes and explains seven requirements of trustworthy AI: (1) Human Agency and Oversight, (2) Technical Robustness and Safety, (3) Transparency, (4) Diversity, (5) Non-discrimination and Fairness, (6) Societal and Environmental Well-being and (7) Accountability. ALTAI consists of more than one hundred questions and as a result, offers a set of recommendations for AI teams and organisations for each requirement. Requirements of trustworthy AI are not chosen arbitrarily and mostly overlap with areas present in other guidelines that focus on ethical principles of AI [13].

However, to this day we are not aware of an established procedure that would help AI practitioners to fully operationalise ALTAI into practice. There are some reports and studies that evaluate ALTAI alone or as a part of EGTAI [14], yet in this case, they mostly offer suggestions on how to adjust ALTAI questions to bring more clarity to the assessment process, or analyse it in light of the relevant legal and policymaking context [15], without necessarily focusing on how to ensure proper implementation. Other studies attempt to contextualise ALTAI with respect to specific domains. These attempts are at this moment mostly coming from the perspective of engineers who are trying to evaluate their AI systems to be compliant with the ALTAI recommendations, while also pointing out certain areas of ALTAI which are not applicable to their specific domains of expertise, such as Advanced Driver-Assistance Systems (ADAS) [16] or in AI systems used in healthcare [17].

We adopt a broader view on ALTAI as ethics-based assessment. Ethics based assessments and audits can be characterised by structured process used by AI teams and organisations to control or influence the behaviour of their AI systems to achieve consistency with ethical principles or norms [6]. We argue that there are at least three areas that need to be extended in order to ensure that the principles of trustworthy AI enshrined in AL-TAI are successfully operationalised. These are (1) stakeholder analysis and segmentation, (2) data flows mapping and (3) management of risks and countermeasures. Some of these elements are already reflected in other ethical frameworks [18, 19, 20, 21] although not sufficiently covered in ALTAI. We are convinced that such extensions would offer a more contextual understanding of the issues related to trustworthy AI that ALTAI has the ambition to open up.

2.1. Stakeholder Analysis and Segmentation

Although ALTAI dedicates a separate section to stakeholder participation in the requirement for *Diversity, Nondiscrimination, and Fairness*, generally it lacks a clear strategy toward various entities that can be affected by unintended harmful consequences of the AI systems. Assessment questions such as "Is the AI system designed to interact, guide or take decisions by human end-users that affect humans ('subjects') or society?" implicitly suggest that there exist a multitude of directly or indirectly affected groups of stakeholders that deserve more subtle distinction than between end-users, subjects and society.

While direct stakeholders represent individuals and groups directly affected by a technology [22] because they design it, use it, or have to manage it, indirect stakeholders are often more difficult to consider since they are affected by the AI system but not in a direct way or immediately. Nonetheless the impacts on indirect stakeholders could be still considerable. Therefore, we suggest that stakeholder analysis and the proper segmentation of direct and indirect stakeholders resemble an essential starting point during the facilitated ethics-based assessment process.

In order to identify a larger set of direct and indirect stakeholders, we propose to enrich stakeholder identification procedure with two additional factors considered:

- 1. a degree of foreseeable benefits and harms of the AI system on the stakeholder,
- 2. a degree to which stakeholder is aware or sensitive to foreseeable benefits and harms. Such analysis, which is inspired by existing stakeholders maps [23], results in a matrix (see Figure 1) that besides better identification of the most vulnerable groups of stakeholders is important for several other reasons: (i) it helps manage the expectations of participants towards the rest of assessment process by providing them with clear categorisation of the affected groups; (ii) from a risk-based perspective, organisations may consider it helpful to know who and to what degree will be affected by the countermeasures that are yet to be implemented; (iii) by establishing separate communication strategies towards various stakeholder groups, the general transparency of the development process can be increased.



Figure 1: Example of Impact/Sensitivity matrix of direct (blue) and indirect (yellow) stakeholders of sentiment analysis tool for content moderation on social media separated into four communication segments

2.2. Data Flows Mapping

Despite the fact that *Privacy and Data Governance* is also amongst the seven requirements of trustworthy AI, AL-TAI on its behalf asks organisations mostly general questions such as "Did you consider the privacy and data protection implications of data collected, generated or processed over the course of the AI system's life cycle?". However, we have often observed that such questions fail to help AI teams and organisation to identify specific ethical issues around data processing. The ethical issues concerning data flows also differ across time, which often makes some of these processes more problematic than others. For example, some data sources, like public data crawled from the internet, might have been generally acceptable by stakeholders a few years ago, but might not be now.

In order to better navigate and address possible risks that emerge from data collection and data processing in general, we found it useful to map out relevant ethical issues on how data moves across the AI systems considering the affected stakeholders. By utilising tools like Data Ecosystem Map [24], a visual representation of the data flows can be easily generated together with additional information like reliability, appropriateness or fair use of data sources. Or it can be supplemented by the analysis of the impacts on individual rights of stakeholders to better understand some of the additional ethical impacts of data collection and data sharing that might have otherwise been overlooked by just following ALTAI questions.

2.3. Management of Risks and Countermeasures

Based on the answers provided by the organisations, ALTAI generates a list of recommendations for each of the seven requirements. However, even if ALTAI questions can sensitize AI teams and organisations to be more aware of possible ethical risks, recommendations such as "Take measures to consider the impact of the AI system on the right to privacy, the right to physical, mental and/or moral integrity and the right to data protection" or "Establish mechanisms to ensure fairness in your AI system" will not help them devise concrete countermeasures that need to be implemented within their AI systems.

When facilitating ethics-based assessments with AI teams and organisations we found it more constructive to use ALTAI questions as a basis for the moral deliberation process leading to the explication of ethical and societal risks concerning AI systems. Therefore we propose to reinforce the risk-based approach in ALTAI as it is already present in other AI ethics assessments [20, 25]. During the assessment process with AI teams and organisations we found it useful to evaluate these risks regarding four dimensions: (1) most affected stakeholder groups; (2) like-

Table 1	
Fragment of a risk matrix for facial recognition system (FF	₹T)

Description	Consequences	Most affected stakeholders	Likelihood (1-5)	Impact (1-5)	Action category	Countermeasures
Forced recognition	People will be forced to biometric identification without exceptions.	Uninformed captured person, Informed captured person	5	4	Influence	Best practice guidelines for integrators (separate entrances for conventional access).
Lack of awareness	People will not be aware of the purpose and aims of facial recognition in particular space.	Informed captured person, Uninformed captured person, Society	4	4	Act	Inform people about the use of FRT technologies before entering the area and explain what a person can expect before opt-in
Over-reliance	People will rely too much on machine decisions and do not use their own reasons and capabilities.	Operators (Users), Third-party integrators	4	3	Influence	Onboarding courses for operators and partners about the the edge cases of facial recognition. Prepare guidelines on system accuracy in specific use cases.

lihood of ethical risks; (3) severity of impacts on stakeholders; and (4) the type of action organisations could provide.

We are of the opinion that such risk matrix (Table 1) better categorises the ethical risks and countermeasures in relation to the specific stakeholders and helps to determine which actionable steps need to be taken foremost. It also navigates AI teams and organisations to create a plan of action after they have worked through all of the questions from ALTAI and proved to be helpful in translating general and unordered recommendations proposed by the ALTAI into a more comprehensive and well-tailored fashion.

3. Ex-ante Assessments in the Context of Current EU Law

We have proposed some adjustments to ALTAI that better translate the principles of ethical and trustworthy AI into practice. But at the same time, we suggest that ALTAI conceived as an ethics-based assessment should also conform with legal requirements as foreseen by applicable or proposed EU law. We are of opinion that our extension of ALTAI is highly compatible with legal requirements in the EU data protection law and proposed regulation in AIA (Figure 2).

Ex-ante assessments are not novel when it comes to the legislation in the law of the European Union (EU), but may differ in the methodology and purposes. For example, the area of product regulation is inherently connected to conducting conformity assessments. This in practice means that when a manufacturer wants to place a product on the EU market, it has to undergo a conformity assessment to ensure compliance with requirements prescribed by law. On the other hand, the fundamental rights impact assessments are part of the binding legislation though limited to the area of data protection (Article 35, GDPR). The purpose of such assessments is to foster accountability of various entities and alter internal processes to mitigate risks for fundamental rights and freedoms. They also function as an "early warning system" in case of deployment of emerging technologies into practice [26]. Even ALTAI mentions fundamental rights impact assessment as one of its prerequisites.

EU General Data Protection Regulation (GDPR) applies to the automatic and manual processing of personal data. In practice, any processing of information about identified or identifiable information triggers the scope of GDPR [27, 28]. The burden of compliance with EU data protection lies primarily on the shoulders of controllers determining purposes of processing and processors processing personal data on behalf of controllers. Within the principle of accountability, the controller shall prior to processing of personal data in specific cases conduct data protection impact assessment or "DPIA" (Article 35, GDPR).

The obligation to conduct DPIA applies in general where "processing is likely to result in a high risk to the rights and freedoms of natural persons" (Article 35 [1], GDPR) or in specific cases stipulated explicitly in the law (Article 35 [3], GDPR). The European Data Protection Board provides criteria to take into account to assess risk [29]. Based on the wording of the GDPR, DPIA shall consist of at least: (1) a systematic description of the envisaged processing operations and the purposes of the processing; (2) an assessment of the necessity and proportionality of the processing operations; (3) an assessment of the risks to the rights and freedoms of data subjects; and (4) the measures envisaged to address the



Figure 2: Process of modified ethics-based assessment established on ALTAI including key legal requirements

risks. Views of involved parties shall be assessed as well (Article 35 (9), GDPR).

The proposal for the Artificial Intelligence Act (AIA) introduced by the European Commission in the spring of 2021 aims to provide the EU with first comprehensive horizontal regulation of AI. Following the risk-based approach, AIA is primarily focused on obligations for high-risk AI systems, either standalone highrisk AI systems, e.g. biometrics systems, or AI systems used as a safety component of products covered by the EU harmonization legislation. The key requirements for high-risk AI systems are embedded in the requirements of compliance with quality management system, including data and data governance (Article 10, AIA), human oversight (Article 14, AIA), or accuracy, robustness, and cybersecurity (Article 15, AIA).

Prior to placing a high-risk AI system on the market, the provider is obliged to undergo a conformity assessment. The majority of providers of high-risk AI systems should follow the procedure for conducting conformity assessment based on internal control as external control is obligatory only in a limited number of cases [30]. Conformity assessment based on the internal control consists of (1) verification of quality management system; (2) examination of the information provided in the technical documentation; and (3) an assessment of the risks to the rights and freedoms of data subjects; and (4) verification of design and development process of the AI system including post-market monitoring plan.

3.1. Stakeholder Analysis and Segmentation

Concerning stakeholder analysis, GDPR requires collection of views of data subjects or their representatives and controller's Data Protection Officer in case of mandatory DPIA (Article 35 [9], GDPR [29]). In AIA the providers of high-risk AI systems are obliged to systematically collect, document, and analyze relevant data provided by business users within post-market monitoring plans (Article 61 [2], AIA). Data from the customers should be covered by this requirement. Therefore, proper stakeholder analysis is not novel within the realm of law. Indeed, it is already part of ex-ante assessments despite focusing on business users. Our approach conceives stakeholder analysis in a broader context but is still well-compatible with DPIA according to GDPR concerning views of data subjects and post-market monitoring plans pursuant to AIA.

3.2. Data Flows Mapping

GDPR and AIA both stipulate that data flows mapping represents an essential part of ex-ante assessments. The crucial part of DPIA consists of "a systematic description of the processing" operations (Article 35 [7] [a], GDPR) including data flows, sources of data and data sharing. Furthermore, AIA proposes mandatory data and data governance measures concerning high-risk AI systems (Article 10, AIA) namely compliance with GDPR, assessment of quality of data, design choices, or examination in view of possible biases. All of these aspects form an integral part of our extension of ALTAI. In general, data flows maps may be subsequently used as a basis for adopting supporting measures pursuant to AIA and form a part of DPIA.

3.3. Management of Risks and Countermeasures

We proposed to supplement ALTAI with a risk matrix to help AI teams and organisations identify and categorise ethical risks around their systems. By GDPR the analysis of privacy and personal data issues may aid controllers to conduct DPIA and mitigate risks for rights, freedoms and interests of data subjects. On the other hand, AIA proposes legal requirements towards human oversight, cybersecurity, robustness or bias mitigation. Evaluating these aspects forms an integral part of conformity assessments pursuant to AIA. Furthermore, one of the prerequisite for conducting self-assessment pursuant to AIA is adherence to aforementioned requirements.

By identifying and managing ethical risks and respective countermeasures we may better mitigate potential unethical and illegal behaviour of AI systems in advance. Additionally, as trustworthiness of AI systems is emphasized in the text of AIA (Recital 5 & 16, AIA) and ALTAI is acknowledged as state-of-the-art minimum requirement towards conformity assessment (Part 5.2.3, AIA Impact Assessment [31]), our extension to ALTAI considering proper management of specific risks and countermeasures represent a starting point for conducting conformity assessments pursuant to AIA.

4. Open Issues and Future Work

We have expressed our opinion that (1) stakeholder analysis and segmentation, (2) mapping of data flows and (3) management of ethical risks and countermeasures have the capacity to transform ALTAI into a more practical ethics-based assessment. However, some aspects of our approach remain to be explored and developed further.

First, additional steps should be devised that will manage the expectations and goals of the AI teams and organisations from early phases of ethics-based assessment, which are already present in other ethical frameworks [19, 32] but their proper implementation into ALTAI still poses an open question for us. ALTAI also lacks transparent methods concerning operationalisation of its general recommendations into specific implementation tasks. We are of opinion that both methods would help the AI teams and organisations to have a clearer idea of the added value of ALTAI conceived as ethics-based assessment.

There is also a challenge in how to promote the direct participation of some of the most affected stakeholders during the assessment process. Furthermore, concerning the transparent communication strategies based on segmentation from the stakeholder matrix, a set of specific practices on stakeholders engagement should be further conceptualized and developed. These should serve as an oversight mechanism that would ensure that the communication strategies are implemented in an appropriate manner.

AI teams and organisations conducting ethics-based assessments and audits also face some considerable risks in attempting to operationalise AI ethics principles on their own. One of such well-known risks is *ethics shopping* [33] which concerns the organisations that selectively choose and adapt only those ethical principles and recommendations that require the least amount of implementation and change of their behavior. The other risk can be described as ethics bluewashing [33]. This risk is tied to pretending to be ethical by implementing superficial measures or making unsubstantiated claims. We perceive both of the risks as imminent threats not only for ALTAI but for all ethics-based assessments. We believe that the strong appeal towards mapping ethical principles and requirements on specific risks and countermeasures implemented into the design of the AI system can at least partially combat these risks. ALTAI partially contributes to the emergence of these two risks by asking vague questions and proposing general recommendations that the AI teams and organisations might struggle to understand and translate into their specific technical domain. Therefore we believe that the presence of the third-party facilitators and experts [18] should play an important role in clarifying ambiguous and contextual questions presented in ALTAI or the ethical trade-offs that might emerge between various requirements.

There is also an open question of whether ALTAI enriched by proposed extensions will be able to form an essential part of future conformity assessment procedures in compliance with AIA which begin to emerge recently [6]. We believe that the assessment of ethical risks should form a binding part of AIA conformity assessments [34] and that ALTAI may take an important role in this process. Still, when conducting any AI ethics-based assessment or audit, one should remain cautious concerning complexity of covered issues. Pursuing trustworthy and ethical AI should not be boiled down to merely following formal procedures and checklists. For this, we would like to emphasize that conducting the ethics-based assessment should be understood as a continuous dialogue present at all stages of development and deployment of AI systems.

5. Conclusion

In this paper we have addressed the gap in operationalising the principles for trustworthy and ethical AI to their real-world practical implementation. We focused on AL-TAI as an example of a well-known ethical framework for assessing AI systems. Drawing from our own empirical experience with conducting ethics-based assessments with AI teams and organisations, we have identified three areas in which ALTAI should be expanded in order to turn general requirements of trustworthy AI into practical implementation. These are namely (1) stakeholder analysis and segmentation, (2) data flows mapping and (3) management of ethical risks and countermeasures. We suggest that such an extension of ALTAI results in a more comprehensible and measurable set of action steps. At the same time, we map out how these areas are already present and supported by existing legal requirements stemming from EU GDPR and AIA. We conclude with general reflection on the possible risks and open issues that emerge when conducting ethics-based assessments and the possible future directions in operationalising AL-TAI requirements.

Acknowledgments

This research was partially supported by TAILOR, a project funded by EU Horizon 2020 research and innovation programme under GA No. 952215.

References

- [1] J. Morley, L. Floridi, L. Kinsey, A. Elhalal, From What to How: An Initial Review of Publicly Available AI Ethics Tools, Methods and Research to Translate Principles into Practices, Science and Engineering Ethics 26 (2020) 2141–2168. doi:10.1007/ s11948-019-00165-5.
- [2] J. C. Ibáñez, M. V. Olmeda, Operationalising AI ethics: how are companies bridging the gap between practice and principles? An exploratory study, AI & SOCIETY (2021). doi:10.1007/ s00146-021-01267-0.
- [3] A. Rességuier, R. Rodrigues, AI ethics should not remain toothless! A call to bring back the teeth of ethics, Big Data & Society 7 (2020). doi:10.1177/ 2053951720942541.
- [4] European Commission, Directorate-General for Communications Networks Content and Technology, The Assessment List for Trustworthy Artificial Intelligence (ALTAI) for self assessment., Publications Office, LU, 2020. URL: https://data.europa.eu/ doi/10.2759/791819.
- [5] J. Ayling, A. Chapman, Putting AI ethics to work: are the tools fit for purpose?, AI and Ethics (2021). doi:10.1007/s43681-021-00084-x.
- [6] L. Floridi, M. Holweg, M. Taddeo, J. Amaya Silva, J. Mökander, Y. Wen, capAI - A Procedure for Conducting Conformity Assessment of AI Systems in Line with the EU Artificial Intelligence Act, SSRN Electronic Journal (2022). doi:10.2139/ ssrn.4064091.
- [7] S. Solarova, J. Podroužek, M. Mesarčík, A. Gavornik, M. Bielikova, Reconsidering the regulation of facial recognition in public spaces, AI and Ethics (2022). doi:10.1007/s43681-022-00194-0.
- [8] M. Tomlein, B. Pecher, J. Simko, I. Srba, R. Moro, E. Stefancova, M. Kompan, A. Hrckova, J. Podrouzek, M. Bielikova, An Audit of Misinformation Filter Bubbles on YouTube: Bubble Burst-

ing and Recent Behavior Changes, Association for Computing Machinery, New York, NY, USA, 2021, p. 1–11. doi:https://doi.org/10.1145/ 3460231.3474241.

- [9] E. Parliament, Regulation (eu) 2016/679 of the european parliament and of the council of 27 april 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing directive 95/46/ec (general data protection regulation), 2016. http://data.europa.eu/eli/reg/2016/679/2016-05-04.
- [10] European Commission, Proposal For a Regulation of The European Parliament and of The Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts, COM/2021/206 final, 2021/0106(COD), 2021. https: //eur-lex.europa.eu/legal-content/EN/TXT/?uri= CELEX:52021PC0206.
- [11] L. Floridi, Establishing the rules for building trustworthy AI, Nature Machine Intelligence 1 (2019) 261–262. doi:10.1038/s42256-019-0055-y.
- [12] European Commission, Directorate-General for Communications Networks, Content and Technology, Ethics guidelines for trustworthy AI, Publications Office, 2019. doi:10.2759/177365.
- [13] J. Fjeld, N. Achten, H. Hilligoss, A. Nagy, M. Srikumar, Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-Based Approaches to Principles for AI, SSRN Scholarly Paper ID 3518482, Social Science Research Network, Rochester, NY, 2020. doi:10.2139/ssrn.3518482.
- [14] L. Pupillo, A. Ferreira, S. Fantin, CEPS Task Force on Artificial Intelligence and Cybersecurity Technology, Governance and Policy Challenges Task Force Evaluation of the HLEG Trustworthy AI Assessment List (Pilot Version). CEPS Task Force Report 22 January 2020, 2020. URL: http://aei.pitt.edu/ 102463/, num Pages: 34.
- [15] C. Stix, Actionable Principles for Artificial Intelligence Policy: Three Pathways, Science and Engineering Ethics 27 (2021) 15. doi:10.1007/s11948-020-00277-3.
- [16] M. Borg, J. Bronson, L. Christensson, F. Olsson, O. Lennartsson, E. Sonnsjö, H. Ebabi, M. Karsberg, Exploring the Assessment List for Trustworthy AI in the Context of Advanced Driver-Assistance Systems, arXiv:2103.09051 [cs] (2021). doi:https: //doi.org/10.48550/arXiv.2103.09051.
- [17] R. V. Zicari, J. Brusseau, S. N. Blomberg, H. C. Christensen, M. Coffee, M. B. Ganapini, S. Gerke, T. K. Gilbert, E. Hickman, E. Hildt, S. Holm, U. Kühne, V. I. Madai, W. Osika, A. Spezzatti, E. Schnebel, J. J. Tithi, D. Vetter, M. Westerlund, R. Wurth, J. Amann,

V. Antun, V. Beretta, F. Bruneault, E. Campano, B. Düdder, A. Gallucci, E. Goffi, C. B. Haase, T. Hagendorff, P. Kringen, F. Möslein, D. Ottenheimer, M. Ozols, L. Palazzani, M. Petrin, K. Tafur, J. Tørresen, H. Volland, G. Kararigas, On Assessing Trustworthy AI in Healthcare. Machine Learning as a Supportive Tool to Recognize Cardiac Arrest in Emergency Calls, Frontiers in Human Dynamics 3 (2021) 673104. doi:10.3389/fhumd.2021.673104.

- [18] R. V. Zicari, J. Brodersen, J. Brusseau, B. Dudder, T. Eichhorn, T. Ivanov, G. Kararigas, P. Kringen, M. McCullough, F. Moslein, N. Mushtaq, G. Roig, N. Sturtz, K. Tolle, J. J. Tithi, I. van Halem, M. Westerlund, Z-Inspection[®] : A Process to Assess Trustworthy AI, IEEE Transactions on Technology and Society 2 (2021) 83–97. doi:10.1109/ TTS.2021.3066209.
- [19] IEEE Recommended Practice for Assessing the Impact of Autonomous and Intelligent Systems on Human Well-Being, IEEE, S.I., 2020. URL: https: //ieeexplore.ieee.org/servlet/opac?punumber = 9084217, oCLC: 1156335739.
- [20] D. Anderson, J. Bonaguro, M. McKinney, A. Nicklin, J. Wiseman, Ethics & algorithms toolkit, 2019. URL: https://ethicstoolkit.ai/.
- [21] Personal Data Protection Commission Singapore, Model ai governance framework, 2020. URL: https: //www.pdpc.gov.sg/help-and-resources/2020/01/ model-ai-governance-framework.
- [22] B. Friedman, D. G. Hendry, A. Borning, A survey of value sensitive design methods, Foundations and Trends[®] in Human–Computer Interaction 11 (2017) 63–125. doi:10.1561/1100000015.
- [23] D. L. Airi, A. Porcari, D. Pimponi, E. Borsella, E. Mantovani, Prisma rri-csr roadmap (2019).
- [24] O. D. Institute, Data Ecosystem Mapping: tool and guide, 2019. URL: https://theodi.org/article/dataecosystem-mapping-tool/.
- [25] Digital Dubai Office, AI System Ethics Self-Assessment Tool, 2020. URL: https: //www.digitaldubai.ae/self-assessment.
- [26] D. Kloza, N. van Dijk, R. Gellert, I. Böröcz, A. Tanas, E. Mantovani, P. Quinn, Data protection impact assessments in the european union: Complementing the new legal framework towards a more robust protection of individuals, 2017. doi:10.31228/ osf.io/b68em.
- [27] The Court of Justice of the European Union C-582/14 Patrick Breyer (ECLI:EU:C:2016:779), 2016.
- [28] The Court of Justice of the European Union C-434/16 Peter Nowak (ECLI:EU:C:2017:994), 2017.
- [29] European Data Protection Board, Guidelines on data protection impact assessment (dpia) and determining whether processing is "likely to re-

sult in a high risk" for the purposes of regulation 2016/679, 2017. URL: https://ec.europa.eu/ newsroom/article29/items/611236/en.

- [30] M. Veale, F. J. Z. Borgesius, Demystifying the draft EU artificial intelligence act, CoRR abs/2107.03721 (2021). doi:doi.org/10.48550/ arXiv.2107.03721.
- [31] European Commission, Impact assessment accompanying the proposal for a regulation of the european parliament and of the council laying down harmonised rules on artificial intelligence (artificial intelligence act) and amending certain union legislative acts, com/2021/206 final, 2021/0106(cod), 2021. https://ec.europa.eu/info/ law/better-regulation/have-your-say/initiatives/ 12527-Artificial-intelligence-ethical-and-legalrequirements_en.
- [32] Government of Canada, Algorithmic impact assessment, 2020. URL: https://open.canada.ca/aia-eiajs/?lang=en.
- [33] L. Floridi, Translating Principles into Practices of Digital Ethics: Five Risks of Being Unethical, Philosophy & Technology 32 (2019) 185–193. doi:10.1007/s13347-019-00354-x.
- [34] M. Mesarčík, S. Soľárová, J. Podroužek, M. Bielikova, Stance on The Proposal for a Regulation Laying Down Harmonised Rules on Artificial Intelligence – Artificial Intelligence Act, preprint, SocArXiv, 2022. doi:10.31235/osf.io/yzfg8.