Knowledge Representation for Algorithmic Auditing to Detangle Systemic Bias

Paola Di Maio[a]

[a]Center for Systems, Knowledge Representation and Neuroscience

Abstract

Knowledge is central to cognition, and adequate knowledge representation is necessary to ensure the correct functioning of intelligent systems leveraging knowledge. Knowledge Auditing methodologies in use, however do not typically address the requirements for evaluating the consistency of logical and functional knowledge representation constructs, as used in AI, that support the correct execution of system processes. This paper puts forward the notion that algorithmic bias results from the lack of adequate knowledge representation mechanisms, and is part of systemic bias. The notion of systemic bias is identified, characterized, and described as an emergent relation resulting from the combination of other factors between knowledge representation (KR) and bias in AI.These factors are here named conceptualised, related to each other and visualised.

The resulting artefacts are then incorporated in the KAF (Knowledge Audit Framework) methodology (first published in 2012) and adopted as a reference model for understanding and representing bias as a systemic emergent phenomenon.

1. Introduction and Motivation

Despite the proliferation of Knowledge Based Technologies and Knowledge Representation (KR) methods and techniques in AI, their impact on system fairness and algorithmic bias has not been studied. This research considers bias as resulting from the lack of adequate knowledge representation, which is essential to ensure the logical and functional consistency and integrity of intelligent systems. This paper investigates the relation between KR (Knowledge Representation) and bias. It provides a definition and a model representation of systemic bias. A set of criteria and a checklist for the evaluation of KR quality and adequacy are provided as a guideline to complement algorithmic audits aimed at identifying bias and its possible causes and dependencies at system level. A module KR is added to the KAF methodological framework which is also briefly summarized.

2. Semantics and Knowledge Representation

In recent years research in semantic technology has been influenced by the development of semantic web standards such as OWL and RDF, however the relationship between semantics and knowledge representation at large is historical and long standing both in classical sciences such as mathematics and linguistics, as well as in computer science and other spheres of human thinking, pre-dating by several decades the semantic web and related published standards [1]. Semantics relates to the ability of words to capture and convey meaning thanks to the relations between them. This can be done using more than one modelling approach, for example also using markup languages such as SysML [2, 3].

KR, especially symbolic KR, which relies on natural language representations as a mechanism to capture the relationship between words and concepts, is intended to formalize and make explicit the semantics of natural languages. In this paper, semantics is considered inherent to all Knowledge Representation methods and techniques and it is not discussed in relation to any specific implementation language nor standard. Ultimately any valid first order logic construct can be mapped to OWL and RDF semantics anyway [1].

3. Accountability And Algorithmic Bias

Intelligent systems powered by algorithms have become pervasive and embedded, and algorithmic accountability is an imperative quality criterion for AI, whereby plausible justifications are necessary for reliable automated system's outputs. Many types of bias are identified and discussed in a growing body of Machine Learning literature [4] and are being addressed in academic research as well in the praxis with initiatives that aim to disseminate principles and good practices

[1]OWL Working Group. 'OWL 2 Web Ontology Language Document Overview: W3C Recommendation 27 October 2009'; (2009)

International Semantic Intelligence Conference (ISIC 2021), Feb 25-27, 2021, New Delhi, India
EMAIL: paola.dimaio@gmail.com (P.D. Maio)
ORCID: 0000-0003-0770-3713 (P.D. Maio)

CEUR Workshop Proceedings (CEUR-WS.org)

summarised as FAT (Fair Accountable and Transparent) . These include

- `Responsibility` : explicit measures to guarantee accountability
- `Explainability` : clarity of exposure of technical content for a variety of stakeholders.
- `Accuracy` : mechanism to ensure semantic and computational precision and minimise error rates.
- `Auditability` : ability to access, open inspect and verify content, data and technical systems implementations.
- `Fairness` : equanimity, balanced and unbiased processes and outcomes.

Several attempts are underway to devise a suitable conceptual structure to represent bias, however given that innumerable distortions and flaws characterize human thinking and consequently systems and machines designed by humans, the resulting list and categorization mechanism for bias is long and unsorted. There is a wealth of research literature documenting various kinds of biases, and several attempts to group them into categories.However there is also category bias [5], which limits the usefulness of categorization efforts. Although biases can be grouped according to different needs and views, the choice of taxonomic arrangement is determined by the adopted worldview, by the tasks, the norms in the field of practice or discipline in which it is applied, the organisational context in which it occurs.

There is a continuum of dependencies and a circularity among various factors that express and contribute to bias. When taking a systemic, high level view of what different types of bias consist of and what causes them, most types of biases are co-related and co-occur-rent. In this paper I demonstrate how these factors can be arranged in a top level cluster, resulting in a type of bias which is defined here as 'systemic'. Systemic bias is therefore any bias that impacts the accuracy of a representation fairness and operation of a system functioning and outcomes and that cannot be ascribed to a single factor [6].

3.1. Disentangling Systemic Bias

There exist different ways to categorize bias, reflecting a diversity of epistemological perspectives, technical goals, constraints and other factors. Some types of bias are well understood and exist aside from AI implementation such as cognitive bias, research bias etc.

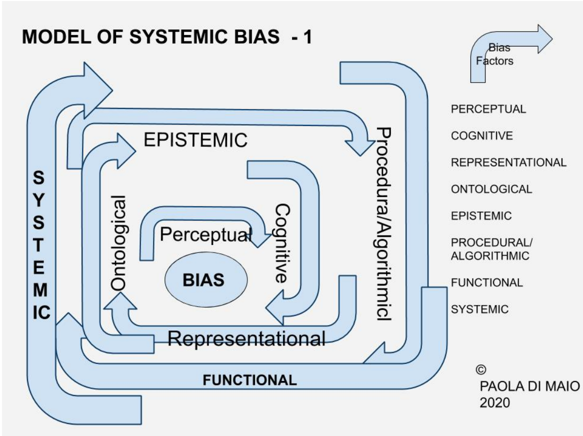

Figure 1: Systemic Bias

Algorithmic and Machine Learning biases are central to AI [7, 8, 9].

However, given algorithms of different kinds with implementation that may or may not resort to the use of ML, research attempting to study algorithmic bias focussing only on ML inevitably falls short and does not cover the entire range of possible biases in AI driven systems.

In the course of analysis carried out during scholarly endeavors, as well as within the IEEE Taxonomy of bias P7003 WG (2020) a high level representation of bias factors and causes has been proposed that can be visualised as a mindmap and a taxonomy.

Initially a concept map is devised showing that bias starts from a distortion in perception and cognition and continues to develop through knowledge misrepresentation, becomes established in biased ontology and epistemology and ends up becoming operationalised into biases at procedural, algorithmic, functional and systemic levels.

These factors are conceptualised in figure 1 and and can be described as follows:

- `Perceptual` : Given that organs and senses of perception are limited and imperfect, perceptual bias corresponds to limitations in the perceptual apparatus.
- `Cognitive` : Many kinds of bias generated by flaws and limitations in cognition [4].
- `Representational` : Caused by limitations in the ability to capture, represent and communicate the representation of knowledge [10] can be characterised further as a type of knowledge misrepresentation i.e., if some fact is false, partially true, incomplete or placed out of context.

- `Epistemic` : Relating to demarcation, validation of truth and belief and choice of paradigm.

- `Ontological`: bias relating to assumptions, axioms, boundary, definitions, relations and assertions.

- `Procedural` : Procedural bias relates to imperfections and flaws in the explicit representation of procedures, or the incorrect implementation and management.

- `Functional` : a malfunction or dysfunction of the system, whether intentional or accidental.

- `Systemic` : Any emergent bias that cannot be ascribed to a single cause and is generated by a combination of the other types of bias listed above.

3.2. Knowledge Representation

Knowledge representation (KR), in particular Symbolic KR, consists of techniques and methods devised to capture system logic so that it can be leveraged by algorithms. KR can also be useful in mitigating at least in part the lack of transparency of machine learning and algorithmic opacity, a role which is not fully explored. In addition to the well documented canonical roles of KR in supporting the development of AI and well documented in classical textbooks [11]. The term KR is generalized to indicate:

- A multifaceted discipline

- A field of study in Computer Science, jurisprudence, learning and other fields KR as a technique/process

- A set of activities in Knowledge Modeling and Knowledge Engineering applied to systems development KR as artifacts

- Knowledge encoded and implemented as a product of KR activities (for example a graph, an ontology or an algorithm)

In attempting to identify and resolve bias, to which systemic dysfunction is here ascribed, it is necessary to resort to Knowledge Representation. First, a concept map is proposed that supports the navigation of bias based on symptoms, causes, types, mitigation and strategies as a vertical dimension, and the factors listed above as a horizontal dimension.

Secondly, the factors identified in the concept map shown in figure 1, are reconfigured as a Relational set

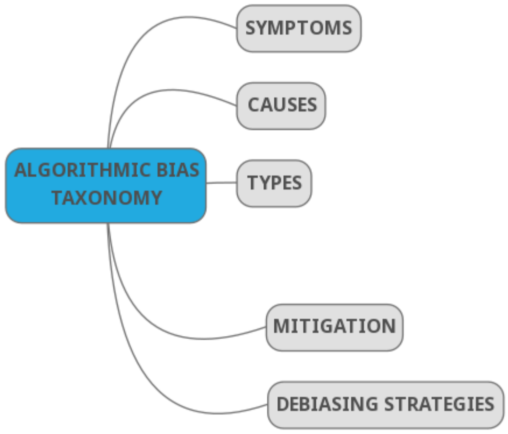

Figure 2: Example of an information KR to represent and detangle bias (P Di Maio Research notes IEEE P7003 work group on Taxonomy Bias, 2020)

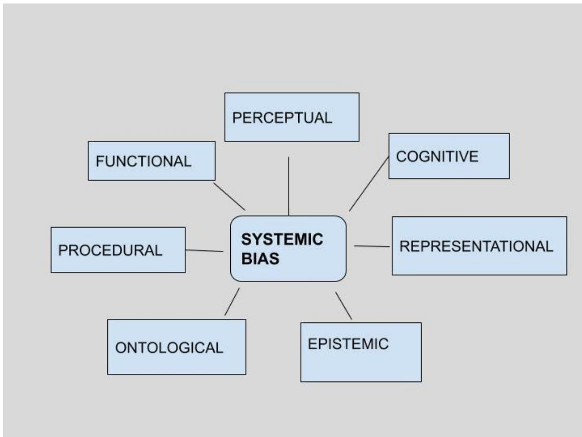

Figure 3: Systemic Bias and its contributing factors represented as a faceted structure (Di Maio 2020)

in a table, so that concepts from any system domain can be mapped to either factors and causes

Finally these relations can be visualised as taxonomies, either faceted or networked for example (figure 2 and 3)

Taxonomies are information structures used in categorization [12].

3.3. Deep fakes As Knowledge Misrepresentation

The social, technical and economic impact of bias, especially in the case of systemic, entangled biases cannot be estimated. Debiasing can add considerable costs

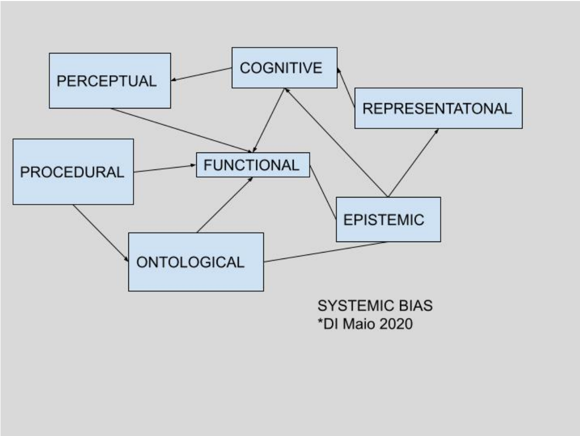

Figure 4: networked taxonomy

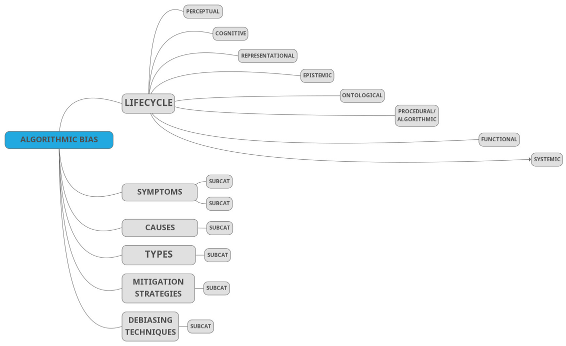

Figure 5: Towards a Taxonomy of Bias (Di Maio, 2020)

in terms of time and money to the software and systems development. In AI, and specifically AI powered by Machine Learning, bias can generally be directly or indirectly mapped to knowledge misrepresentation. This is demonstrated by the case of Deep fakes [13] where hidden algorithmic manipulations deliberately distort knowledge representation and produce a false output intended to manipulate public opinion, create misinformation and confusion which costs time and money to rectify. Technically speaking, an algorithm can function correctly and no error of flaw is apparent without considerable expert human discernment and skill to investigate what type of hidden automated permutations deliver the capability of fictional outputs.

Knowledge misrepresentation, in different forms can cause ‘Systemic Deviation‘ a behavioral and functional shift of a system from its intended purpose. Systemic deviation can occur when [14] :

1. A system is designed with the intent to disguise its real aim to achieve a different goal from what it appears to be designed for. For example in the case of software or hardware designed to spy but marketed as technology to protect privacy is considered systemic deviation.
2. When a system is deliberately designed to function in a different way from what is stated. The system says it offers localization, but it is used for tracking and passes the tracking data to unauthorized third parties.
3. A system designed to function in a certain way for a certain (ethical) purpose, is misused and made to function to fulfill the exact opposite (unethical) purpose as in the case of health monitoring software designed to save patient lives, which can be used by medics to make end of life decisions based on sociological or economic factors (such as not providing Covid treatment to a tetraplegic patient on some statistical rationale) [2].
4. A system knowledge representation is deliberately manipulated to produce an output that does not conform correctly to the intended KR model

Eventually, these dimensions can be used to create a relational structure to plot concepts from any given domain, say for example the system or conceptual domain under study to attempt to establish, extricate and disentangle the various factors so that they can be addressed accordingly.

4. Auditing: Data, Knowledge, Algorithms

There are many kinds of system compliance auditing processes and methods used in technology engineering and management to ensure that systems do what they say they do. They tend to be very complex, because they tackle different system levels without disambiguating them.

Audits are defined as “independent evaluations of conformance of software products and processes to applicable regulations, standards, guidelines, plans, specifications, and procedures“ []. In organisational settings they are defined as tools for interrogating complex processes, often to determine whether they comply with company policy, industry standards or regulations [15].

[2] W. Post, article, 2020. URL: https://www.washingtonpost.com/health/2020/07/ 05/coronavirus-disability-death/

4.1. Data Auditing

In software engineering, databases, data dictionaries and catalogues contain data sets and their properties and attributes maintained by developers and administrators. From an organisational point of view, data audits are inventories published by institutions and comprising digital assets. One example of published methodology for Data Audit is DAF which was scoped to handle mostly research data assets and constitutes the methodological basis for knowledge auditing methodologies which evolved from it [16].

4.2. Knowledge Auditing

A Knowledge Audit consists typically of an inventory, a knowledge map and a knowledge flow. [17] Knowledge Auditing (KA) is an established practice in Knowledge Management, having the purpose to provide an inventory of knowledge assets and identify knowledge gaps. It is typically devised to answer questions such as:

What knowledge exists? And what knowledge is missing?

Additional KA techniques include Knowledge surveys, Assets mapping, Intellectual Capital (IC) Inventorying, Knowledge Landscape Mapping, Competitive Knowledge Analysis, Knowledge Flow charting and Analysis (KFA) and others. Knowledge Audits are typically carried out on site whereby auditors visit the premises of a company or organisation owning the - knowledge Assets, often for the purpose of valuation on behalf of a buyer in relation to a takeover and are carried out behind closed doors, their outcomes are not public. Knowledge Auditing however has been used also, as described in the following section 4.3 as a baseline method to audit Open Data, Open Knowledge and Open Access resources, produced by the increasingly widely accepted policies which man date the accessible and transparent access of resources produced by publicly funded research. In public sector affairs, KA can be used to provide a measure of the extent to which a public administration publishes knowledge and to what extent is this knowledge accessible, answering the question:

What knowledge generated by public funding can be openly (without payment or subscription required) accessed online?

4.3. The Knowledge Audit Framework

Initially devised to fulfil the requirements for doctoral research [18], the initial purpose of KAF was to enable the systematic evaluation of Knowledge Sharing Artefacts and Practices adoption, with particular emphasis on reuse. The need for remote auditing originally arose during the doctoral research because to obtain physical access to premises where knowledge was held would require overcoming bureaucracy and red tape to obtain clearances and permissions. Later however, during the current Covid 19 Pandemics at the time of writing, remote knowledge auditing method makes it possible to conduct audits online even during lockdowns.

KAF was devised as an instrument to carry out remote knowledge audits and to enable an objective evaluation of what knowledge resources are publicly shared online, their ease of accessibility via open web searches, where they are located, who is (or is not) responsible for their occurrence, validity, maintenance. The main strengths of KAF are that

1. it is the first knowledge auditing methodology that supports remote audits which can be carried out via online searches and do not require physical access to documents.
2. the results of the audits are replicable - given a standardized set of search criteria two different auditors would come up with very similar results.

Integrating Knowledge Representation Auditing into KAF can serve as a model for standardizing algorithmic auditing methodologies and processes.

4.4. Algorithm Auditing

Algorithmic auditing has entered the software development practice and is used generally as a posterior, i.e. after the algorithm has been compiled, to evaluate its fairness and identify possible biases. But in systems engineering terms, this is not optimal. Innumerable examples of algorithmic bias are provided in literature and frameworks attempting to mitigate such biases however address the artifact level (the knowledge artefacts) rather than the reasoning and inferences contained in the algorithm [19].

Algorithmic auditing methods have been evaluated empirically but not been systematically bench-marked at the time of writing. Examples with different emphases, weaknesses and strengths are summarized in 1 [20].

None of the algorithmic auditing methods surveyed tackles the bias deriving from Knowledge Misrepresentation and the lack of Knowledge Representation

Table 1
source [20] Sandvig et al 2014

Type of Algorithmic Auditing	Strength	Weakness
Code Audits	Straightforward	Code not shared
Non Invasive User Audits	Ease of use	Sampling limited
Scraping Audits	Easily automated	Breach of terms of uses and CFAA regulations
Sock puppet	Provides anonymity	Fake users/fabrication/reliability
Crowd sourced/collaborative	High volume	High cost/requires organisational effort

adequacy. This paper places Knowledge Representation of Bias, using a concept map and a faceted taxonomy as illustrated in figure 1 and 2 in the context of algorithmic auditing and adds it as an auditing instrument to KAF, the Knowledge Audit Framework [21]. The extended rationale for this contribution is provided in 5 below.

5. Knowledge Representation For Algorithmic Audits

In relation to Knowledge Representation, a Knowledge Audit seeks to answer questions such as:

Is knowledge adequately represented to fulfil fairness requirements within the scope and goals of the system?

5.1. Criteria for Auditing KR

The body of knowledge of KR in AI is vast, providing innumerable computational techniques and methods for the modelling of Knowledge in intelligent systems, yet the fundamental purpose of KR can easily become lost in the computation: to represent knowledge adequately, to state things as they are and not otherwise. This results in great mixups.

Understanding and addressing the different types of adequacy in AI is not trivial. The first and foremost risk of any representational device is to either not represent knowledge (omission) or to mis-represent knowledge (distortion). Both types of misrepresentation can occur very easily and cause malfunctions and result in the incorrect outputs. Malicious engineers know how to engineer bias and deviation by exploiting misrepresentation. Principles and criteria to evaluate KR are shared with, and sometimes derived from, neighbouring disciplines, such as for example Formal Logic, Information Science and Ontology Engineering. Adding KR quality and adequacy criteria to algorithmic auditing is therefore desirable and feasible, and delivers the benefit of supporting the clear distortions (bias) that can lead to systemic deviation and malfunction. This section provides a set of criteria with corresponding justification and a checklist to facilitate this task for both developers and auditors.

5.2. Levels of Representation

Knowledge representations range from computer-oriented forms to conceptual ones nearer to those present in our internal world models. Five knowledge levels are identified in literature [22]:

- `Implementational`: It includes data structures such as atoms, pointers, lists and other programming notations.
- `Logical` : Symbolic logic is inside this level. Thus, symbolic logic propositions, predicates, variables, quantifiers and Boolean operations are included.
- `Epistemic` : A level for defining concept types with sub-types, inheritance, and structuring relations.
- `Conceptual` : The level of semantic relations, linguistic roles, objects and actions.
- `Linguistic` : Deals with arbitrary concepts, words and expressions of natural languages.

A systemic view of KR levels therefore emerges from practice to evaluate adequacy of KR that identifies KR as Task, Domain, System and System of Systems Level, which correspond to different levels of functionality, This view, shown in the image 6 below shows how KR can contribute to the disentanglement of systemic bias.

5.3. Adequacy

To be useful, KR needs to adhere to certain criteria and requirements, which at least in part are common to most types of information systems, knowledge bases and ontologies.

In relation to expert systems two types of adequacy criteria for of KR were identified [22], such as:

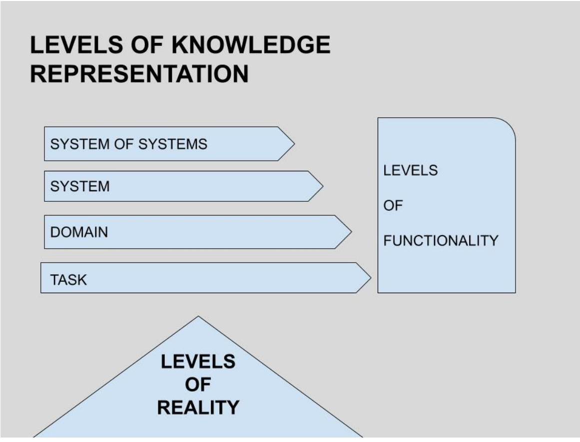

Figure 6: Level of KR adequacy: Di Maio 2020

- `Terminological adequacy` the ability to form the appropriate kind of technical vocabulary and understand the dependencies among the terms.

- `Assertional adequacy` involves the ability to form the kind of theory appropriate to the world knowledge of a system and understand the implications of the theory [23].

  Other forms of quality metrics for KR [11] include:

  1. Representational Adequacy – the ability to represent all the different kinds of knowledge that might be needed in that domain.
  2. Inferential Adequacy – the ability to manipulate the representational structures to derive new structures (corresponding to new knowledge) from existing structures.
  3. Inferential Efficiency – the ability to incorporate additional information into the knowledge structure which can be used to focus the attention of the inference mechanisms in the most promising directions.
  4. Acquisitional Efficiency– the ability to acquire new information easily. Ideally the agent should be able to control its own knowledge acquisition, but direct insertion of information by a 'knowledge engineer' would be acceptable.

Finding a single system that automates the optimization of these for all possible parameters is not always feasible, however researchers and developers can develop and apply their own evaluation schemas for KR using as a reference and guidelines the criteria listed here.

RELATIONAL STRUCTURE FOR THE MINDMAP (TO BE FILLED WITH KEYWORDS FROM THE KNOWLEDGE BASE)

	PERCEPTUAL	COGNITIVE	REPRESENTATIONAL	EPISTEMIC	ONTOLOGIC	PROCEDURAL	FUNCTIONAL	SYSTEMIC
SYMPTOMS								
CAUSES								
TYPES								
MITIGATION								
DEBIASING								

Figure 7: Towards a Taxonomy of Bias P Di Maio 2020

In the light of the recent emphasis on AI safety, adequate KR should also address Fairness accountability, transparency, and trustworthiness criteria and the identification of Risks and corresponding mitigation strategies as well as the identification of bias, Sources of bias and Debiasing techniques. These top level dimensions are identified and related to dimensions in the mindmap in a table

It is therefore proposed here that quality evaluation for Algorithmic KR can take place across many dimensions:

- Adequacy (Variety of Expressiveness, Modularity, Semantics, a Organization of Knowledge, conformance to a correct model of reality and other criteria);

- Inference Methods (Reasoning Strategies, Data, Control and Search Strategies);

- Inference Requirements (Computational Efficiency, Transparency of line of control, Completeness, and Consistency);

- Ability to use a prior knowledge, and update it with newly acquired knowledge.

- Dealing with incomplete and imperfect knowledge.

- Correctness Evaluation - the KR should include the ability to estimate its representational correctness.

6. Algorithmic Auditability Checklist based on KR

Algorithmic auditing may not be entirely automated due to the inherent lack of replicability of certain ML methods, using a combination of different KR techniques in triangulation facilitates the explicit representation and replication, even partial, of artificial neural network computation. A sample checklist is synthesized from good practices in KR and offered here as a set of heuristic criteria to guide developers who may want to attempt creating an algorithm for algorithmic auditability, as well for auditors and developers of auditing methodologies.

Algorithmic Audit should identify and makes explicit:

- Individuals/Components/Entities of the Algorithm
- Axioms/Laws/Constraints/Limitations
- Truth Values (consistency, persistence, conflicts)
- Processes/Functions
- Inputs/Outputs/Outcomes
- Risks and Mitigation
- Bias/Debiasing options
- Type Of Reasoning/Inference
- Relevant Variables
- Behaviours
- Structure/Patterns
- Levels Of Predictability Of The Behaviours (Probability, Randomness) Influence Factors
- Variables That May Influence Factors
- Interactions
- Type Of Notation/Encoding
- Knowledge Level (Logical, Implementational, Epistemological, Conceptual, Linguistic, Task, Domain, System)
- Overall Scientific/Computational Paradigm
- Complete/Sufficiently Describing what it represents
- Should Allow Manipulation Of The Representation For Testing/Evaluation Purposes
- Human And Machine Readable
- General FAT areas of concerns (Fairness Accountability Transparency)

6.1. Adding KR To KAF

The checklist, which evolves from general good practices in KR, can be used to inform system stakeholders, such as owners, managers and developers, and is the basis for the systemization of algorithmic audits awareness of possible areas of risk, where the algorithm can go wrong. It has been added to KAF in the form of a questionnaire to be compiled by at least four roles

1. the system designer
2. the coders/programmers
3. the independent auditors
4. stakeholders and system users at large

7. Conclusions and Future Work

This paper puts forward considerations and analyses in Systems research with a focus on Systemic bias. It synthesises perspectives and techniques from AI, KR, and knowledge auditing. It contributes a definition and a knowledge representation (a relational structure and a taxonomy) for systemic bias, which emerges from the combination of contributing factors which are also identified, described and characterized. A checklist is offered to support algorithmic auditings, and that can also be used as a guide to inform and systematize algorithmic awareness in different groups of stakeholders, as such, it is integrated in KAF, a online Knowledge Audit Framework. Future work includes further formalization of systemic bias, adding new dimensions (such as agency, temporality etc) and case studies.

References

[1] H. Hermann, Knowledge representation and the semantics of natural language (2006). doi:10.1007/3-540-29966-1.

[2] L. Wouters, S. Creff, E. E. Bella, A. Koudri, Towards semantic-aware collaborations in systems engineering, in: 24th Asia-Pacific Software Engineering Conference (APSEC), Nanjing, China, 2017.

[3] Boute, Systems semantics: principles, applications, and implementation, ACM Transactions on Programming Languages and Systems 10 (1998).

[4] R. M. A. Wilke, Encyclopedia of Human Behavior, 2nd. ed., 2012.

[5] D. Smaal, E. A, Biases in categorization, Advances in psychology 68 (1990) 349–386.

[6] Principles for accountable algorithms and a social impact statement for algorithms :: Fat ml, 2001. URL: https://www.fatml.org/resources/principles-for-accountable-algorithms.

[7] M. N., G. A., et al., A survey on bias and fairness in machine learning. (2019). doi:arXivpreprintarXiv:1908.09635.

[8] Baeza-Yates, Ricardo, Data and algorithmic bias in the web, Proceedings of the 8th ACM Conference on Web Science (2016).

[9] Danks, David, A. John, Algorithmic bias in autonomous systems, IJCAI (2017).

[10] U. P., Incremental induction of decision trees, Machine Learning 4 (1989) 161–186.

[11] R. Davis, H. Shrobe, P. Szolovits, What is a knowledge representation?, AI Magazine 14 (1993) 17–33.

[12] B. Denise, Evaluating classification schema and classification decisions, Bulletin of the American Society for Information Science and Technology (2013) 13–21.

[13] R. Chesney, D. K. Citron, Deep fakes: A looming challenge for privacy, democracy, and national security, SSRN Electronic Journal (2018). doi:10.2139/ssrn.3213954.

[14] P. D. Maio, Systemic deviation, the evil in the machine, in: Proceedings Of The 60th Annual Meeting Of The International Society For The Systems Sciences (Isss), volume 1, 2016.

[15] J. Liu., The enterprise risk management and the risk oriented internalaudit, Ibusiness 4 (2012) 287.

[16] Jones, Sarah, The data audit framework: A first step in the data management challenge, Int. J. Digit Curation 3 (2008) 112–120.

[17] P. Soltero, P.-M. J. T., et al., Knowledge audit methodology with emphasis on core processes, In European and Mediterranean Conference on Information Systems 13 (2006) 1–10.

[18] P. D. Maio, Towards shared systems knowledge, British Library (2012).

[19] Raji, I. Deborah, Closing the ai accountability gap: defining an end-to-end framework for internal algorithmic auditing, Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency (2020).

[20] CSandvig, KHamilton, K. Karahalios, Auditing algorithms: Research methods for detecting discrimination on internet platforms, Proceedings "Data and Discrimination: Converting Critical Concerns into Productive Inquiry," a preconference at the 64th Annual Meeting of the International Communication Association (2014).

[21] P. D. Maio, Knowledge audit framework (2011). doi:arXivpreprintarXiv:1108.1490.

[22] Brachman, R.J., On the epistemological status of semantic networks, Associative Networks: Representation and Use of Knowledge by Computers (1979) 3–50.

[23] P. D. Maio, Knowledge representation system level, Knowledge Representation and Systems Proceedings (2020).