Felzmann, H., Fosch-Villaronga, E., Lutz, C., & Tamo-Larrieux, A. (2019). Robots and Transparency: The Multiple Dimensions of Transparency in the Context of Robot Technologies. *IEEE Robotics & Automation Magazine, 26*(2), 71-78. doi:10.1109/MRA.2019.2904644

Robots and Transparency

The multiple dimensions of transparency in the context of robot technologies

Heike Felzmann
NUI Galway, Galway, Ireland
heike.felzmann@nuigalway.ie

Eduard Fosch-Villaronga
University of Leiden, Leiden, Netherlands
eduard.fosch@gmail.com

Christoph Lutz
BI Norwegian Business School, Oslo, Norway
christoph.lutz@bi.no

Aurelia Tamo-Larrieux
University of Zurich, Zurich, Switzerland
aurelia.tamo-larrieux@uzh.ch

The second author is the corresponding author.

The article was published in *IEEE Robotics & Automation Magazine*. Please cite as follows:

Felzmann, H., Fosch-Villaronga, E., Lutz, C., & Tamo-Larrieux, A. (2019). Robots and Transparency: The Multiple Dimensions of Transparency in the Context of Robot Technologies. *IEEE Robotics & Automation Magazine, 26*(2), 71-28. DOI: 10.1109/MRA.2019.2904644

Robots and Transparency

The multiple dimensions of transparency in the context of robot technologies

Introduction

Transparency is often seen as a means to provide accountability and to show that something is done with due diligence. This approach to transparency regards it as a remedy to hidden (potentially malevolent) practices. We therefore require transparency from manufacturers of products and services. While the outcry for more transparency often occurs in response to a particular scandal (e.g., the various controversies surrounding Facebook in 2018), the General Data Protection Regulation (GDPR) includes transparency as a proactive requirement for information technologies that process personal data.

Since the functioning of many robots depends on processing personal data, the GDPR becomes applicable[1] and the challenge of applying the principle of transparency arises.

When users interact with a robot, it might not be clear that the robot is not the only relevant entity in the interaction; third parties also provide significant aspects of its functioning (Fosch-Villaronga et al., 2018). Complex cyber-physical environments challenge a straightforward notion of transparency, especially when autonomous social robots make decisions and perform actions of social and moral significance for users.[2]

In this contribution, we investigate what achieving transparency could mean for the field of robotics, keeping in mind that the operation of autonomous systems should be transparent to a broad range of stakeholders.[3] We first introduce the concept of transparency and outline different expectations regarding transparency. Second, we provide an overview of the general ethical underpinnings of transparency, which connect to autonomy and informed consent. Third, we outline the transparency requirement of the GDPR, which demands from data controllers easily accessible and understandable information, as well as communication of how personal data is

[1] Assuming that data processing falls under the territorial and material scope of the GDPR.
[2] See https://explainableroboticsystems.wordpress.com/
[3] See http://sites.ieee.org/sagroups-7001/.

being processed (Rec. 58 of the GDPR). Fourth, we summarize current findings of human-robot interaction (HRI) research in the field of transparency to show how transparency works in practice. Finally, we conclude by proposing a checklist for designers that outlines a step-by-step guide which may help robot developers implement the transparency requirement set by the law. Future work will address the application to individual use cases.

Different Expectations of Transparency

Transparency usually refers to things and concepts that are easy to perceive or detect. In the context of computing, however, it refers counterintuitively to processes or interfaces that function without a user being aware of them (Oxford Dictionary, 2018). This latter understanding of transparency contrasts with the GDPR, which demands transparency for information technologies in the sense of making data processing explicit to the user. Such standards could entail a barrier to the deployment of robotic systems that process personal data in Europe or of European citizens.

The GDPR is intended to be a technology-neutral piece of legislation, meaning that no specific technology should be the target of the law. Instead, it should apply to all possible technologies at large. The GDPR’s strength lies in providing general legal requirements across technologies. However, the its lack of recognition for specific technologies and context factors risks neglecting crucial elements in protecting users' data-related rights. This challenge also arises in the determination of the requirements of transparency, which will need to be molded according to the characteristics of the technology - in this case, a robot. Moreover, complex technologies such as artificial intelligence (AI) and robotic systems raise particular challenges not only due to their information processing nature and contexts of use, but also due to the multitude of stakeholders potentially affected by the transparency requirements.

Distinct from many other information technologies, the end user in a human-robot interaction context is not the only user who is engaging with the system. The broader context of robot deployment, involving different roles and responsibilities among various stakeholders, demands a comprehensive understanding of transparency (IEEE P7001). A typical robot ecosystem, in the healthcare sector for instance, would involve the healthcare organization’s management who initially decided to deploy the robot, healthcare staff who implement the robot in therapies or daily care, family members who make decisions about their relative’s engagement with the robot, the end user; the robot's developer, and infrastructure providers (Fosch-Villaronga et al., 2018; Lutz & Tamo, 2018).

Weller's (2017) investigation into the roles and types of transparency in the context of human intelligibility of robotic systems explores this issue. However, further differentiation among the various stakeholders in the field of assistive robots awaits development:

Transparency in the context of robotics and AI	
For a...	**Transparency serves to...**
Developer	Understand whether their system is working properly, in order to identify and remove errors from the system or improve it
User	Provide a sense for what the system is doing and why, to enable intelligibility of future unpredicted actions circumstances and build a sense of trust in the technology
	Understand why one particular decision was reached
	Allow a check that the system worked appropriately
	Enable meaningful challenge (e.g. credit approval or criminal sentencing)
Society broadly	Understand and become comfortable with the strengths and limitations of the system
	Overcome a reasonable fear of the unknown
Expert/Regulator	Provide the ability to audit a prediction or decision trail in detail, particularly (un)intended harmful actions, e.g. a crash by an autonomous car.
Deployer	Make a user feel comfortable with a prediction or decision, so that they keep using the system
	Lead a user into some action or behavior, e.g. Amazon might recommend a product while providing an explanation in order that the user then clicks through to make a purchase

Table 1. Transparency expectations for different stakeholders (adapted from Weller, 2017).

Different stakeholders have different roles, information needs, background knowledge and abilities. Accordingly, the transparency requirement needs to be tailored to the types of users (in light of their roles, responsibilities, and interests) and to their level of capacity and vulnerability.

Assistive robotics, as a field, frequently targets vulnerable users such as elderly individuals with dementia or children with autism. Even for non-vulnerable users, understanding the information provided about a specific robot is a non-trivial task. The inclusion of vulnerable users creates particular challenges for transparency, since strategies regarding transparency which work for non-vulnerable users may not work for vulnerable populations. Evidence about effective strategies which do justice to the specific needs of vulnerable populations still needs to be gathered.

One particular challenge is that, as HRI research has shown, users intuitively relate to robots as if they were living beings (Jeong et al., 2018). The information processing capacities of such robots may not be apparent to users. In particular, it may not be evident to users that their data may be collected and analyzed. Accordingly, without prompting, users may not expect that such information processing can raise concerns and require explanation.

Ethics of Transparency

The ethical need for transparency can be understood as closely linked to the value of human autonomy. Autonomy requires that users have the opportunity to interact with their environment on their own terms. Transparency gives users of technologies an understanding of what will be happening with their data; having such information facilitates an informed consent process that allows them to make meaningful choices about their use of these technologies. According to Beauchamp and Childress (2012), there are several elements of consent that need to be realized for an autonomy-respecting informed consent process. Users must not be coerced into consenting; they must have the capacity to consent; information that is relevant to understanding the nature and potential impact of the technology, including practical implications, risks, costs, and benefits, needs to be disclosed understandably to the user; users must be given opportunities to achieve understanding; and users have to authorize interventions actively.

Given the complexity and opacity of information processing in information technologies, achieving meaningful informed consent to information technologies is challenging. Achieving consent *merely* through notice and consent, via the simple acceptance of Terms and Conditions for general processing purposes that are designed to meet the legal minimum of disclosure, is ethically insufficient according to these criteria. Not only is there often little choice available to users, but frequently such terms and conditions remain vague and unclear on those aspects that users would consider essential to know while providing lengthy and detailed information that is not designed to enable users' engagement. In short, they often do not give the users understandable information on what the system is doing and how it potentially may affect them at an appropriate level of specificity that is informative without being overly demanding.

Demands for transparency need to take into account what information would be of value and interest to potential users to underpin meaningful decision-making and to help them engage with it. Importantly, a lack of transparency may affect the perceived trustworthiness of those responsible for the provision of such information. While the ethical literature on transparency emphasizes the complexity of potential positive and negative effects of transparency (Heald 2006, O'Neill 2002), it is generally acknowledged that transparency conveys at least the willingness to be open to scrutiny and to be held accountable. It indicates trustworthiness to potential users, even if in practice transparency may not always result in increased accountability or more significant experience of trust among recipients of transparency information (O'Neill 2002).

Transparency and the Law

While from an ethical point of view transparency can be conceived as a way to achieve autonomy via informed consent and to convey trustworthiness, the legal field has specified the conditions of how to obtain consent. It has defined how data controllers have to inform data subjects about the processing of personal data in a transparent manner. Today, transparency is a core principle enshrined in Art. 5 (1)(a) of the GDPR stating that personal data must be "processed lawfully, fairly and in a transparent manner in relation to the data subject."

Not only has transparency become a principle of data protection, but the term has also been specified within the GDPR. One aspect of transparency is the broadened information duties of data controllers (defined in Arts. 13 and 14 of the GDPR). Data controllers must at least inform data subjects who they are, what quantity and quality of personal data they process, and when, for how long, why and for what purposes they handle said data. Recitals 39 and 58 of the GDPR provide some guidance on how to implement transparency in different systems. In particular, data controllers must inform their customers about their data processing practices through concise, easily accessible, easy to understand and clear and plain language (where appropriate with visualization). The information must be provided in writing or, where necessary, by electronic means, and the information must come in an intelligible and easily accessible form (in particular when data controllers target children; see Art. 12 of the GDPR).

Apart from prospective transparency, where a data subject is informed about the data processing beforehand, transparency requires retrospective transparency, meaning the ability to follow the data processing step-by-step, for audit purposes for example (Paal & Pauly, 2017). Article 22 of the GDPR gives the right to data subjects not to be subject to a decision based solely on automated processing that significantly affects them. Moreover, Recital 71 of the GDPR gives the subject the right to obtain human intervention, to express his or her point of view and to obtain an explanation of the decision.

Although the GDPR contains specific passages that explain how complex data collection and transformation processes should be made accessible to the data subjects (Arts. 12-22, 34 GDPR), these descriptions are still subject to interpretation. Additional specification and evidence on effective strategies are required to support engineers and HRI experts to realize the transparency requirements from the legal and also ethical point of view. In particular, improvements are necessary to the current practice of reducing consent to the presentation of complex information to users followed by a simple tick box. New strategies are needed concerning adaptation to users' information needs and differentiated preferences, to allow the transparency requirement to contribute to the facilitation of meaningful choice. In the next section, we, therefore, discuss the empirical evidence on transparency and transparency effects found in HRI research.

HRI Transparency Realization

Sociological and psychological studies have explored transparency effects (Kim & Hinds, 2006) and expectations in the context of robotics (Berkelaar 2014). This research investigates the user effects of information provision, in the sense of explanations of how and why the robot does what it does. It shows that user perceptions of and attitudes towards transparency differ substantially depending on the technologies and services investigated, the tasks given and other contextual factors. In some cases, robot transparency has limited impact, for example on attributions of credit and blame (Kim & Hinds, 2006), or on assessments of competence (Kwon, Ferguson & Knepper, 2018). In other cases, robot transparency leads to poorer perceptions of the robot (Petisca, Dias & Paiva, 2015).

How transparency is enacted can lead to different outcomes (Wang et al., 2018). The situational importance of transparency has also been pointed out, suggesting that technology should be transparent and able to explain itself in critical states. However, this may not be as advantageous when everything is running as usual (Huang et al., 2018).

General research on transparency in intelligent systems shows similarly mixed results. It does not permit drawing generalized design recommendations for transparency and explainability (Felzmann et al., 2018). While the requirement for transparency has strong ethical and rights-based support and is now a legal requirement in the GDPR, these mixed results indicate that from a purely pragmatic and user-centered perspective an increase in transparency is not always clearly desirable. Accordingly, from an industry point-of-view, investment in transparency by developers could be costly, with unclear effects and benefits, and there may even be a risk that at times transparency could backfire by decreasing user trust (Eiband et al., 2018).

Fischer (2018), for example, points to the detrimental effects transparency might have in the context of assistive robots. If such robots are too transparent about their information processing capacities, such transparency might impede natural and seamless human-robot interaction. The desirability of transparency might depend heavily on the application domain and the particular type of human-robot interaction. Accordingly, when engaging with the demand for transparency any transparency measures need to be designed with due regard to the specific characteristics of human-robot interaction in that use context.

What transparency means for the field of robotics is still underexplored. This is despite recently intensifying research efforts in the form of a dedicated IEEE group on the transparency of autonomous systems (IEEE-P7001)[4] or a recent workshop on explainable robotic systems (see table below for a summary of the contributions):[5]

[4] *See* http://sites.ieee.org/sagroups-7001/

[5] *See* https://explainableroboticsystems.wordpress.com/

Study	Topic of study	Transparency outcomes investigated	Contribution	Key findings
Holder and Marge	Role of intent in HRI to improve explainability	Robot design to support human understanding	Taxonomy for intent in HRI in a military context	Intent statement needs additional background information; context-dependent.
Hellström and Bensch	Meaning of understandability and its formalization	Human understanding of robot's behavior; Robot design to support human understanding	Model of interaction for understanding	Conceptual separation of information to be communicated from the means of communication.
Gong and Zhang	Reciprocal understanding between robot and human	Human understanding of robot's behavior; Robot design to support human understanding	Formulation of human interpreting robot's actions as a labeling process	Label process: check if the action is explainable; if not, search for optimal timing and content to signal its intentions.
Thellman and Ziemke	Prediction and explanation of robot actions	Human understanding of robot's behavior	Observation - Ascription - Inference Model	People's predictions and explanations are constrained by theoretical presuppositions regarding robots' role-specific goals, morphological and environmental constraints, sensory capabilities and situated perspective on the world.
Avrunin, Rosenthal and Simmons	Appropriate level of detail in explanations	Robot design to support human understanding	Coverage-based explanation reduction algorithm	Robots can create reduced explanations that approximate the true robot policy but are more memorable or understandable to a human.
Bekele, Lawson, Horne and Khemlani	Human-level explanatory biases for explainability (in person re-identification)	Robot design to support human understanding	Creation of a multi-attribute residual network for explanatory re-identification	Deep learning systems capable of mimicking human explanatory biases can provide meaningful and interpretable explanations of their own internal operations.
Huang, Bhatia, Abbeel and Dragan	Establishing trust via critical states	Robot design to support human understanding	Computation and use of critical states	The end-user does not need to know what the robot would do in all states; the robot action matters only in critical states. Showing end-users how the robot acts in critical states gives them a better understanding of what it has learned, and enables them to decide in which situations they can trust the robot.

Korpan and Epstein	Natural explanations of robot navigation plans	Robot design to support human understanding	Method that compares the perspectives of an autonomous robot and a person when they plan a path for navigation	By explaining the context of the robot's most recent action and its long-range perspective, robots can produce meaningful, human-friendly explanations quickly in natural language.
Chiyah Garcia, Robb, Liu, Laskov, Patron and Hastie	Natural language interface for remote autonomy explanations	Robot design to support human understanding	Model to allow 'on-demand' queries for status and explanations of behaviour	If the expert is from the same pool of end-users (i.e. operators), explanations are likely to align with their mental models and assumptions about the system.
Ghayoumi	Cognitive-based emotion model for social robots	Robot design to support human understanding	Architecture for emotion and social robot model incorporating a knowledge-based system and the cognitive appraisal	Applying this model in healthcare allows the robot to report emergency and non-emergency cases to the experts and to communicate with patients with a proper expressive face.
Erel , Hoffman and Zuckerman	Interpreting non-anthropomorphic robots' social gestures	Human understanding of robot's behavior	One study had gestures designed by performance artists and movement experts, and in the other gestures were purely mechanistic. Both suggested specific loci of focus for gesture design.	Limited gestures of non-anthropomorphic robots can be consistently interpreted as social interaction cues. Specific movement components (e.g. vertical axis), have more profound effects on the emotion perceived from the robot.
Gutierrez, Chu, Short, Niekum and Thomaz	Learning task decompositions for the construction of generalizable primitives and task models	Robot design to understand humans	Keyframe demonstrations provide an intuitive interface for teaching robots to perform tasks.	If the robot's task model conforms with people's intuitions of task decomposition, the robot's behavior will be more readily understandable and aid in the update of the robot's task models.
Hastie, Lohan, Chantler, Robb, Petrick, Lane, Ramamoorthy and Vijayakumar	Increase transparency to increase trust in robot's actions	Robot design to support human understanding	Understanding the user's mental model of the systems is key to providing just the right information at the right time.	Balancing information presentation and explanations with the user's actual needs and cognitive load in a dynamic, fast moving environment will be essential to successfully deploy robotics and AI offshore robotics and elsewhere
Racca and Kyrki	Challenges of transparency for learning robots	Robot design to support human understanding	Transparency has enormous potential to improve the reliability and	Exposure of the analytical and the task models is a challenge (especially in black-box learning

			performance of learning robots.	techniques). The modelling of the human state is important, but research is needed to develop methods to reliably detect the human state and models to aid the decision making of learning systems during training.
Fischer	Transparency in human-robot interactions and its desirability	Negative outcomes and pitfalls of transparency such as misunderstandings and unnatural interaction flow	Transparency can enhance explainability and predictability but is still not desirable in HRI due to three main reasons.	Three downsides of transparency in HRI: 1) Transparency can destroy the illusion that robots are similar to humans and thus inhibit seamless HRI. 2) Transparency about robots' high-level capabilities will create assumptions that the robot also has lower-level capabilities, which might not necessarily be true and create misunderstandings. 3) Transparency about low-level capabilities of a robot results in unwanted inferences about the robot's capabilities.
Broz, Ghosh, Keller, McKenna, Rajendran, Aylett	Design of a robot for behavioral skills training to tackle autism spectrum disorder in workplaces	Robot design to support understanding of social signals	System architecture to enable behavioral skills training through a robot for individuals with ASD.	A solid system architecture for behavioral skills training for individuals with ASD involves user-centered elements and modelling of workplaces.
Kwon, Ferguson, Mann, Knepper	Implicit impressions of robot competence	Transparency understood in terms of warnings and investigation of their effect on implicit impressions	Measurement of implicit impressions, rather than explicit impressions, through the affect misattribution procedure	Transparency in the form of a warning did not change competence judgment in the short run but might have an effect in the longer run.
Marmpena, Dahl, Lim	Human perception of robot expressions and emotions and how they are influenced by transparency	Empathic and anthropomorphic weight attributed to the robot by a human	Linking transparency and emotional perception: Does transparency affect how robots are perceived emotionally?	No empirical findings on the transparency-emotions link but solid taxonomy of emotional cues of robots and their assessment in a two-dimensional space (valence, arousal).
Wang, Barnes, Pynadath, Hill	Automatic explanations of complex decisions	Trust in a robot, as affected through explanations; Team performance	Contrasting two types of explanation: numeric information on uncertainty and textual information about sensor	Numeric information explanations was seen as more open and competent but textual explanations were seen as more predictable.

			readings	
Oliveira, Arriaga, Correia, Paiva	Display of stereotype content, especially warmth and competence, by robots and their effect	Emotions; Behavior; Intention for future interaction with the robot	Applying the topic of stereotypes in a group interaction setting	No empirical findings at this point.

Table 2. HRI research on the transparency principle in technical terms[6]

From this table, we can see that HRI research on transparency has been focusing primarily on the explainability of the systems, either from the robot (intelligibility) or the user perspective (understandability). Their conclusions vary, with some studies showing no relevant findings and some identifying the downsides of transparency. Such disadvantages include the inhibition of seamless HRI and the creation of misunderstandings and unwanted inferences about a robot's capabilities (Fischer, 2018). These findings are in line with other results which highlight that transparency has technical limitations (Koops & Leenes, 2014), that it can create false binaries and be harmful, and that it could be used to prioritize seeing over understanding (Ananny & Crawford, 2018).

Checklist for Implementing Transparency in Robot Development

The following checklist translates the general considerations outlined above into a step-by-step guide for robot developers. Its goal is to provide user-centered guidance on how to design for transparency.

To implement transparency in a given AI environment, we suggest to:
I. Identify general transparency obligations;
II. Identify the different transparency needs and expectations of the involved stakeholders;
III. Translate the transparency requirements to the level of understanding of a target group;

[6] Supra note 2.

IV. Conduct user testing concerning some transparency related parameters;

V. Guide users concerning available transparency functions.

General task	Steps	Sources
I. Identify general transparency obligations	Create a data flow map that identifies the collection of personal data (who collects what and how?), transfer (how is the data transferred; where to; which additional considerations regarding transfer, such as security and contracts, need to be taken into account?), processing (who processes the data and what safeguards are in place?), storage (where is the data stored internally or externally? For how long? How secure is it?), erasure (when is the data erased?)	Consult with experts in GDPR
II. Differentiate transparency needs of stakeholders	Identify likely contexts of use (e.g., domestic, hospital, nursing home, rehabilitation)	Refer to design use cases
	Identify likely core user groups (e.g., health care manager, nurse, physiotherapist, patient, informal carer)	Consult with experts familiar with context of use
	Identify likely relevant background knowledge of core user groups (educational background, IT expertise level, familiarity with robots, familiarity with healthcare processes, familiarity with data protection)	Consult with experts familiar with context of use Consult with stakeholder representatives
	Identify likely transparency information needs and interests of core user groups (e.g., general functionality of robot, data security, health record compatibility, user privacy)	Consult with experts familiar with context of use Consult with stakeholder representatives
	Identify stakeholder characteristics relevant for transparency communication (e.g. education level, physical or sensory impairments, cognitive limitations)	Consult with experts familiar with context of use Consult with stakeholder representatives
III. Adapt transparency communication to user groups	Provide differentiated access to transparency communications tailored to user groups	HRI expertise Explain the General Data Protection Regulation (GDPR) to different target audiences
	Identify suitable modality for transparency communication for user groups given user characteristics and context of	HRI expertise

	use	Consult with experts familiar with context of use Consult with stakeholder representatives
	Ensure easy accessibility of transparency communication for users	HRI expertise Consult with stakeholder representatives
IV. User testing	Track use of transparency functions	HRI expertise User behaviour
	Observe user behaviour in response to transparency communication	HRI expertise User behaviour
	Evaluate effectiveness of transparency communication	HRI expertise Assess users' understanding
	Evaluate user satisfaction with transparency communication	HRI expertise Consult users
	Adapt and integrate users' feedback to increase positive results	HRI expertise Results of user testing process
V. Provide guidance for users on transparency functions	Provide understandable and easily accessible information on transparency functions for different user groups	Arts. 12-15, 22 GDPR
	Allow options and facilitate choice with regard to users' engagement with transparency communications	HRI expertise Informed consent requirements

Table 3 Tabular checklist for implementing transparency in robot development

Conclusions

This article has offered an overview of the transparency requirement and has explained the main dimensions of the transparency principle in the context of robotics. The implementation of legal transparency requirements requires interdisciplinary collaboration between legal, social science and technology experts to avoid overlooking ethical and societal aspects and to create an evidence base that will be essential for engineers and industry in designing transparency measures that are effective and meet legal requirements (Ausloos et al., 2018; Miller, 2017).

Future HRI research on transparency should investigate the situational and contextual value of transparency, user awareness and needs, and design-related questions such as how transparency can best be implemented in assistive robots. In addition to experiments, HRI research could use ethnographic and observational approaches, interface studies and reverse engineering. Doing so will require applying the findings of this article to concrete use cases, thereby providing more practical guidance on how to implement transparency in a given context. The checklist in the previous section could offer direction in developing such use cases in practice.

Takeaway Messages

- Transparency is an ethical requirement based on the value of autonomy and is essential for meaningful informed consent. Data subjects must be informed about how controllers process their data in a concise, easily accessible, and understandable manner.
- Transparency is also a legal requirement, binding in the EU and for the processing of personal data of EU citizens or in the EU territory. Engineers whose products might be used in the EU need to become familiar with and abide by the transparency requirements of the GDPR.
- Transparency in HRI is underexplored, and the findings on the user value of transparency are mixed. More research is needed to understand what constitutes effective versus counterproductive implementations of transparency.
- More interdisciplinary collaboration is needed to translate the relatively abstract and technology-agnostic GDPR into the design of robot technology and to understand how technical measures for transparency could be implemented to achieve compliance with the law.
- A procedural checklist for designers might be a suitable instrument to guide the design process concerning meeting transparency requirements.

References

Ananny, M., & Crawford, K. (2018). Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability. *New Media & Society, 20*(3), 973-989.

Ausloos, J., Dewitte, P., Geerts, D., Valcke, P., & Zaman, B. (2018). Algorithmic Transparency and Accountability in Practice. In *Proceedings of the 36th Annual ACM Conference on Human Factors in Computing Systems* (CHI’18). ACM.

Beauchamp, T., & Childress, J. (2012) *Principles of Biomedical Ethics*, 7th ed., New York: Oxford University Press.

Berkelaar, B. L. (2014). Cybervetting, online information, and personnel selection: New transparency expectations and the emergence of a digital social contract. *Management Communication Quarterly*, *28*(4), 479-506.

Cas, J., Strauss, S., Amicelle, A., Ball, K., Halliman, D., Friedewald, M., & Szekely, I. (2015). Social and economic costs of surveillance. In Wright, D., & Kreissl, R. (Eds.). (2015). *Surveillance in Europe*. Routledge.

Eiband, M., Schneider, H., Bilandzic, M., Fazekas-Con, J., Haug, M., & Hussmann, H. (2018, March). Bringing Transparency Design into Practice. In *23rd International Conference on Intelligent User Interfaces* (pp. 211-223). ACM.

Felzmann, H., Fosch Villaronga, E., Lutz, C., & Tamò-Larrieux, A. (2018, October). How transparent is your AI? Ethical, legal, and societal issues of the transparency principle in cyber-physical systems. In *Amsterdam Privacy Conference 2018*, working paper.

Fosch-Villaronga, E., Felzmann, H., Ramos-Montero, M., & Mahler, T. (2018, October). Cloud services for robotic nurses? Assessing legal and ethical issues in the use of cloud services for healthcare robots. In *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS),* 290-296.

Fischer, K. (2018, March). When Transparent does not Mean Explainable. In *Explainable Robotic Systems - Workshop in conjunction with HRI 2018*. ACM/IEEE. https://explainableroboticsystems.files.wordpress.com/2018/03/1-fischer-being-honest-about-transparency-final.pdf

Heald, D. (2006). Transparency as an Instrumental Value. In Hood, C. & Heald, D. (eds.) *Transparency: The Key to Better Governance*. *Proceedings of the British Academy 135* (pp. 59-73). Oxford: Oxford University Press.

Huang, S. H., Bhatia, K., Abbeel, P., & Dragan, A. D. (2018, March). Establishing (Appropriate) Trust via Critical States. In *Explainable Robotic Systems - Workshop in conjunction with HRI 2018*. ACM/IEEE. https://explainableroboticsystems.files.wordpress.com/2018/03/2-huang-et-al-establishingappropriatetrust_hri2018workshop-final.pdf

Jeong, S., Breazeal, C., Logan, D., & Weinstock, P. (2018, April). Huggable: The Impact of Embodiment on Promoting Socio-emotional Interactions for Young Pediatric Inpatients. In

Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (p. 495). ACM.

Kim, T., & Hinds, P. (2006). Who should I blame? Effects of autonomy and transparency on attributions in human-robot interaction. In *The 15th IEEE International Symposium on Robot and Human Interactive Communication, 2006. ROMAN 2006.*, pp. 80-85.

Koops, B. J., & Leenes, R. (2014). Privacy regulation cannot be hardcoded. A critical comment on the ‘privacy by design’ provision in data-protection law. *International Review of Law, Computers & Technology*, *28*(2), 159-171.

Kwon, M., Ferguson, M., & Knepper, R. (2018, March). An Exploration of Implicit Attitudes Towards Robots Using Implicit Measures. In *Explainable Robotic Systems - Workshop in conjunction with HRI 2018*. ACM/IEEE. https://explainableroboticsystems.files.wordpress.com/2018/03/1-kwon-et-al-an-exploration-of-implicit-attitudes-towards-robots-final.pdf

Lutz, C., & Tamò, A. (2018). Communicating with robots: ANTalyzing the interaction between healthcare robots and humans with regards to privacy. In A. Guzman (Ed.), *Human-Machine Communication: Rethinking Communication, Technology, and Ourselves* (pp. 145–165). Bern: Peter Lang.

Miller, T. (2017). Explanation in artificial intelligence: insights from the social sciences. *arXiv preprint arXiv:1706.07269*.

O’Neill, O. (2002). *A Question of Trust. The BBC Reith Lectures 2002*. Cambridge: Cambridge University Press.

Paal, P. & Pauly, D., *Datenschutzgrundverordnung Kommentar*, C.H. Beck, 2017 (cited Paal/Pauly/author, Art., N.), Paal/Pauly/Frenzel, Art. 5, N. 21.

Petisca, S., Dias, J., & Paiva, A. (2015, October). More social and emotional behaviour may lead to poorer perceptions of a social robot. In *International Conference on Social Robotics* (pp. 522-531). Springer, Cham.

Transparent definition, Oxford Dictionary, 2018, last accessed December 21, 2018, https://en.oxforddictionaries.com/definition/transparent

Wang, N., Pynadath, D. V., Barnes, M. J., & Hill, S. G. (2018, March). Comparing Two Automatically Generated Explanations on the Perception of a Robot Teammate. In *Explainable Robotic Systems - Workshop in conjunction with HRI 2018*. ACM/IEEE. https://explainableroboticsystems.files.wordpress.com/2018/03/1-wang-et-al-comparing-two-automatically-generated-explanations-on-the-perception-of-a-robot-teammate-final.pdf

Weller, A. (2017). Challenges for transparency. arXiv preprint arXiv:1708.01870.