Governmental transparency in the era of artificial intelligence

van Engers, T.M.; de Vries, D.M.

DOI
10.3233/FAIA190304
Publication date
2019
Document Version
Final published version
Published in
Legal Knowledge and Information Systems

Legal Knowledge and Information Systems
M. Araszkiewicz and V. Rodríguez-Doncel (Eds.)

doi:10.3233/FAIA190304

Governmental Transparency in the Era of Artificial Intelligence

Tom M. van ENGERS [a] and Dennis M. de VRIES [b]
[a] *Leibniz Center for Law, University of Amsterdam*
[b] *Informatics Institute, University of Amsterdam*

Abstract. In the last years governments started to adapt new types of Artificial Intelligence (AI), particularly sub-symbolic data-driven AI, after having used more traditional types of AI since the mid-eighties of past century. The models generated by such sub-symbolic AI technologies, such as machine learning and deep learning are generally hard to understand, even by AI-experts. In many use contexts it is essential though that organisations that apply AI in their decision-making processes produce decisions that are explainable, transparent and comply with the rules set by law. This study is focused on the current developments of AI within governments and it aims to provide citizens with a good motivation of (partly) automated decisions. For this study a framework to assess the quality of explanations of legal decisions by public administrations was developed. It was found that communication with the citizen can be improved by providing a more interactive way to explain those decisions. Citizens could be offered more insights into the specific components of the decision made, the calculations applied and sources of law that contain the rules underlying the decision-making process.

Keywords. Artificial Intelligence, XAI, Transparency, Explanations, Government

1. Introduction

All over the world, governments have started to adopt new Artificial Intelligence (AI) technologies to improve the efficiency and effectiveness of public administration. Automated Decision-making Systems (ADS), for example, can help governmental agencies with various tasks such as deciding on tax assessment and student finance. In these application domains, the citizens' stakes are high. Therefore, it is of great importance that those (partly) automated decision systems are transparent on their reasoning mechanisms and carefully explain their decisions. This study focus on the improvement of explanations of governmental agencies' communications regarding (partly) automated decisions.

Over the years, a substantial number of studies have been published on opening the black boxes of artificial intelligence [1,2]. Only a few studies suggest procedures on how decisions made by artificial intelligence should be explained in a proper manner [3,4]. Besides some studies of prestigious consulting firms, academic research on how to improve explainability and transparency of automated decision-making systems in governments is laying back [5,6,4].

1.1. Governments adopting Artificial Intelligence

AI-based systems have been used by governments for decision-making purposes since the mid-eighties of the 20th century. Most of these systems were and still are rule-based systems, with the 'rules' elicited from (legal) experts [7,8]. The primary purpose for the government to invest in AI systems is to provide better services and improve the effectiveness and efficiency of public administration [9], e.g. in application domains such as optimising traffic flows [10], tax assessment [11], assessing visa applications [12] and crime prevention [13].

Many of those AI systems are used for decision-support and use a rule-based reasoning mechanism using determined rules to come to a specific decision [14,15,16]. With the increase of computer power, sheer unlimited data availability, and the boost of the internet, new AI-technologies have emerged and become popular. Particularly data-driven, sub-symbolic AI technologies, that are known under various names, such as machine learning, deep learning, and neural nets, became popular again in the 21st century [17]. Since the end of the nineteen-nineties, e.g. machine vision methods were used for various pattern recognition task including that of handwritten addresses from envelopes [18]. Contrary to symbolic AI that is typically connected to deductive approaches, sub-symbolic AI is typically connected to inductive approaches. This focuses on learning systematic patterns from the data, and then apply those learned patterns on new input determining the appropriate output [19]. The use of AI in fields where the stakes are high, however come with some worries.

1.2. Challenges of AI

Ever since the introduction of AI-technologies people have feared the lack of human touch and empathy, the lack of transparency and unfairness when smart AI-components replace the human in the loop [13].

The COMPAS system developed for predicting the likelihood of recidivism of criminals for example became infamous for its bias against Afro-Americans [20]. Such bias against a specific group within society could easily lead to more segregation and then decreasing opportunities for that specific group and as a result produce a self-fulfilling prophecy [21]. In order to be able to trust organisations in taking (legally) justified decisions, these decisions when produced by AI applications need to be explained and argued for in such a way that the persons subjected to those decisions at least understand what the decision is based upon.

The main challenge that is addressed in this study, is providing insight into the reasoning mechanisms of AI-algorithms for citizens. This is needed in order to check their correctness, fairness, normative compliance and sensitivity to potential biases in their judgements.

1.3. A Renewed Interest in Explainable AI

Data-driven AI-technologies that 'learn' from data, are vulnerable for bias and the models induced from the data are generally hard to understand even for experts. This is even getting worse if the AI-algorithms keep 'learning', i.e. adapting their models, while being used. Because of the lack of transparency it is hard to 'trust' those AI algorithms hence

the recent demand 'Responsible AI', a term that includes explainable AI (XAI) and fairness, but is somewhat ambiguous as responsibility could refer both to the AI-technology itself as well to the developers and organisations exploiting these technologies. Holding AI responsible for anything, i.e. attributing some kind of personality to it, would bring us back to dark ages, so let's keep the human stakeholders responsible, like we do with all other artefacts! The call for XAI has become louder after a few scandals, and it is needless to say that specifically governmental agencies that deploy AI to support their tasks have to meet the traditional government requirements for explainability, transparency, accountability and auditability [6].

In order to try to protect some essential social fundamental values, The Dutch Council of State (advisory body to the government) published a report on the influence of new technologies on constitutional relations [22]. The Council advises the government to pay closer attention to the motivation of their automated decisions. They demand that it should be clear which decision-rules (algorithms) and data the governmental authority used for a specific decision. Furthermore, it should be made clear which data is taken from other governmental authorities. Explainability in Europe further pushed by the General Data Protection Regulation (GDPR) [23] that is applicable since May 25th 2018 in all European member states. The GDPR includes Article 22 on *'Automated individual decision-making, including profiling'* forcing organisations to be transparent about the decision-making process of their algorithms.

2. Literature Review

As stated in the previous section the need for explainable AI is not an entirely new topic; it has been addressed in many reports and academic papers and is discussed at plenty of conferences such as those of ACM's CHI community [24,25,26]. The increased popularity of sub-symbolic AI has just put the topic back on the agenda again.

2.1. Why Explanations Matter

One key part of XAI is the explanation itself, The Oxford English Dictionary defines EXPLANATION as: 1) *'A statement or account that makes something clear'* and 2) *'A reason or justification given for an action or belief'* [27]. Therefore, an explanation mainly aims to answer the *how* and *why* questions, which can be useful to clarify or justify the behaviour of an AI agent respectively [28]. Within our daily lives, explanations are used by humans to share information and in order to better understand each other. Therefore, explanations lead to better acceptations about specific statements [29]. Over the years, studies from various disciplines suggest that providing explanations on the mechanisms of AI systems improve the acceptance of the user in regards to the decisions, conclusions and recommendations of those systems [30,28,31,24,32]. As a result, systems that provide better explanations on their reasoning will improve the acceptance by citizens in the outcome of those systems. Other studies suggest that explanations from AI systems help to acquire or maintain trust from the user in the accuracy of those systems [33,18,34,26,35].

2.2. Explaining Good Explanation

Research into explanations has a long history. Early examples of research in this subject include topics such as logic, causality and human discourse [36,37]. Related work can be found in various areas such as philosophy and psychology. Based on earlier studies, an evaluative framework that enables to evaluate the quality of XAI was developed. In literature, several criteria have been described that can be used to determine the satisfaction of an explanation. The framework presented in this paper includes those criteria that are most frequently mentioned and extensively discussed in the field of cognitive sciences and AI literature. Below we'll present six primary quality criteria for explanations and references to preliminary research on these criteria:

The first quality criterion for explanation is called **EXTERNAL COHERENCE** [38]. Some researchers suggest that the likelihood of acceptance of a decision increases when the explanation is consistent with one's former beliefs [39]. This means that explanations should be compatible with what the reader already knows in the specific context at hand [40].

The second quality criterion is **INTERNAL COHERENCE**. This concept points out the sense of how good the several elements of an explanation fit together [40]. There should be a logical relation between propositions to improve the completeness of the explanation and improve the perceived understanding [41,38].

The third quality criterion is **SIMPLICITY**. Two studies tested the theory of Thagard on Explanatory Coherence [38] and found that people preferred explanations that invoke fewer causes [42,43].

The fourth quality criterion is **ARTICULATION**. One particular study presents several linguistic markers that examine clear articulation of a letter [40]. One of the three elements is the number of words used in the explanation. Another one is the average word length of the statement. The median word frequency of the text can also be used as an indicator [40].

The fifth quality criterion is **CONTRASTIVENESS**. This criterion expresses the clarity of the arguments that explain why event P happened rather than event Q [39,44]. This specific factor also emphasises questions such as what would happen when a particular condition in the process is changed [45].

Finally, some research mentions that the user's satisfaction with an explanation might increase when the possibility for **INTERACTION** between the explainer and explainee is provided [46]. What is needed for an explanation also depends on what the explainee already knows and specifically; still wants to know [47]. This criterion proposes new opportunities in the field of Human-Computer Interaction (HCI) [39]. By providing interactive dialogue, the satisfaction of the user might increase.

Several criteria for the layout of a letter (used fonts, use of color, etc.) might influence the receiver as well. This research was scoped to mainly focuses on the structure of a letter and therefore the layout criteria have been left out.

The evaluative framework described here will be used to analyse a specific ADS-generated governmental decision later. Thereafter, the framework will be used to create an alternative presentation format for that decision with the main goal to enhance the citizen's satisfaction and acceptance of the decision.

3. Case Study on Student Loans in the Netherlands

Ideally, this study would focus on an AI application that is representative of approaches that raised the issue of explainability, in other words deep-learning or similar sub-symbolic technologies. However, The Council of State said that with the less complex technologies, problems still emerge. The absence of sub-symbolic tools in administrative practice means that a decision was made to look into popular, rule-based, symbolic AI.

The case selected is the application that is used to decide on student loans, deployed by the Education Executive Agency (referred to as DUO in Dutch), an administrative agency that falls under the responsibility of the Dutch Ministry of Education. The ADS for deciding on student loans uses symbolic AI. More specifically, it is a rule-based system that contains different rules that are evaluated when deciding on the entitlement of students to financial support.

3.1. Designing a Conceptual Disposal

After analysing the original letter from DUO, an improved version of the presentation format was developed using the principles described in the framework from section 2.2. This conceptual online letter was set up with the main goal of providing better insight into the reasoning mechanisms of the algorithm, the data used to make the decision, and the presentation of the decision in a clearer way. The six criteria for explanation, as defined earlier, were used to improve the letter in the following ways. First, the letter contains a section that informs the receiver about the change in address that affects the student's monthly loan (external coherence criterion). The order of messages, one per section, was reorganised to give a better relation between the various parts of the letter (internal coherence criterion). Different from the original letter, the conceptual letter explains the reasoning that led to the decision. As in the original letter, only one cause (change in address) was presented to explain the change in the loan to the student (simplicity criterion). The number of words in the letter was reduced for the conceptual disposal (articulation criterion). Furthermore, the student's old situation and new situation were presented together in a contrastive table (contrastiveness criterion). By offering the user the possibility to learn more about the decision via hyperlinks to more elaborated information, the student's understanding of the situation might increase as well (interaction criterion).

4. Methodology

The case selected is a symbolic AI decision-making tool used for deciding on student loans provided by DUO. The original and conceptual versions of the presentation format were subjected to an A/B test. The A/B test was included in an online survey using Qualtrics. Half of the subjects received the survey that included version A, the other half version B. Besides questions about the explainability of the presented version, the survey included questions that were used to measure the students' attitudes towards the use of ADS in the Dutch government.

Chat service WhatsApp was used for contacting around 100 students, being the target audience for the application studied. Some of the students forwarded the questionnaire to other students, resulting in 133 students who completed the survey.

4.1. Hypotheses

The case study was used to test the following hypotheses:

- There is no relation between one's trust in government and trust in computer systems within the government.
- The citizen's support for the deployment of AI by the government does not vary by case.
- The presentation format of a governmental decision will have no influence on the citizen's perceived satisfaction about that decision.
- The presentation format of a governmental decision will have no influence on the chance a citizen will accept that decision.
- The presentation format of a governmental decision will have no influence on the citizen's urge to object to or appeal that decision.

4.2. Outline of the Survey

First, the subjects were shown an introductory text that explained the current situation of AI use by the Dutch government and the purpose of the research.

Thereafter, a five-point Likert scale, ranging from strongly disagree (1) to strongly agree (5), was used to determine the participants' attitudes. The participants were asked to rate how strongly they agreed with specific statements on the use of symbolic AI in government.

Subsequently, participants were asked to evaluate a disposal of an automated decision from DUO. One original disposal was obtained from the agency itself; the other one was a more interactive disposal that was created specifically for this study and included all factors that, according to theory, would enhance explainability. The participants were randomly assigned to one of the two versions and were then asked questions to survey their satisfaction with the disposal.

Before distribution, the survey was checked by three individuals to ensure understandability.

4.3. Participants

For finding subjects for the A/B test and the survey, a convenience sample was taken. The sample selection resulted in 133 subjects responding and completing the survey. The students recruited were enrolled in various universities and colleges in the Netherlands. From the total group, 60 students (45.1%) were female, and 73 students (54.9%) were male. All the participants were aged between 18 and 30, with an average age of 23.46 years (SD = 1.78). Most of the students were currently enrolled in an academic master's programme (49.6%), followed by academic bachelor students (28.6%), and 14 respondents were enrolled in a bachelor's programme at a university of applied sciences (10.5%). Additionally, there was one student enrolled in an applied sciences master's programme (0.8%) and one student from college (0.8%). Thirteen participants noted that they were currently not in school (9.8%). The next section discusses the data preparation and analysis, and the results are then discussed.

5. Analysis and Main Findings

The 133 students involved in the A/B test were split into one group of 68 persons who received the survey on the original letter and 65 who received the survey on the conceptual letter. A check for sampling independence between the two groups was then performed. No difference in gender ($\chi^2(1) = 0.013$, p = .910), age (t(131) = 0.662, p = .509) or education level between the groups ($\chi^2(5) = 5.161$, p = .397) was found.

Our first hypothesis was rejected as we found a correlation between the *trust in government* and the *trust in computer systems within government* (F(1,131) = 14.137, p < .0005, R^2 = .097, b = 0.333, t(131) = 3.760, p < .0005).

The respondents were asked for what tasks they support the deployment of computer systems for governmental use. Students stated that they support the use of computer systems for the optimisation of traffic flows (91.7%), the calculation of student finance (84.2%) and the calculation of tax assessment (80.5%). Only 34.6% of the students have the opinion that automated systems should be used for the rejection or grant of visas. Cochran's Q shows that agreement ratios for these four purposes are not identical (Cochran's Q(3) = 144.437, p < .0005). Post-hoc McNemar tests with Bonferroni correction showed that the students' support for automated systems for visa decisions is significantly lower than the three other variables. Therefore, we colclude that the students' support for the deployment of AI in government varies by use.

Since the dependent variables do not follow a normal distribution in either condition (*Original Disposal*: Shapiro-Wilk W(68) = .941, p = .003, *Conceptual Disposal*: Shapiro-Wilk W(65) = .936, p = .002), the t-test cannot be used. Therefore, a non-parametric Mann-Whitney U test is preferred to analyse the difference between the clarity of the two letters. One of the major findings of this study is that students are more satisfied with the conceptual disposal than the original disposal (U = 1082.5, z = 5.112, p < .0005). Furthermore, respondents also agreed with the statement *'I prefer an interactive (clickable) letter.'* With an average score of 3.80 on the five-point Likert scale, this was also significantly higher than the neutral value of 3.0 on the five-point Likert scale (t(132) = 10.805, p < .0005). Therefore, this study finds that students will be more satisfied with a more interactive letter than the original letter from DUO.

Furthermore, it is shown that the letter type (original or conceptual) has a significant influence on the acceptance of the decision. Respondents agree significantly more easily with the statement *'The content of the letter convinces me to agree with the decision.'* when receiving the conceptual letter (U = 1550, z = 3.331, p = .001). Therefore, the letter type, the presentation format of the governmental decision, has a significant influence on the acceptance of the decision by the student.

No significant difference between the two letter conditions was found in the urge to object to or appeal the decision (U = 1967, z = 1.186, p = .235). However, the explanation in the conceptual letter was found to be more beneficial for the support and argumentation of a potential objection or appeal (U = 1577, z = 2.979, p = .003). Also studied was the way in which the students agreed with the statement that a good explanation of the decision would help to reduce the chance of objection or appeal. With an average score of 4.02 on the five-point Likert scale, this is significantly higher than neutral, which has the value 3.0 (t(132) = 13.319, p < .0005). Therefore, it can only be stated that the citizen's willingness to object to or appeal the decision might only be reduced by offering a better explanation.

6. Conclusion

The adoption of new AI technologies by governments bring challenges such as the potential bias in the algorithms exploited and, certainly in case of data-driven, sub-symbolic AI approaches, the general lack of explainability of the decision-making processes supported by those algorithms. As a result, a renewed interest in XAI emerged. Equally important is the transformation in the way governments interact with their citizens thriving for higher effectivity and costs reduction leading to AI-usage in a wide variety of previously manually operated tasks. This study aims to contribute to this growing area of research by exploring the principles of explanations, and it offers a framework that strives to assess the quality of a given explanation. When analysing a Dutch disposal, it seems that the government is already doing a great job with a bright, interactive and straightforward letter. However, the way the government currently interacts with the citizens can be significantly improved.

In order to achieve a better understanding of the citizen, a digital letter should be compatible with existing knowledge of the citizen; the parts of the letter have to fit together and use as few causes possible; and the letter should be written clearly, provide contrastive information and offer the opportunity to interact.

Several conclusions can be drawn from the quantitative study. A significant relation between one's trust in the government and the trust in computer systems used by that government was found. The citizen's support for the deployment of AI by the government varies per use or case, and more research is necessary to better understand why. This research demonstrates that students will be more satisfied with a more interactive letter than the current original letter from DUO. Furthermore, it can be concluded that a clearer explanation of the decision will lead to a greater likelihood of accepting that decision, which also confirms the previous studies as discussed in section 2.1. Therefore, governments can increase the acceptance rate of citizens by improving the clarity of their explanations, and this can create a new field of interest in *explanation optimisation*. Lastly, the study found that letter type has no significant influence on the urge to object to or appeal the governmental decision. On the contrary, a good explanation of an automated governmental decision was found to help to reduce the citizen's willingness to object to or appeal that decision.

The study also reconfirms that while investments in AI supporting various tasks of public administrations are merely driven by the need for improving efficiency and effectiveness. It is important to keep in mind that explainability, transparency, accountability and auditability are essential to governmental processes.

7. Discussion

There are several limitations that need to be addressed for this study. First, this study mainly focuses on the adoption of rule-based AI-systems within the Dutch government. Data-driven, sub-symbolic AI technologies have become more popular but have even larger problems with explainability and fairness. At this moment very few governmental agencies within the Netherlands make use of data-driven sub-symbolic AI-technologies for their decision-making. Governmental agencies such as the Dutch Tax and Customs Administration (De Belastingdienst) stated that they were using sub-symbolic AI for var-

ious fields such as the prediction of fraud, and other agencies are either exploiting or considering the use of such technologies for similar purposes. This authority however did not want to provide materials on their reasoning mechanisms for this research because they were perceived to be confidential (intended lack of transparency). Therefore, a decision was made to collaborate with DUO, which provided materials on the reasoning mechanisms of their algorithms. The further adoption of data-driven AI-technologies would only raise the importance of XAI. Future studies in this field should also include such data-driven AI-technologies, as they are the most problematic in terms of explainability, fairness and transparency.

Acknowledgments. We would like to thank the Canadian Research Council sponsoring the ACT Project.

References

[1] W. Samek, T. Wiegand and K.-R. Müller, Explainable artificial intelligence: Understanding, visualizing and interpreting deep learning models, *arXiv preprint arXiv:1708.08296* (2017).
[2] D. Gunning, Explainable Artificial Intelligence, *Defense Advanced Research Projects Agency (DARPA)* (2017).
[3] D. Baehrens, T. Schroeter, S. Harmeling, M. Kawanabe, K. Hansen and K.-R. Muller, How to explain individual classification decisions, *Journal of Machine Learning Research* **11**(Jun) (2010), 1803–1831.
[4] M. van Kempen, Motivering van automatisch genomen besluiten, *Knowbility* (2019).
[5] A. Dhasarathy, S. Jain and N. Khan, When governments turn to AI: Algorithms, trade-offs, and trust, *McKinsey&Company* (2019). https://www.mckinsey.com/industries/public-sector/our-insights/when-governments-turn-to-ai-algorithms-trade-offs-and-trust.
[6] M. Carrasco, S. Mills, A. Whybrew and A. Jura, The Citizens Perspective on the Use of AI in Government, *Boston Consulting Group* (2019).
[7] J.C. Giarratano and G. Riley, *Expert Systems*, PWS Publishing Co., 1998. ISBN ISBN 0534950531.
[8] AINED, *AI voor Nederland: vergroten, versnellen en verbinden*, 2018.
[9] V. Homburg, *Understanding e-government: Information systems in public administration*, Routledge, 2008. ISBN ISBN 1134085028.
[10] Y. Lv, Y. Duan, W. Kang, Z. Li and F.-Y. Wang, Traffic flow prediction with big data: a deep learning approach, *IEEE Transactions on Intelligent Transportation Systems* **16**(2) (2014), 865–873.
[11] B. Corydon, V. Ganesan, M. Lundqvist, E. Dudley, D.-Y. Lin, M. Mancini and J. Ng, Transforming Government Through Digitization, *McKinsey & Company* (2016). https://www.mckinsey.com/˜/media/McKinsey/Industries/Public Sector/Our Insights/Transforming government through digitization/Transforming-government-through-digitization.ashx.
[12] Dutch Digital Government, NL DIGIbeter, *Digital Government Agenda* (2018). https://www.nldigitalgovernment.nl/document/digital-government-agenda-2/.
[13] M. Reid, Rethinking the Fourth Amendment in the Age of Supercomputers, Artificial Intelligence, and Robots, *West Virginia Law Review* **119** (2017), 863–890.
[14] R.V. Schuwer, *Het nut van kennissystemen*, 1993. doi:10.6100/IR394707.
[15] J. Haugeland, *Artificial Intelligence: The Very Idea*, MIT Press, 1985. ISBN ISBN 0262580950.
[16] P. Smolensky, Connectionist AI, symbolic AI, and the brain, *Artificial Intelligence Review* **1**(2) (1987), 95–109. doi:10.1007/BF00130011.
[17] H. Lieberman, Symbolic vs. Subsymbolic AI, *MIT Media Lab* (2016). http://futureai.media.mit.edu/wp-content/uploads/sites/40/2016/02/Symbolic-vs.-Subsymbolic.pptx.pdf.
[18] H. Mehr, Artificial Intelligence for Citizen Services and Government, *Harvard Ash Center Technology & Democracy* (2017). https://ash.harvard.edu/files/ash/files/artificial_intelligence_for_citizen_services.pdf.
[19] T.D. Kelley, Symbolic and Sub-Symbolic Representations in Computational Models of Human Cognition: What Can be Learned from Biology?, *Theory & Psychology* **13**(6) (2003), 847–860.
[20] J. Angwin, J. Larson, S. Mattu and L. Kirchner, Machine Bias, *ProPublica* (2016). https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing.

[21] G. Sileno, A. Boer and T. van Engers, The Role of Normware in Trustworthy and Explainable AI, 2018. http://arxiv.org/abs/1812.02471.
[22] Raad van State, Ongevraagd advies over de effecten van de digitalisering voor de rechtsstatelijke verhoudingen, *Kamerstukken II 2017/18, 26643, nr. 557* (2018). https://www.raadvanstate.nl/adviezen/zoeken-in-adviezen/tekst-advies.html?id=13065.
[23] European Union, Regulation 2016/679: General Data Protection Regulation, *Official Journal of the European Communities* (2016), 1–88. ISBN 9251032718. doi:http://eur-lex.europa.eu/pri/en/oj/dat/2003/l_285/l_28520031101en00330037.pdf.
[24] E.H. Shortliffe and B.G. Buchanan, *Rule-based expert systems: the MYCIN experiments of the Stanford Heuristic Programming Project*, Addison-Wesley Publishing Company, 1985. ISBN ISBN 0201101726.
[25] E. Horvitz, D. Heckerman, B. Nathwani and L. Fagan, The use of a heuristic problem-solving hierarchy to facilitate the explanation of hypothesis-directed reasoning, in: *Proceedings of Medinfo, Washington, DC*, 1986, pp. 27–31.
[26] J.E. Mercado, M.A. Rupp, J.Y.C. Chen, M.J. Barnes, D. Barber and K. Procci, Intelligent Agent Transparency in Human-Agent Teaming for Multi-UxV Management, *Human Factors* **58**(3) (2016), 401–415. doi:10.1177/0018720815621206.
[27] English Oxford Dictionaries, Definition of 'explanation' in English, 2019. https://en.oxforddictionaries.com/definition/explanation.
[28] R. Neches, W. Swartout and J. Moore, Enhanced Maintenance and Explanation of Expert Systems Through Explicit Models of Their Development, *IEEE Transactions on Software Engineering* **SE-11**(11) (1985), 1337–1351. doi:10.1109/TSE.1985.231882.
[29] T.R. Roth-Berghofer and J. Cassens, Mapping goals and kinds of explanations to the knowledge containers of case-based reasoning systems, in: *International Conference on Case-Based Reasoning*, Springer, 2005, pp. 451–464.
[30] J.L. Herlocker, J.A. Konstan and J. Riedl, Explaining collaborative filtering recommendations, in: *Proceedings of the 2000 ACM conference on Computer supported cooperative work*, ACM, 2000, pp. 241–250. ISBN ISBN 1581132220.
[31] F. Sørmo, J. Cassens and A. Aamodt, Explanation in case-based reasoning-perspectives and goals, *Artificial Intelligence Review* **24**(2) (2005), 109–143.
[32] R. Ye and P. Johnson, The Impact of Explanation Facilities on User Acceptance of Expert Systems Advice, *MIS Quarterly* **19**(2) (1995), 157–172. doi:10.2307/249686. http://www.jstor.org/stable/249686.
[33] D. Doran, S. Schulz and T.R. Besold, What does explainable AI really mean? A new conceptualization of perspectives, *arXiv preprint arXiv:1710.00794* (2017).
[34] W. Pieters, Explanation and trust: what to tell the user in security and AI?, *Ethics and information technology* **13**(1) (2011), 53–64.
[35] J.Y. Chen, K. Procci, M. Boyce, J. Wright, A. Garcia and M. Barnes, Situation awareness-based agent transparency, Technical Report, 2014.
[36] A. Falcon, Aristotle on Causality, in: *Stanford Encyclopedia of Philosophy*, 2008.
[37] S.E. Toulmin, The Uses of Argument, *Cambridge University Press* (1958). ISBN 0521827485.
[38] P. Thagard, Explanatory coherence, *Behavioral and brain sciences* **12**(3) (1989), 435–467.
[39] T. Miller, Explanation in artificial intelligence: Insights from the social sciences, *Artificial Intelligence* (2018).
[40] J.C. Zemla, S. Sloman, C. Bechlivanidis and D.A. Lagnado, Evaluating everyday explanations, *Psychonomic Bulletin and Review* **24**(5) (2017), 1488–1500. doi:10.3758/s13423-017-1258-z.
[41] N. Pennington and R. Hastie, *The story model for juror decision making*, Cambridge University Press Cambridge, 1993. ISBN ISBN 0521419883.
[42] S.J. Read and A. Marcus-Newhall, Explanatory coherence in social explanations: A parallel distributed processing account, *Journal of Personality and Social Psychology* **65**(3) (1993), 429.
[43] T. Lombrozo, Simplicity and probability in causal explanation, *Cognitive Psychology* **55**(3) (2007), 232–257.
[44] P. Lipton, Contrastive explanation, *Royal Institute of Philosophy Supplements* **27** (1990), 247–266.
[45] D.J. Hilton, Conversational processes and causal explanation, *Psychological Bulletin* **107**(1) (1990), 65.
[46] I. Nunes and D. Jannach, A systematic review and taxonomy of explanations in decision support and recommender systems, *User Modeling and User-Adapted Interaction* **27**(3–5) (2017), 393–444.
[47] R. Hoffman, S. Mueller, G. Klein and J. Litman, Metrics for Explainable AI: Challenges and Prospects, *XAI Metrics* (2018).