

GOVERNANCE OF A DAO FOR FACILITATING DIALOGUE ON HUMAN-ALGORITHM INTERACTION AND THE IMPACT OF EMERGING TECHNOLOGIES ON SOCIETY

A PREPRINT

 **Julião Braga***

Centro de Matemática, Computação e Cognição
Universidade Federal do ABC
Santo André, SP, BR
juliao.braga@ufabc.edu.br

 **Francisco Regateiro†**

Instituto Superior Técnico
Universidade de Lisboa
Lisboa, PT
francisco.regateiro@tecnico.ulisboa.pt

 **Itana Stiubiener‡**

Centro de Matemática, Computação e Cognição
Universidade Federal do ABC
Santo André, SP, BR
itana.stiubiener@ufabc.edu.br

 **Juliana Cristina Braga§**

Centro de Matemática, Computação e Cognição
Universidade Federal do ABC
Santo André, SP, BR
juliana.braga@ufabc.edu.br

June 27, 2023

ABSTRACT

Human-algorithm interaction is a crucial issue for humanity in light of the impacts of the recent release of ChatGPT3 and 4, among others. These advanced chatbots provoked a worldwide debate in March/2023, when a manifesto signed by several stakeholders was published and widely discussed in the media and academia. This work assumes that human-algorithm interactions are influenced by a context of diverse interests and perspectives, which adds high complexity to the problem. Therefore, this work proposes a solution to enable the effective participation of stakeholders from different domains and society in a constructive dialogue, using digital platforms as a medium. Inspired by the successful governance of the Internet infrastructure ecosystem, the proposal involves the creation of an Autonomous Decentralized Organization (DAO) implemented in the blockchain environment of the Ethereum network. However, before implementing the DAO, it is necessary to build a knowledge base, that is, an ontology, which guides its development in a safe and adequate way. A preliminary version of this knowledge base was manually built using Protégé with over 4,000 axioms.

Keywords internet infrastructure · web3 · ontology · machine learning · blockchain · decentralized computing · decentralized science

1 Introduction

This paper presents the first results of a proposal outlined in a previous article. The proposal involved brainstorming the creation of an environment for teachers, researchers, thinkers, and other stakeholders to govern artificial intelligence (AI) algorithms and data [1].

The main motivation for proposing the creation of this environment was the complexity of the problem. Regulating or establishing rules for AI algorithm developers is a huge task that requires a debate among a diverse group of stakeholders.

*<http://lattes.cnpq.br/7092085044582071>

†<https://fenix.tecnico.ulisboa.pt/homepage/ist13522>

‡<http://lattes.cnpq.br/4008970012663480>

§<http://lattes.cnpq.br/7111526592323456>

This is because it involves ethical, social, human, technical, private and public policy issues, among many other areas of human knowledge. Moreover, it is a long debate that requires proper organization to ensure the persistence of the facts produced by a multitude of stakeholders, who represent humanity as a whole.

A problem of similar complexity can be considered: the protocols that ensure the proper functioning and dynamic nature of the Internet over time (past, present, and future). This example comes from the ecosystem surrounding the Internet Engineering Task Force (IETF). Thousands of stakeholders meet in person three times a year and continue to work persistently through email working groups for the rest of the year. It is a large and effective organization that ensures the functioning of the Internet as we know it today.

The Internet, with its undeniable importance to humanity, justifies the existence of its own ecosystem [2]. Governance of AI algorithms is a more complex issue than the Internet. Records show that governance of AI algorithms is indeed more complex than the Internet because it involves implications that are not only offensive but also deadly to human beings. In general, regardless of the application, these systems are considered a black box, resulting in asymmetric information between their developers and their consumers [3]. One of the saddest examples highlighting the consequence of this asymmetry is the design of the MCAS⁵ system on the Boeing 737 MAX, which led to two accidents with 346 deaths in October 2018 (Lion Air) and March 2019 (Ethiopian Airlines). When the angle of attack sensor failed, the built-in algorithms forced the plane to lower its nose, resisting repeated attempts by confused pilots to turn the nose up. Ben Shneiderman, in his book *Human-Centered AI* [4], comments on the two Boeing 737 MAX crashes and considers that the future of these AI algorithms is human-centered. They should primarily become super tools that amplify human abilities and empower people in remarkable ways while ensuring human control. Ben named these algorithms HCAI, an acronym for the title of his book.

There are countless other applications, both using AI and not, that behave disproportionately. A detailed description of the so-called algorithmic biases can be found in Safiya Noble's book, *Algorithms of Oppression*, and in other sources [5] [6] [7].

Asymmetric information, biases, and other pertinent issues are concerning developers, researchers, and other interested parties as they try to figure out what is missing [8]. Perspectives associated with ethics [9] [10] [11] [12] [13], regulations [14] [15] [16] [17] [18], governance [19] [20] [21] [22] [23] [3] [24] and many other issues [25] [26] [27] [28] [29] [30] [31] [32] [33] are on the agenda of all stakeholders in search of appropriate alternatives. For example, these issues are widely discussed in Shneiderman's book *Human-Centered AI* [4]. There is extensive literature on the topic presented in the unpublished work that gave rise to the approaches in this article [1].

If we had as strong a motivation as the Internet does, with its immense global reach, the IETF model with its broad stakeholder participation (as seen in Figure 1) could be a solution that would certainly address the issues involving algorithm and data governance.

There are recent signs in AI activities that suggest a growing understanding of the technology by the world's population. This is due to the immense success of ChatGPT, a tool that is popularizing AI and expanding its prominence. This has occurred at an unimaginable speed despite ChatGPT's own limitations in natural language understanding, its reliance on training data, and the possibility of biases. In addition to ChatGPT and other generative AIs, there is the fact that they can propagate biases. Generative AIs are influenced by their training data, which can lead to biased or discriminatory responses. The danger increases with the possibility of the training data being influenced during interaction with its users, i.e., groups with ulterior motives can influence the training data. The refinement of generative AIs is leading to an expansion of their capabilities, including the use of multimodal resources such as ChatGPT-4⁶ and DALL-E 2⁷. In fact, the community involved believes that after such recent releases, we are in for a significant shift in AI. Bill Gates, for example, reacts through a seven-chapter document entitled "The Age of AI has begun" [35]. In Chapter 7, he concludes with "The Age of AI is filled with opportunities and responsibilities." Other concerns come from neuroscientist Miguel Nicolelis. In his most recent book, "The true creator of everything: How the human brain sculpted the universe as we know it" [36], he exposes in the final two chapters the serious risks that humanity will face in the coming years as a result of our increasing interaction and dependence on digital systems. This establishes a true symbiosis that can deeply affect the brain through the phenomenon of neural plasticity. Basically, almost continuous coexistence with computers can affect the way the brain works and, in the limit, turn us into mere digital zombies. Moreover, one cannot forget the stories that are told again and again about famous and recent generative AIs [37].

This work proposes the creation of a DAO called GHAI DAO⁸ as an original governance mechanism. To develop this proposal, the authors created a knowledge base about DAOs, which is available, including its updates, in a

⁵Acronym for Manoeuvring Characteristics Augmentation System

⁶<https://openai.com/product/gpt-4>

⁷<https://openai.com/product/dall-e-2>

⁸<https://ghaia.pt>

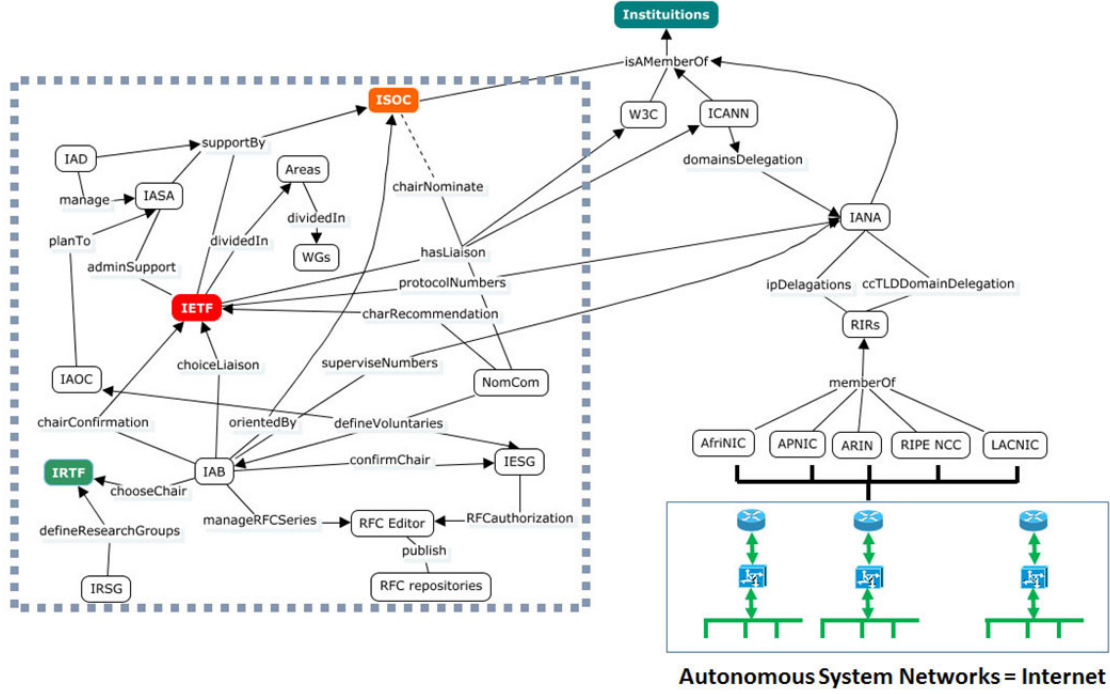


Figure 1: Internet Governance Ecosystem. Source: [34]

public environment of the Open Science Framework (OSF) [38]. Additionally, this work proposes the creation of the GHAIA DAO to support an environment for debate and registration of stakeholder opinions on issues involving the regulation of AI algorithms and data, as well as issues involving human-algorithm interactions. These issues are widely discussed in the Cátedra Oscar Sala Chair⁹ at IEA-USP¹⁰, whose holder is Prof. Dr. Virgílio Almeida (ORCID: 0000-0001-6452-0361¹¹).

In addition to this Introduction section, this paper discusses DAOs and their variety of governance in Section 2. In Section 2.1 the knowledge base built on Protégé, a free and open-source ontology editor and framework for building intelligent systems [39], is presented. This section also displays alternatives for using this knowledge base, in particular, guidance on the use of the *SPARQL* language as a tool for ontology searches [40, 41]. In Section 5, the proposal for the creation of the GHAIA DAO is presented, which includes an original governance mechanism. Section 8 addresses the conclusions of this stage of the work and recommends future activities to follow. Finally, the bibliography is provided.

2 DAOs

A Decentralized Autonomous Organization (DAO) is a form of organization based on blockchain technology that is generally governed by its members, who hold tokens [42]. Tokens, a type of cryptocurrency (among other meanings), can be acquired or received in some form by any person. As the owner of these tokens, the person gains the right to vote on matters directly related to the governance of the DAO. The governance rules of DAOs are characterized through computer programs known as smart contracts, which are executed and validated within the blockchain of the Ethereum network through a resource called the Ethereum Virtual Machine (EVM). The features of smart contracts, such as a distributed blockchain database, cause the rules of the organization to be enforced by the very code that defines the DAO, thus making it self-governed.

Therefore, DAOs are different from traditional organizations because they are self-governing and function autonomously in a decentralized manner without the need for intermediaries. In contrast, traditional organizations are subject to rights and responsibilities defined by the legal system of the environment in which they operate.

⁹<https://bit.ly/cos-usp>

¹⁰<http://www.iea.usp.br/>

¹¹<https://orcid.org/0000-0001-6452-0361>

2.1 Getting to know implemented DAOs in detail through ontologies

There are many types and functions performed by DAOs. As of April 2023, approximately 150 DAOs have been implemented on the Ethereum network. Due to the diversity of implementations, it was decided to create a knowledge base (or ontology), referred to as KB, to understand and maintain permanent information about all the DAOs. To create this knowledge base, the Protégé software was used [39]. Protégé, developed by Stanford University¹², is a free and open-source ontology editor and knowledge management system.

Two ontologies were created using different techniques. Both are available in the public environment of the project on the Open Science Framework (OSF) [38]. We separated from these two, the one that best represented the expected knowledge. It was named *decom.ttl*, and is detailed in Figure 2.

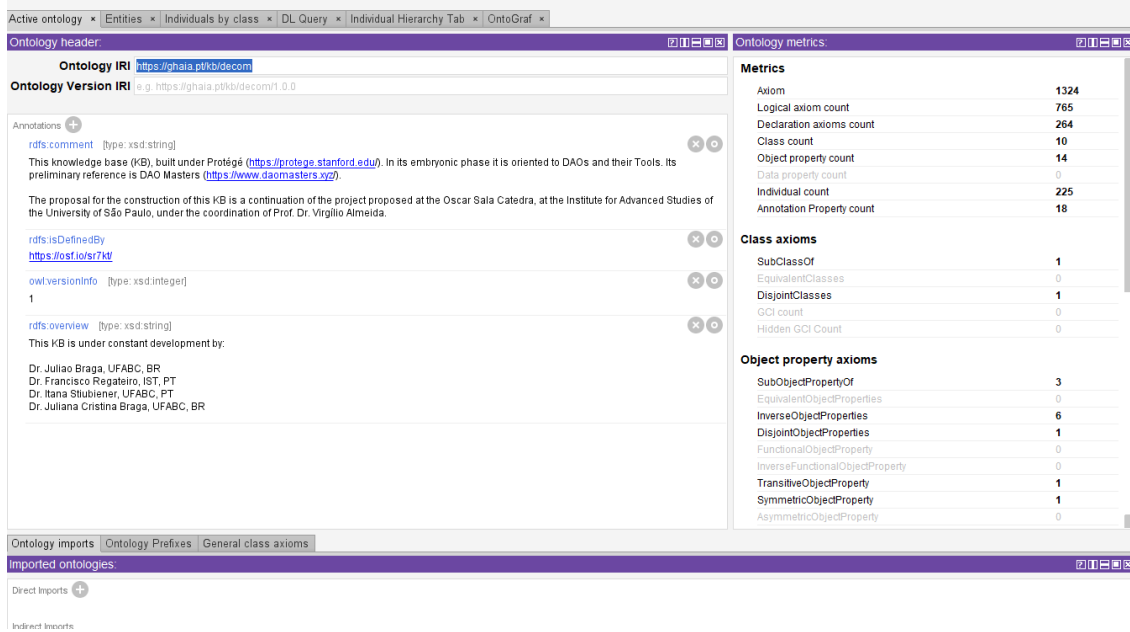


Figure 2: Features of the decom.owl ontology, with 4,004 axioms

The *.ttl* extension, named Turtle, is a syntax for expressing the set of axioms represented in OWL in a text file and is appropriate for display on the Web [43].

To develop the ontology, its scope in the domain of DAOs was first identified. The figures 3, 4 and 5 characterize the main terms of the ontology proposed.

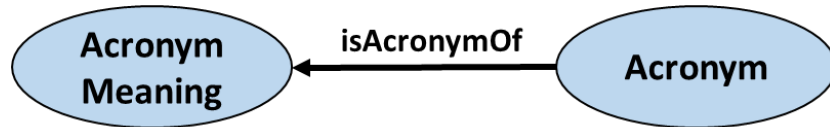


Figure 3: Acronyms that will appear in the development of the proposed ontology.

The ontology resulting from these studies is available on the GHAIA DAO website¹³.

3 How to Search the Knowledge Base

Once the KB is complete with a substantial number of axioms, the main human interest turns to searching the KB. One tool for this is the well-known SPARQL Protocol and RDF Query Language (SPARQL) [44] [45] [46]. Protégé

¹²<https://protege.stanford.edu/>

¹³<https://ghaia.pt/kb/decom.ttl>

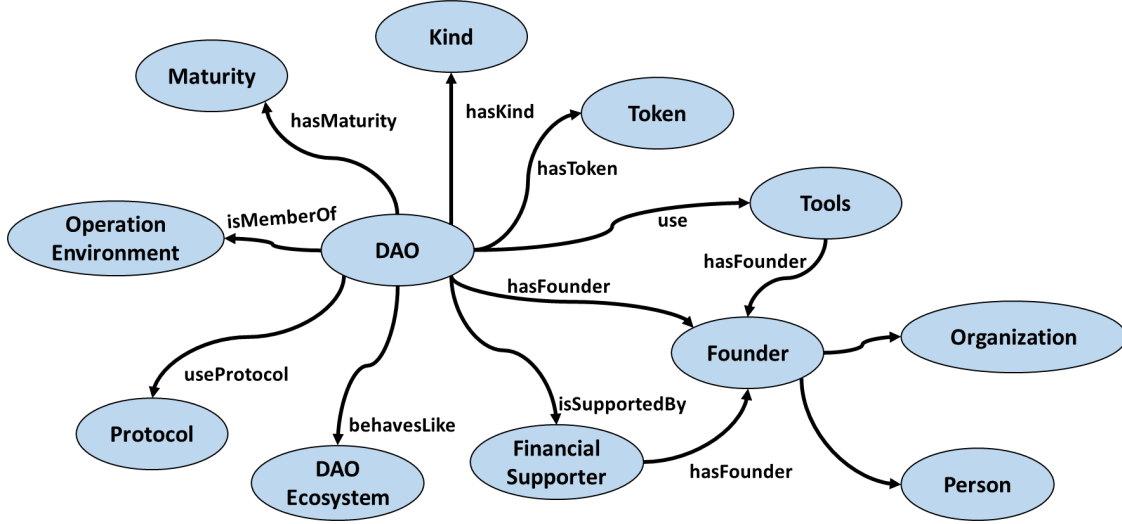


Figure 4: Ontology to be implemented in Protégé.

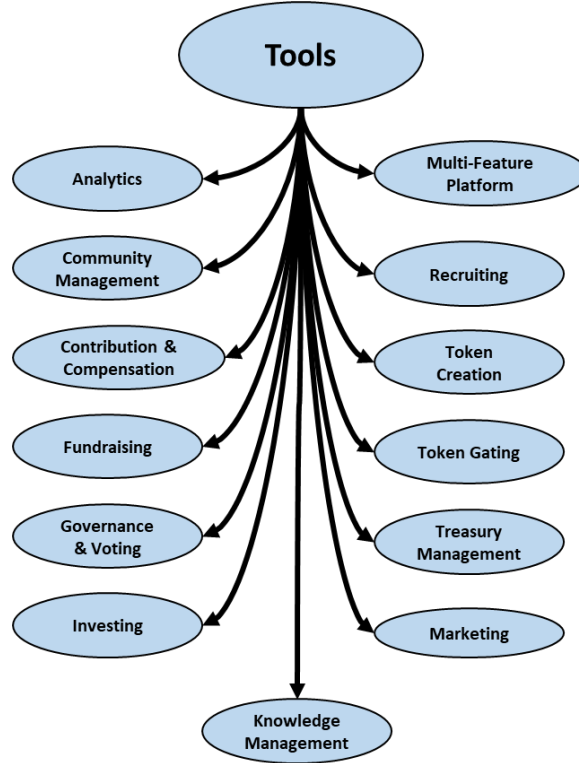


Figure 5: Complement to the Ontology in the figure above, identifying the various tools for creating DAOs.

provides facilities for using SPARQL, as does Apache Jena¹⁴. In addition to these two examples, DBpedia¹⁵ and Wikidata¹⁶ provide public interfaces to SPARQL, among many others. In all the tools used except Protégé, the URLs <https://ghaia.pt/kb/decom.owl> or <https://ghaia.pt/kb/decom.ttl> are used as the entry point. In Protégé, SPARQL acts on the loaded ontology.

¹⁴<https://jena.apache.org/tutorials/sparql.html>

¹⁵<http://virtuoso.openlinksw.com/dataspace/doc/dav/wiki/Main/VOSSPARQL>

¹⁶<https://w.wiki/rL>

Protégé offers other resources to search the ontology produced through it. For example, the OntoGraph provides images such as the one shown in Figure 6.

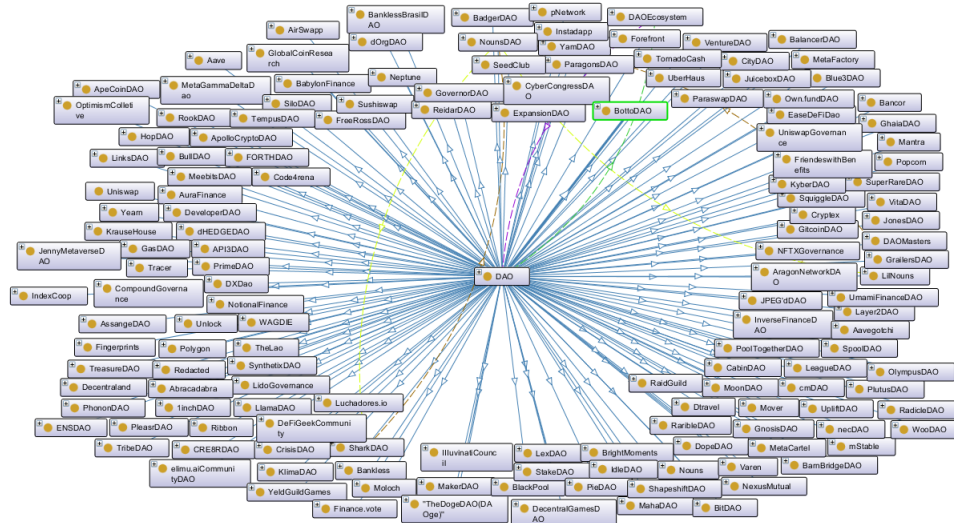


Figure 6: Graphical view of knowledge base DAOs

4 An AI Algorithms and their biases

The Catedra Oscar Sala focused on “Algorithms, artificial intelligence (AI), robots and machines operated by algorithms that increasingly mediate our social, cultural, economic and political interactions” in its Human-Algorithm Interactions Project.

For the Catedra¹⁷, "there are three basic motivations for the proposed focus of the Catedra Oscar Sala in 2022/23:

1. There are many different types of algorithms in operation that play an increasing role in society and in the daily activities of citizens.
2. The complexity of algorithms and systems integrating multiple algorithms is increasing rapidly. New models and massive amounts of data make these algorithms and systems opaque, making it very difficult to understand their behavior.
3. Understanding the social, political, and economic impacts, whether positive or negative, is a research challenge.

Prof. Virgilio anticipated a global concern over opaque algorithms, mostly those of AI. This has proven to be an issue of high priority for humanity, culminating very recently with a manifesto produced by the Future of Life Institute regarding concerns over ChatGPT (Chat Generative Pre-trained Transformer) [47]. The manifesto proposes a six-month moratorium on training it and other similar algorithms, especially those using large language models [48].

Books, scientific papers, newspaper articles, and many other forms of expression have laid out their opinions, concerns, and recommendations to counter the unregulated invasion of algorithms with biases of all kinds. There have been serious offenses and even criminal offenses. However, the texts before the manifesto described in the previous paragraph were intense.

In addition to the above documents, many others, especially recent ones with their references, address the same issue [49] [50] [51] [52] [53] [54] [55] [56] [57] [58] [59] [60] [61].

The proposal of the Catedra Oscar Sala for activities for the year 2022/2023 showed its immense academic diversity. In the opinion of the authors, it is a subject of quite high complexity deserving special attention and continuous debate with the participation of a large number of stakeholders. An adequate form of governance, copying the proposal implemented in the IETF by the Internet Society (ISOC), is suggested.

¹⁷<https://bit.ly/cosvirgilioalmeida>

Therefore, based on ISOC’s concerns about the ecosystem that governs the Internet, the authors propose the creation of a DAO named GHAIA DAO, which is presented in the following sections.

5 GHAIA DAO Constitution

One of the main concerns of a DAO is its governance, which must be efficient and democratic [62] [63]. The first DAOs set out to admit that their governance was done by their members, who held the tokens. A DAO member would have a number of votes equivalent to the number of tokens they owned. As time went on, it became apparent that those interested in controlling the votes and inducing the governance to work the way they wanted could easily do this by buying enough tokens. This began to happen, and such participants were called whales.

In an attempt to improve the governance process, a voting scheme was used in which each participant would have one vote. The whales were unrelenting and created phantom participants or used proxies to increase their participation power. A new scheme was adopted, the so-called quadratic scheme, ensuring that voting would be decentralized [64]. In this proposal, voting continues per member, but each member receives a number of votes equal to twice the number of their tokens, which can additionally be used in voting. Suppose that member *A* has 5 tokens and member *B* has 10 tokens. Then, *A* will have 10 votes that they can use 5 as credits to vote on one proposal and 5 credits to vote on another proposal. Member *B* will have 20 votes and can use 10 credits for one proposal and 10 credits for another proposal. The member who has more tokens can spend their votes on a proposal, but if the other member has more members supporting their proposal, they can win the vote. Quadratic voting allows users to “pay” for additional votes on a given proposal to more strongly express their support for certain issues. This results in voting outcomes aligned with the highest willingness to participate (or pay), rather than just the outcome preferred by the majority, regardless of the intensity of individual preferences. This question about democracy exercised in a DAO leads to the fundamental dilemmas of societies with their paradoxes and behavior of the individuals who constitute them, particularly in those who live in societies in which democratic institutions function [65].

Other DAOs, adopting one of the above criteria, establish a Governance Council that will look after their governance for a previously agreed period.

6 The GHAIA DAO

DeSci (Decentralized Science) is a recent movement that aims to use new technologies, such as blockchain or the Web3 environment¹⁸, to address some of the problematic points of scientific research, silos, and bottlenecks. It is an open and global alternative to the modern scientific system that faces many challenges. It extends the idea of open science, allowing scientists to raise funds, share experimental data, and get ideas. One of the most interesting examples is that of a DeSci to tailor peer review [66] [67] [66] [68] [69] [70].

The GHAIA DAO is a DeSci and will use ORCID ID for its governance, which is described in the following section.

6.1 ORCID

ORCID stands for Open Researcher and Contributor ID and is a global, non-profit, fee-supported organization of its member organizations. They form a community built and governed by a Board of Directors representative of members with broad stakeholder representation [71].

The ORCID ID is a unique, persistent, free identifier for individuals to use while engaging in research, scholarship, and innovation activities. ORCID offers a set of Application Programming Interfaces (APIs).

As of early April 2023, the statistics of ORCID¹⁹ indicated something like nine million four hundred and ten researchers enrolled, spread across 56 countries. Brazil was the third country with the most registrants (361,900), after the United States (794,493) and China (412,925).

6.2 The Governance of the GHAIA DAO

When implemented, the GHAIA DAO will be a DeSci that provides a multi-disciplinary discussion environment among stakeholders in the issues surrounding algorithmic human interaction. It will follow, in part, the IETF and IRTF²⁰ (Internet Research Task Force) discussion model. The following rules will model the GHAIA DAO:

¹⁸Not a very acceptable generic name given to the blockchain

¹⁹<https://info.orcid.org/orcid-statistics/>

²⁰<https://irtf.org>

1. It will have two types of tokens: OR and NOR, both initially with a value of one (1) USD.
2. Its members will be of two types: those who have an ORCID and those without an ORCID.
3. Stakeholders with an ORCID will receive 1 OR free of charge. The OR is equivalent to 1 NOR, which on the implementation date of the GAIA DAO will be equivalent to one (1) USD. Stakeholders without an ORCID need to purchase NORs at market value to become members.
4. One OR is equivalent to one NOR. ORs are not tradable, but NORs are freely tradable.
5. The holder of an OR will be entitled to receive a NOR twenty-one (21) months after having received the OR without losing the right to vote.
6. The GHAIA DAO treasury must have collateral equivalent to the value of the number of NORs distributed in ORs. In other words, you cannot distribute an OR without having the equivalent collateral in the GHAIA DAO treasury.
7. Participants holding an OR can vote and can be voted for a twenty-one (21) member board, which will handle the governance of the GHAIA DAO.
8. Participants who do not hold an OR, that is, do not have an ORCID, do not have voting rights but can be voted for.
9. Participants with an OR also vote for the Control Council, consisting of seven (7) members whose purpose is to ensure that there are no excesses on the part of the Governance Council. See Figure 7.
10. Outside the blockchain, the Governance Board, through Technical Support, will maintain email lists equivalent to the IETF/IRTF working groups (WGs²¹) and other similarities to make effective debate around Human-Algorithm Interactions.
11. The GHAIA DAO Internet environment should host a repository of documents equivalent to IETF RFCs (Request for Comments) that will be developed by its members.
12. Other rules related to the social and ethical behavior of both participants should be defined.
13. Technical and Operational Support is composed of technical, administrative and other personnel who are adequately remunerated.

Figure 7 displays the proposed governance for the GHAIA DAO.

This figure abstracts from implementation details on the Ethereum network and the Internet resources outside the blockchain that are necessary to meet the goals of the GHAIA DAO.

7 Related Literature

Table 1 references the literature used to understand the mechanism of algorithm and data governance and allows for a comparison of the recommended proposals. Most of these references were originally collected in the proposal prepared as a requirement of the preliminary phase of the Oscar Sala Chair and not published [1].

The references are classified into seven categories and are not exhaustive in the list presented in this proposal:

- (a). **Internet:** These include references that address the topic of Internet governance.
- (b). **Algorithms:** These are references that display AI algorithms in various application areas.
- (c). **DAO:** References that address DAOs and the respective techniques on which they are built (blockchain and cryptocurrencies).
- (d). **Economics:** References that address issues related to the economics of algorithms and their environments.
- (e). **Others:** A set of references that describe the involvement of AI algorithms in subjects such as Bots, Discrimination, Software Engineering, AI, Games, Robotics, and Security.
- (f). **RLiterature:** Literature review papers, including systematic reviews.
- (g). **Social:** Papers that reference the social, ethical, and philosophical aspects of algorithms.

²¹Work Groups

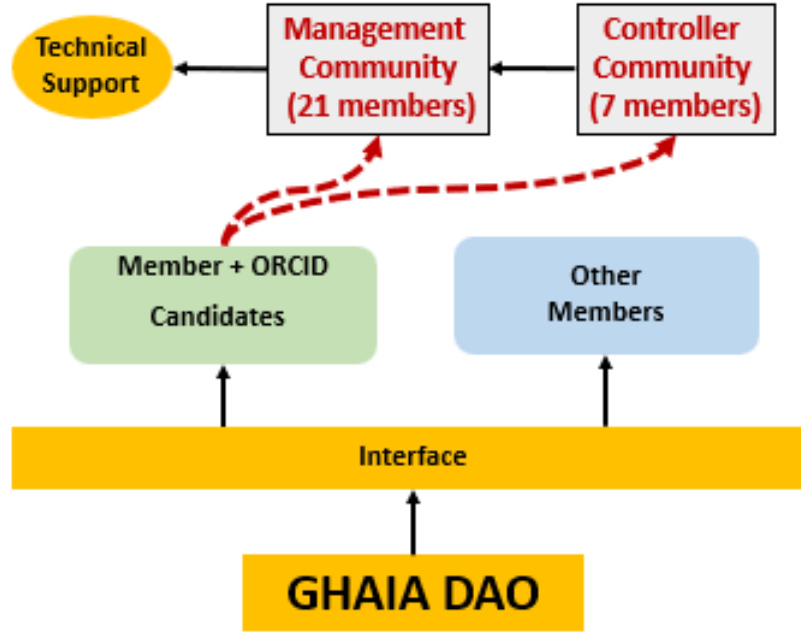


Figure 7: GHAIA DAO's proposed governance structure

8 Conclusions and future work

There is much work to be done to set up the GHAIA DAO. The current unavailability of resources has prevented the implementation of the GHAIA DAO, but it is considered a momentary hindrance. The KB of DAOs proved to be an appropriate solution for learning about DAOs while simplifying their presentation.

In the future, the following tasks were considered mandatory:

- i. ORCID should be informed of the intentions to use the ORCID ID to identify future voting members of the GHAIA DAO.
- ii. The economic model of the GHAIA DAO's operating structure must be formally constructed prior to its establishment. This model, among many other outcomes, must estimate the safe collateral for its initiation. The base parameter for this formulation at the beginning of April 2023 is the 9,410,000 researchers registered in ORCID, plus the costs involved in maintaining Technical and Operational Support. It is hoped that stakeholders in the project can develop papers in this direction.
- iii. After the economic model has been defined and all the operating rules for the DAO have been established, one or more smart contracts will be developed to ensure the self-governance of the GHAIA DAO.
- iv. DAOs are not yet regulated in many countries, which can create legal uncertainty. Interested parties with expertise in law, particularly international law, should study this issue.
- v. The ontology created using Protégé and stored in *decom.ttl* must be evaluated, checked, and compared with other similar ontologies to ensure its validity. However, the manual process of building an ontology can be time-consuming and exhausting, and may not always produce accurate results. As such, it is recommended to use semi-automatic techniques that involve a combination of manual input and automated processes to develop and update the ontology. These techniques, which are constantly being improved, generally involve the use of deep learning and text capture from the web [462, 463, 464, 465].
- vi. A text detailing the use of SPARQL over the two bases should be developed in tutorial form to spread the work developed and useful for the interested community.
- vii. A companion paper presenting the graphics produced extensively by Protégé is available at the OSF of the project²².

²²<https://bit.ly/daoKBinGraphics>

Table 1: Primary and secondary studies. by subject

#	References	Classification
1.	[72],[73],[74],[75],[76],[77],[78],[79],[80],[81],[82],[83],[84],[85],[86]	Internet
2.	[87],[88],[89],[90],[91],[92],[93],[94],[95],[96],[97],[98],[99],[100],[101],[102]	Algorithms
3.	[103],[104],[105],[106],[107],[108],[109],[110],[111],[112],[113],[114],[115]	
4.	[116],[117],[118],[119],[120],[121],[122],[123],[124],[125],[126],[127],[128]	
5.	[129],[130],[131],[132],[133],[134],[135],[136],[137],[138],[139],[140],[141]	
6.	[142],[143],[144],[145],[146],[147],[148],[149],[150],[151],[152],[153],[154]	
7.	[155],[156],[157],[158],[159],[160],[161],[162],[163],[164],[165],[166],[167]	
8.	[168],[169],[170],[171],[172],[173],[174],[175],[176],[177],[178],[179],[180]	
9.	[181],[182],[183],[184],[185],[186],[187],[188],[189],[190],[191],[192],[193]	
10.	[194],[195],[196],[197],[198],[199],[200],[201],[202],[203],[204],[205],[206]	
11.	[207],[208],[209],[210],[211],[212],[213],[214],[215],[216],[217],[218],[219]	
12.	[220],[221],[222],[223],[224],[225],[226],[227],[228],[229]	
13.	[230],[231],[232],[233],[234],[235],[236],[237],[238],[239],[240],[241],[242]	DAO
14.	[243],[244],[245],[246],[247]	
15.	[248],[249],[250],[251],[252],[253],[254],[255],[256],[257],[258],[259],[260]	Economics
16.	[261],[262],[263],[264],[265],[266],[267],[268],[269],[270],[271],[272],[273]	
17.	[274],[275],[276],[277],[278],[279]	
18.	[280],[281],[282],[283],[284],[285],[286],[287],[288],[289],[290],[291],[292]	Others
19.	[293],[294],[295],[296],[297],[298],[299],[300],[301],[302],[303],[304],[305]	
20.	[306],[307],[308],[309],[310],[311],[312],[313],[314],[315],[316],[317],[318]	
21.	[319],[320],[321],[322],[323],[324],[325],[326],[327],[328],[329],[330],[331]	
22.	[332],[333],[334],[335],[336],[337],[338],[339]	
23.	[340],[341],[342],[343],[344],[345],[346],[347],[348],[349],[350],[351],[352]	Governance
24.	[353],[354],[355],[356],[357],[358],[359],[360],[361],[362],[363],[364],[365]	
25.	[366],[367],[368],[369],[370],[371],[372],[373],[374],[375],[376],[377],[378]	
26.	[379],[380],[381],[382],[383],[384],[385],[386],[387],[388],[389],[390],[391]	
27.	[392],[393],[394],[395],[396],[397],[398],[399],[400],[401],[402],[403],[404]	
28.	[405],[406],[407],[408],[409],[410],[411],[412],[413],[414],[415],[416],[417]	
29.	[418],[419],[420],[421],[422],[423],[424],[425],[426],[427],[428],[429],[430]	
30.	[431],[432],[433]	
31.	[434],[435],[436],[437]	
32.	[438],[439],[440],[441],[442],[443],[444],[445],[446],[447],[448],[449],[450]	RLiterature
33.	[451],[452],[453],[454],[455],[456],[457],[458],[459],[460],[461]	Social

- viii. For the success of the project, it is appropriate to have an unlimited presence of stakeholders from the most varied and immense areas of knowledge, including society in general.
- ix. It is expected that the DAO will host interested parties in extending human-algorithm interactions to the context of all digital platforms.

References

- [1] Juliao Braga, Francisco Regateiro, Itana Stiubiener, and Juliana C Braga. A proposal to improve research in AI algorithm and data governance, Sep 2022. Portuguese version: <https://osf.io/xcpsd>.
- [2] Wolfgang Kleinwachter and Virgilio AF Almeida. The internet governance ecosystem and the rainforest. *IEEE Internet Computing*, 19(2):64–67, 2015.
- [3] Urs Gasser and Virgilio A.F. Almeida. A Layered Model for AI Governance. *IEEE Internet Computing*, 21(6):58–62, 2017.
- [4] Ben Shneiderman. *Human-Centered AI*. Oxford University Press, 2022.
- [5] Safya. Noble and Felipe Damorim. *Algoritmos da Opressão: Como os mecanismos de busca reforçam o racismo*. Editora Rua do Sabão, Rio de Janeiro, 1 edition, 2022.
- [6] Cathy O’Neil, editor. *Algoritmos de Destruição em Massa*. Editora Rua do Sabão, Santo André, SP, 2020.
- [7] Magaly Prado. *Fake news e inteligência artificial: O poder dos algoritmos na guerra da desinformação*, volume 1. Almedina Brasil, 2022.
- [8] M Mitchell Waldrop. What are the limits of deep learning? *Proceedings of the National Academy of Sciences*, 116(4):1074–1077, 2019.

- [9] Katina Michael, Diana Bowman, Meg Leta Jones, and Ramona Pringle. Robots and socio-ethical implications [guest editorial]. *IEEE Technology and Society Magazine*, 37(1):19–21, 2018.
- [10] Andrea Censi, Konstantin Slutsky, Tichakorn Wongpiromsarn, Dmitry Yershov, Scott Pendleton, James Fu, and Emilio Frazzoli. Liability, ethics, and culture-aware behavior specification using rulebooks. In *2019 International Conference on Robotics and Automation (ICRA)*, pages 8536–8542. IEEE, 2019.
- [11] Michael Kearns and Aaron Roth. *The ethical algorithm: The science of socially aware algorithm design*. Oxford University Press, 2019.
- [12] Flávio S. Corrêa Silva and Nina S. T. Hirata. Inteligência Ética. *Computação Brasil*, 7:15–18, 2022.
- [13] Paola Cantarini. Por uma ética. *Revista Juridica*, 4(71):892–912, 2022.
- [14] Christianr Mölle and Arnaud Amouroux, editors. *Governing the Internet: Freedom and Regulation in the OSCE Region*. Organization for Security and Co-operation in Europe (OSCE), 2007.
- [15] Eduardo Bismarck. PL 21/2020, 2020.
- [16] Europe Commission. Coordinated Plan on Artificial Intelligence 2021 Review. Technical report, European Commission, Brussels, 2021.
- [17] Margrethe Vestager and Thierry Breton. Uma Europa Preparada para a Era Digital : Comissão propõe novas regras e ações para promover a excelência e a confiança na inteligência artificial. Technical report, Comissão Europeia, 2021.
- [18] Sociedade Brasileira de Computação. Ética e regulação na inteligência artificial. Technical report, Sociedade Brasileira de Computação, 7 2022.
- [19] Solon Barocas, Sophie Hood, and Malte Ziewitz. Governing Algorithms: A Provocation Piece. *SSRN Electronic Journal*, pages 1–12, 2013.
- [20] Florian Saurwein, Natascha Just, and Michael Latzer. Governance of algorithms: Options and limitations. *Info*, 17(6):35–49, 2015.
- [21] Danilo Doneda and Virgilio A.F. Almeida. What Is Algorithm Governance? *IEEE Internet Computing*, 20(4):60–63, 2016.
- [22] Lucas D. Introna. Algorithms, Governance, and Governmentality: On Governing Academic Writing. *Science Technology and Human Values*, 41(1):17–49, 2016.
- [23] Martin Ebers and Marta Cantero Gamito, editors. *Algorithmic Governance and Governance of Algorithms: Legal and Ethical Challenges*. Springer, 2021.
- [24] Fernanda Bruno. Rastros digitais sob a perspectiva da teoria ator-rede. *Revista Famecos*, 19(3):681–704, 2012.
- [25] Adam D.I. Kramer, Jamie E. Guillory, and Jeffrey T. Hancock. Experimental evidence of massive-scale emotional contagion through social networks. In *Proceedings of the National Academy of Sciences of the United States of America*, volume 111, pages 8788–8790, 2014.
- [26] David Lazer. The rise of the social algorithm. *Science*, 348(6239):1090–1091, 2015.
- [27] Matthew O Jackson. *The human network: How your social position determines your power, beliefs, and behaviors*. Pantheon Books, 2019.
- [28] Eytan Bakshy, Solomon Messing, and Lada A Adamic. Exposure to ideologically diverse news and opinion on facebook. *Science*, 348(6239):1130–1132, 2015.
- [29] Jean-François Bonnefon, Azim Shariff, and Iyad Rahwan. The social dilemma of autonomous vehicles. *Science*, 352(6293):1573–1576, 2016.
- [30] David MJ Lazer, Matthew A Baum, Yochai Benkler, Adam J Berinsky, Kelly M Greenhill, Filippo Menczer, Miriam J Metzger, Brendan Nyhan, Gordon Pennycook, David Rothschild, et al. The science of fake news. *Science*, 359(6380):1094–1096, 2018.
- [31] Mengyi Wei and Zhixuan Zhou. Ai ethics issues in real world: Evidence from ai incident database. *arXiv preprint arXiv:2206.07635*, 2022.
- [32] European Data Protection Supervisor (EDPS). Towards a new digital ethics. Technical Report September, European Organization, 2015.
- [33] Anne Magaly de Paula Canuto. Ética no Uso de Dados Biométricos: Histeria ou uma Preocupação Coerente? *Computação Brasil*, 7:36–39, 2022.

- [34] Juliao Braga, Jeferson Campos Nobre, Lisandro Zambenedetti Granville, and Marcelo Santos. Como Protocolos Inovadores são Criados e Adotados em Escala Mundial: Uma visão sobre o Internet Engineering Task Force (IETF) e a Infraestrutura da Internet. In Taisy Silva Weber and Claudia Aparecida Martins, editors, *Jornadas de Atualização em Informática 2020*, page 45. Sociedade Brasileira de Computação, Cuiabá, MT Brazil, 2020. Available in: <https://doi.org/10.5753/sbc.5728.3.2>.
- [35] Bill Gates. The Age of AI has begun, 2023. Last accessed 23 march 2023.
- [36] Miguel Nicolelis. *O Verdadeiro Criador de Tudo: Como o Cerebro Humano Moldou o Universo Tal Como o Conhecemos*. ELSINORE, Brasil, 2023.
- [37] Bruno Garattoni. GPT-4 tenta assumir o controle de outro computador – e digitar “como escapar” no Google. Super Interessante, Publicado em 23 mar 2023, 15h08, 2023. Last accessed 24 march 2023.
- [38] Juliao Braga, Francisco Regateiro, and Itana Stiubiener. Human-algorithm: Governance, Sep 2022. Accessed in 03/09/2022.
- [39] Mark A Musen. The protégé project: a look back and a look forward. *AI matters*, 1(4):4–12, 2015.
- [40] Steve Harris and Andy Seaborne. SPARQL 1.1 Query Language. W3c working draft, W3C, march 2023. <https://www.w3.org/TR/sparql11-query/>.
- [41] W3C. SPARQL, March 2023. [Online; accessed 24-March-2023].
- [42] Carlos Santana and Laura Albareda. Blockchain and the emergence of decentralized autonomous organizations (daos): An integrative model and research agenda. *Technological Forecasting and Social Change*, 182:1–15, 2022.
- [43] Caroline Burle, Bernadette Farias Loscio, and Newton Calegari. Data on the web best practices. W3C recommendation, W3C, January 2017. <https://www.w3.org/TR/2017/REC-dwbp-20170131/>.
- [44] Andy Seaborne and Steven Harris. SPARQL 1.1 query language. W3C recommendation, W3C, March 2013. <https://www.w3.org/TR/2013/REC-sparql11-query-20130321/>.
- [45] Bob DuCharme. *Learning SPARQL: querying and updating with SPARQL 1.1*. " O'Reilly Media, Inc.", 2013.
- [46] Jędrzej Potoniec. Learning sparql queries from expected results. *Computing and Informatics*, 38(3):679–700, 2019.
- [47] OpenAI. Gpt-4 technical report. *arXiv*, 2023.
- [48] Future of Life Institute. Pause giant ai experiments: An open letter, 2023.
- [49] Emily M Bender, Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell. On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? In *Proceedings of the 2021 ACM conference on fairness, accountability, and transparency*, pages 610–623, 2021.
- [50] Edd Gent. A cryptocurrency for the masses or a universal id?: Worldcoin aims to scan all the world’s eyeballs. *IEEE Spectrum*, 60(1):42–57, 2023.
- [51] Nick Bostrom. *Superintelligence: Paths, Dangers, Strategies*. Oxford University Press, 2016.
- [52] Benjamin S Bucknall and Shiri Dori-Hacohen. Current and near-term ai as a potential existential risk factor. In *Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society*, pages 119–129, July 2022.
- [53] Joseph Carlsmith. Is power-seeking ai an existential risk? *arXiv preprint arXiv:2206.13353*, 2022.
- [54] Brian Christian. *The Alignment Problem: Machine Learning and Human Values*. Norton & Company, 2020.
- [55] Michael Cohen, Marcus Hutter, and M Osborne. Advanced artificial agents intervene in the provision of reward. *AI Magazine*, 43(3):282–293, 2022.
- [56] T Eloundou et al. Gpts are gpts: An early look at the labor market impact potential of large language models. *arXiv preprint arXiv:2303.10130*, 2023.
- [57] Dan Hendrycks and Mantas Mazeika. X-risk analysis for ai research. *arXiv preprint arXiv:2206.05862*, 2022.
- [58] Richard Ngo. The alignment problem from a deep learning perspective. *arXiv preprint arXiv:2209.00626*, 2022.
- [59] Stuart Russell. *Human compatible: Artificial intelligence and the problem of control*. Penguin, 2019.
- [60] Max Tegmark. *Life 3.0: Being Human in the Age of Artificial Intelligence*. Knopf, 2017.
- [61] Lukas Weidinger et al. Ethical and social risks of harm from language models, 2021.
- [62] Galia Kondova and Renato Barba. Governance of decentralized autonomous organizations. *Journal of Modern Accounting and Auditing*, 15(8):406–411, 2019.

- [63] Usman W Chohan. The decentralized autonomous organization and governance issues. *Available at SSRN* 3082055, pages 1–16, 2017.
- [64] Steven P Lalley and E Glen Weyl. Quadratic voting: How mechanism design can radicalize democracy. In *AEA Papers and Proceedings*, volume 108, pages 33–37, 2018.
- [65] Paulo Trigo Pereira. *Prisioneiro, o amante e as sereias: instituições econômicas, políticas e democracia*. Almedina, 2008.
- [66] Bianca Trovò and Nazzareno Massari. Ants-review: A privacy-oriented protocol for incentivized open peer reviews on ethereum. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, volume 12480 LNCS, 2021.
- [67] Sarah Hamburg. Call to join the decentralized science movement, 2021.
- [68] Richard Walker and Pascal Rocha da Silva. Emerging trends in peer review - a survey. *Frontiers in Neuroscience*, 9, 2015.
- [69] Richard Smith. Peer review: A flawed process at the heart of science and journals. *Journal of the Royal Society of Medicine*, 99, 2006.
- [70] Robert E. Gropp, Scott Glisson, Stephen Gallo, and Lisa Thompson. Peer review: A system under stress. In *BioScience*, volume 67, 2017.
- [71] ORCID. Open Researcher and Contributor ID, 2023. <https://info.orcid.org/what-is-orcid/>.
- [72] Tom Standage. *The Victorian Internet: The remarkable story of the telegraph and the nineteenth century's online pioneers*. Phoenix London, 1998.
- [73] Janet Abbate. *Inventing the Internet*. MIT Press, 1999.
- [74] Michele Banko and Eric Brill. Scaling to very very large corpora for natural language disambiguation. In *Proceedings of the 39th annual meeting of the Association for Computational Linguistics*, pages 26–33, 2001.
- [75] Lee A. Bygrave and Jon Bing. *Internet Governance: Infrastructure and Institutions*. Oxford University Press, New York, 2009.
- [76] Wolfgang Kleinwachter. The History of Internet Governance. In *Governing the Internet: Freedom and Regulation in the OSCE Region*, pages 41–65. OSCE Region, Vienna, 2007.
- [77] Christian Moller. Governing the Domain Name System: An Introduction to Internet Infrastructure. In *Governing the Internet: Freedom and Regulation in the OSCE Region*, pages 29–39. OSCE, Vienna, 2007.
- [78] Diego Rafael Canabarro and Flavio Rech. A Governança da Internet: Definição, Desafios e Perspectivas. In *9o ENCONTRO DA ABCP*, page 17, 2014.
- [79] Diego Rafael Canabarro. *Governança global da internet: tecnologia, poder e desenvolvimento*. Doutoral thesis, Federal University of Rio Grande do Sul, 2014.
- [80] Jovan Kurbalija. *An introduction to internet governance*. Diplo Foundation, 2016.
- [81] Alexandre Arns Gonzales. *Quem Governa a Governança da Internet? Uma análise do papel da Internet sobre os rumos do sistema-mundo*. Dissertação de mestrado, Universidade Federal do Rio Grande do Sul, 2016.
- [82] Fabrício Pasquot Bertini Polido and Lucas Costa Dos Anjos, editors. *Marco Civil E Governança Da Internet: Diálogos Entre O Doméstico E O Global*. Faculdade de Direito da UFMG, 2016.
- [83] Fabrício Bertini Pasquot Polido, Lucas Cosda dos Anjos, and Luíza Couto Chaves Brandão, editors. *Tecnologias e Conectividade: Direito e Políticas na Governança das Redes*. IRIS, 2017.
- [84] Lucas Andrade, Juliao Braga, Stefany Pereira, Rafael Roque, and Marcelo Santos. In-Person and Remote Participation Review at IETF. In *Proceeding of CSBC 2018 - V Workshop pre IETF*, page 11, Natal, RN Brazil, July 2018. To be published. Available at: <http://braga.net.br/papers/In-Person%20and%20Remote%20Participation%20Review%20at%20IETF.pdf>.
- [85] Alexandre Pacheco da Silva, Ana Paula Camelo, Diego R. Canabarro, and Flavio Rech Wagner, editors. *Estrutura e funcionamento da internet : aspectos técnicos, políticos e regulatórios*. FGV, 2021.
- [86] Elizabeth Machado Veloso. Legislação sobre Internet no Brasil. Technical report, Camara dos Deputados, 2009.
- [87] Warren S McCulloch and Walter Pitts. A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, 5(4):115–133, 1943.
- [88] Jürgen Schmidhuber. Deep learning in neural networks: An overview. *Neural Networks*, 61:85–117, 2015.
- [89] Janet Abbate*. L’histoire de l’Internet au prisme des STS. *Le temps des médias*, 18(1):170–180, 2012.

- [90] Martin Adam, Michael Wessel, and Alexander Benlian. Ai-based chatbots in customer service and their effects on user compliance. *Electronic Markets*, 31(2):427–445, 2021.
- [91] Nadav Aharony, Wei Pan, Cory Ip, Inas Khayal, and Alex Pentland. Social fmri: Investigating and shaping social mechanisms in the real world. *Pervasive and mobile computing*, 7(6):643–659, 2011.
- [92] Katti Faceli, Ana Carolina Lorena, João Gama, and André Carlos Ponce de Leon Ferreira de Carvalho. *Inteligência artificial: uma abordagem de aprendizado de máquina*. LTC, Rio de Janeiro, RJ, Brasil, 2011.
- [93] Saar Alon Barkat and Madalina Busuioc. Human-ai interactions in public sector decision-making: Automation bias and selective adherence to algorithmic advice. *Accepted Manuscript*, 2022.
- [94] Saleema Amershi, Dan Weld, Mihaela Vorvoreanu, Adam Fourney, Besmira Nushi, Penny Collisson, Jina Suh, Shamsi Iqbal, Paul N Bennett, Kori Inkpen, et al. Guidelines for human-ai interaction. In *Proceedings of the 2019 chi conference on human factors in computing systems*, pages 1–13, 2019.
- [95] Saleema Amershi, Andrew Begel, Christian Bird, Robert DeLine, Harald Gall, Ece Kamar, Nachiappan Nagappan, Besmira Nushi, and Thomas Zimmermann. Software engineering for machine learning: A case study. In *2019 IEEE/ACM 41st International Conference on Software Engineering: Software Engineering in Practice (ICSE-SEIP)*, pages 291–300. IEEE, 2019.
- [96] Dario Amodei, Chris Olah, Jacob Steinhardt, Paul Christiano, John Schulman, and Dan Mané. Concrete problems in ai safety. *arXiv preprint arXiv:1606.06565*, 2016.
- [97] Julia Angwin, Jeff Larson, Surya Mattu, and Lauren Kirchner. Machine Bias, 2016.
- [98] George A. Akerlof and Rachel E. Kranton. *Identity Economics: How Our Identities Shape Our Work, Wages, and Well-Being*. Princeton University Press, Princeton, 1 edition, 2010.
- [99] Kenneth C Arnold, Krysta Chauncey, and Krzysztof Z Gajos. Predictive text encourages predictable writing. In *Proceedings of the 25th International Conference on Intelligent User Interfaces*, pages 128–138, 2020.
- [100] Edmond Awad, Sohan Dsouza, Richard Kim, Jonathan Schulz, Joseph Henrich, Azim Shariff, Jean-François Bonnefon, and Iyad Rahwan. The moral machine experiment. *Nature*, 563(7729):59–64, 2018.
- [101] Lisanne Bainbridge. Ironies of automation. In *Analysis, design and evaluation of man-machine systems*, pages 129–135. Elsevier, 1983.
- [102] Reuben Binns, Max Van Kleek, Michael Veale, Ulrik Lyngs, Jun Zhao, and Nigel Shadbolt. 'it's reducing a human being to a percentage' perceptions of justice in algorithmic decisions. In *Proceedings of the 2018 Chi conference on human factors in computing systems*, pages 1–14, 2018.
- [103] Tolga Bolukbasi, Kai-Wei Chang, James Y Zou, Venkatesh Saligrama, and Adam T Kalai. Man is to computer programmer as woman is to homemaker? debiasing word embeddings. *Advances in neural information processing systems*, 29, 2016.
- [104] Josh Bongard, Victor Zykov, and Hod Lipson. Resilient machines through continuous self-modeling. *Science*, 314(5802):1118–1121, 2006.
- [105] Tawfiq Ammari, Jofish Kaye, Janice Y Tsai, and Frank Bentley. Music, search, and iot: How people (really) use voice assistants. *ACM Trans. Comput. Hum. Interact.*, 26(3):17–1, 2019.
- [106] N Bolstrom. *Superintelligence. Paths, dangers, strategies*. Oxford University Press, United Kingdom, 2014.
- [107] Jeffrey M Bradshaw, Robert R Hoffman, David D Woods, and Matthew Johnson. The seven deadly myths of "autonomous systems". *IEEE Intelligent Systems*, 28(3):54–61, 2013.
- [108] Elizabeth Broadbent. Interactions with robots: The truths we reveal about ourselves. *Annual review of psychology*, 68(1):627–652, 2017.
- [109] Connor Brooks and Daniel Szafrir. Visualization of intended assistance for acceptance of shared control. In *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 11425–11430. IEEE, 2020.
- [110] Connor Brooks and Daniel Szafrir. Balanced information gathering and goal-oriented actions in shared autonomy. In *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 85–94. IEEE, 2019.
- [111] Erik Brynjolfsson and Tom Mitchell. What can machine learning do? Workforce implications. *Science*, 358(6370):1530–1534, 2017.
- [112] Joy Buolamwini and Timnit Gebru. Gender shades: Intersectional accuracy disparities in commercial gender classification. In *Conference on fairness, accountability and transparency*, pages 77–91. PMLR, 2018.

- [113] Aylin Caliskan, Joanna J Bryson, and Arvind Narayanan. Semantics derived automatically from language corpora contain human-like biases. *Science*, 356(6334):183–186, 2017.
- [114] Rafael A Calvo, Dorian Peters, Karina Vold, and Richard M Ryan. Supporting human autonomy in ai systems: A framework for ethical enquiry. In *Ethics of Digital Well-Being*, pages 31–54. Springer, 2020.
- [115] Neil A. H. Campbell. The Evolution of Flight Data Analysis. In *Proceedings of Australian Society of Air Safety Investigators*, pages 1–22, 2003.
- [116] Juliano Cappi. *Internet, Big Data e discurso de ódio: reflexões sobre as dinâmicas de interação no Twitter e os novos ambientes de debate político*. Doctoral thesis, Pontifícia Universidade Católica de São Paulo, 2017.
- [117] Felix Carros, Johanna Meurer, Diana Löffler, David Unbehaun, Sarah Matthies, Inga Koch, Rainer Wieching, Dave Randall, Marc Hassenzahl, and Volker Wulf. Exploring human-robot interaction with the elderly: results from a ten-week case study in a care home. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pages 1–12, 2020.
- [118] Miles Brundage, Shahar Avin, Jasmine Wang, Haydn Belfield, et al. Toward Trustworthy AI Development: Mechanisms for Supporting Verifiable Claims. *arXiv preprint*, 2020.
- [119] Radhika Chemuturi, Farshid Amirabdollahian, and Kerstin Dautenhahn. Adaptive training algorithm for robot-assisted upper-arm rehabilitation, applicable to individualised and therapeutic human-robot interaction. *Journal of NeuroEngineering and Rehabilitation*, 10(1), 2013.
- [120] Bo Chen, Chunsheng Hua, Bo Dai, Yuqing He, and Jianda Han. Online control programming algorithm for human–robot interaction system with a novel real-time human gesture recognition method. *International Journal of Advanced Robotic Systems*, 16(4):1–18, 2019.
- [121] Lu Cheng, Kush R Varshney, and Huan Liu. Socially responsible ai algorithms: Issues, purposes, and challenges. *Journal of Artificial Intelligence Research*, 71:1137–1181, 2021.
- [122] Alexandra Chouldechova, Diana Benavides-Prado, Oleksandr Fialko, and Rhema Vaithianathan. A case study of algorithm-assisted decision making in child maltreatment hotline screening decisions. In *Conference on Fairness, Accountability and Transparency*, pages 134–148. PMLR, 2018.
- [123] Paul F. Christiano, Jan Leike, Tom B. Brown, Miljan Martic, Shane Legg, and Dario Amodei. Deep reinforcement learning from human preferences. In *Advances in Neural Information Processing Systems*, pages 4300–4308, 2017.
- [124] Jacob W Crandall, Mayada Oudah, Fatimah Ishowo-Oloko, Sherief Abdallah, Jean-François Bonnefon, Manuel Cebrian, Azim Shariff, Michael A Goodrich, Iyad Rahwan, et al. Cooperating with machines. *Nature communications*, 9(1):1–12, 2018.
textbfThis study examines algorithmic cooperation with humans and provides an example of methods that can be used to study the behaviour of human–machine hybrid systems.
- [125] Antoine Cully, Jeff Clune, Danesh Tarapore, and Jean-Baptiste Mouret. Robots that can adapt like animals. *Nature*, 521(7553):503–507, 2015.
- [126] Nilesh Dalvi, Pedro Domingos, Sumit Sanghai, and Deepak Verma. Adversarial classification. In *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 99–108, 2004.
- [127] Rajarshi Das, James E Hanson, Jeffrey O Kephart, and Gerald Tesauro. Agent-human interactions in the continuous double auction. In *International Joint Conference on Artificial Intelligence*, volume 17, pages 1169–1178. Lawrence Erlbaum Associates Ltd, 2001.
- [128] Jesse Davis and Mark Goadrich. The relationship between precision-recall and roc curves. In *Proceedings of the 23rd international conference on Machine learning*, pages 233–240, 2006.
- [129] Berkeley J Dietvorst, Joseph P Simmons, and Cade Massey. Algorithm aversion: people erroneously avoid algorithms after seeing them err. *Journal of Experimental Psychology: General*, 144(1):114, 2015.
- [130] Douglas C Engelbart and William K English. A research center for augmenting human intellect. In *Proceedings of the December 9-11, 1968, fall joint computer conference, part I*, pages 395–410, 1968.
- [131] Julia Dressel and Hany Farid. The accuracy, fairness, and limits of predicting recidivism. *Science advances*, 4(1):1–5, 2018.
- [132] Douglas C Engelbart. Augmenting human intellect: A conceptual framework. *Menlo Park, CA*, page 21, 1962. Reprinted in Packer, R. amd Kprdam. L.; eds; (2–1). *Multimedia: From Wagner to Virtual Reality*. New York: W. W. Norton, 64-90.
- [133] Danielle Ensign, Sorelle A Friedler, Scott Neville, Carlos Scheidegger, and Suresh Venkatasubramanian. Runaway feedback loops in predictive policing. In *Proceedings of Machine Learning Research*, pages 1–12, 2018.

- [134] Ziv Epstein, Blakeley H Payne, Judy Hanwen Shen, Abhimanyu Dubey, Bjarke Felbo, Matthew Groh, Nick Obradovich, Manuel Cebrian, and Iyad Rahwan. Closing the ai knowledge gap. *arXiv preprint arXiv:1803.07233*, 2018.
- [135] Michael Feldman, Sorelle A Friedler, John Moeller, Carlos Scheidegger, and Suresh Venkatasubramanian. Certifying and removing disparate impact. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 259–268, 2015.
- [136] Tarleton Gillespie. The Relevance of Algorithms. In *Media technologies: Essays on communication, materiality, and society*, pages 167–194. The MIT Press, 2014.
- [137] Enric Galceran, Alexander G Cunningham, Ryan M Eustice, and Edwin Olson. Multipolicy decision-making for autonomous driving via changepoint-based behavior prediction: Theory and experiment. *Autonomous Robots*, 41:1367–1382, 2017.
- [138] Timnit Gebru, Jamie Morgenstern, Briana Vecchione, Jennifer Wortman Vaughan, Hanna Wallach, Hal Daumé Iii, and Kate Crawford. Datasheets for datasets. *Communications of the ACM*, 64:86–92, 2021.
- [139] Alessandro Giusti, Jérôme Guzzi, Dan C Cireşan, Fang-Lin He, Juan P Rodríguez, Flavio Fontana, Matthias Faessler, Christian Forster, Jürgen Schmidhuber, Gianni Di Caro, et al. A machine learning approach to visual perception of forest trails for mobile robots. *IEEE Robotics and Automation Letters*, 1:661–667, 2015.
- [140] Vern L Glaser. *Enchanted algorithms The Quantification of Organizational Decision-Making*. University of Southern California, 2014.
- [141] Kurt Gray and Daniel M Wegner. Feeling robots and human zombies: Mind perception and the uncanny valley. *Cognition*, 125(1):125–130, 2012.
- [142] Victoria Groom and Clifford Nass. Can robots be teammates?: Benchmarks in human–robot teams. *Interaction studies*, 8(3):483–500, 2007.
- [143] Jaron Harambam, Dimitrios Bountouridis, Mykola Makhortykh, and Joris Van Hoboken. Designing for the better by taking users into account: A qualitative evaluation of user control mechanisms in (news) recommender systems. In *Proceedings of the 13th ACM Conference on Recommender Systems*, pages 69–77, 2019.
- [144] Sophie Freeman, Martin Gibbs, and Bjørn Nansen. ‘don’t mess with my algorithm’: Exploring the relationship between listeners and automated curation and recommendation on music streaming services. *First Monday*, 2022.
- [145] Jeffrey Heer. Agency plus automation: Designing artificial intelligence into interactive systems. *Proceedings of the National Academy of Sciences*, 116:1844–1850, 2019.
- [146] Bram Hendriks, Bernt Meerbeek, Stella Boess, Steffen Pauws, and Marieke Sonneveld. Robot vacuum cleaner personality and behavior. *International Journal of Social Robotics*, 3:187–195, 2011.
- [147] Martin Hilbert, Saifuddin Ahmed, Jaeho Cho, Billy Liu, and Jonathan Luu. Communicating with Algorithms: A Transfer Entropy Analysis of Emotions-based Escapes from Online Echo Chambers. *Communication Methods and Measures*, 12(4):260–275, 2018.
- [148] Gunter J Hitsch, Ali Hortaçsu, and Dan Ariely. Matching and sorting in online dating. *American Economic Review*, 100(1):130–63, 2010.
- [149] Shanee Honig, Alon Bartal, Yisrael Parmet, and Tal Oron-Gilad. Using online customer reviews to classify, predict, and learn about domestic robot failures. *arXiv preprint arXiv:2201.03287*, 2022.
- [150] Shanee Honig and Tal Oron-Gilad. Understanding and resolving failures in human-robot interaction: Literature review and model development. *Frontiers in psychology*, 9:861, 2018.
- [151] Xiyang Hu, Yan Huang, Beibei Li, and Tian Lu. Uncovering the Source of Evaluation Bias in Micro-Lending. In *ICIS 2021 Proceedings*, volume 1. Association for Computing Machinery, 2021.
- [152] Yin-Fu Huang and Yi-Hao Li. Sentiment translation model for expressing positive sentimental statements. In *2019 International Conference on Machine Learning and Data Engineering (iCMLDE)*, pages 79–84. IEEE, 2019.
- [153] Lillian Hung, Mario Gregorio, Jim Mann, Christine Wallsworth, Neil Horne, Annette Berndt, Cindy Liu, Evan Woldum, Andy Au-Yeung, and Habib Chaudhury. Exploring the perceptions of people with dementia about the social robot paro in a hospital setting. *Dementia*, 20:485–504, 2021.
- [154] Nicholas R Jennings, Luc Moreau, David Nicholson, Sarvapali Ramchurn, Stephen Roberts, Tom Rodden, and Alex Rogers. Human-agent collectives. *Communications of the ACM*, 57:80–88, 2014.

- [155] Neil Johnson, Guannan Zhao, Eric Hunsader, Hong Qi, Nicholas Johnson, Jing Meng, and Brian Tivnan. Abrupt rise of new machine ecology beyond human response time. *Scientific reports*, 3(1):1–7, 2013.
- [156] Ece Kamar, Severin Hacker, and Eric Horvitz. Combining human and machine intelligence in large-scale crowdsourcing. In *AAMAS*, volume 12, pages 467–474, 2012.
- [157] Sooyeon Jeong, Cynthia Breazeal, Deirdre Logan, and Peter Weinstock. 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*, pages 1–13, 2018.
- [158] William B Kannel and Daniel L McGee. Diabetes and cardiovascular disease: the framingham study. *Jama*, 241(19):2035–2038, 1979.
- [159] Serge Kernbach, Ronald Thenius, Olga Kernbach, and Thomas Schmickl. Re-embodiment of honeybee aggregation behavior in an artificial micro-robotic system. *Adaptive Behavior*, 17:237–259, 2009.
- [160] Peter H Kahn, Nathan G Freier, Takayuki Kanda, Hiroshi Ishiguro, Jolina H Ruckert, Rachel L Severson, and Shaun K Kane. Design patterns for sociality in human-robot interaction. In *Proceedings of the 3rd ACM/IEEE international conference on Human robot interaction*, pages 97–104, 2008.
- [161] Hiroaki Kitano, Minoru Asada, Yasuo Kuniyoshi, Itsuki Noda, and Eiichi Osawa. Robocup: The robot world cup initiative. In *Proceedings of the first international conference on Autonomous agents*, pages 340–347, 1997.
- [162] Jon Kleinberg, Himabindu Lakkaraju, Jure Leskovec, Jens Ludwig, and Sendhil Mullainathan. Human Decisions and Machine Predictions. *The Quarterly Journal of Economics*, 133(1):237–293, 08 2017.
- [163] Jacqueline M Kory Westlund, Sooyeon Jeong, Hae W Park, Samuel Ronfard, Aradhana Adhikari, Paul L Harris, David DeSteno, and Cynthia L Breazeal. Flat vs. expressive storytelling: Young children’s learning and retention of a social robot’s narrative. *Frontiers in human neuroscience*, 11:295, 2017.
- [164] Johannes Kunkel, Claudia Schwenger, and Jürgen Ziegler. NewsViz: Depicting and Controlling Preference Profiles Using Interactive Treemaps in News Recommender Systems. *UMAP 2020 - Proceedings of the 28th ACM Conference on User Modeling, Adaptation and Personalization*, pages 126–135, 2020.
- [165] Angeliki Lazaridou, Alexander Peysakhovich, and Marco Baroni. Multi-agent cooperation and the emergence of (natural) language. In *5th International Conference on Learning Representations, ICLR 2017 - Conference Track Proceedings*, pages 1–11, 2017.
- [166] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. Microsoft coco: Common objects in context. In *European conference on computer vision*, pages 740–755. Springer, 2014.
- [167] Michael L. Littman, Ifeoma Ajunwa, Guy Berger, Craig Boutilier, Morgan Currie, Finale Doshi-Velez, Gillian Hadfiel, Michael C. Horowitz, Charles Isbell, Hiroaki Kitano, Karen Levy, Terah Lyons, Melanie Mitchell, Julie Shah, Steven Sloman, Shannon Vallor, and Toby Walsh. Gathering Strength, Gathering Storms: The One Hundred Year Study on Artificial Intelligence (AI100) 2021 Study Panel Report. *Stanford University, Stanford, CA*, pages 1–82, 2021.
- [168] Gustavo López, Luis Quesada, and Luis A Guerrero. Alexa vs. siri vs. cortana vs. google assistant: a comparison of speech-based natural user interfaces. In *International conference on applied human factors and ergonomics*, pages 241–250. Springer, 2017.
- [169] Tamara Lorenz, Astrid Weiss, and Sandra Hirche. Synchrony and reciprocity: Key mechanisms for social companion robots in therapy and care. *International Journal of Social Robotics*, 8(1):125–143, 2016.
- [170] Irene Lopatovska, Katrina Rink, Ian Knight, Kieran Raines, Kevin Cosenza, Harriet Williams, Perachya Sorsche, David Hirsch, Qi Li, and Adrianna Martinez. Talk to me: Exploring user interactions with the amazon alexa. *Journal of Librarianship and Information Science*, 51(4):984–997, 2019.
- [171] John Markoff. *Machines of loving grace: The quest for common ground between humans and robots*. HarperCollins Publishers, 2016.
- [172] Sean McGregor. Preventing Repeated Real World AI Failures by Cataloging Incidents: The AI Incident Database. *35th AAAI Conference on Artificial Intelligence, AAAI 2021*, 17B:15458–15463, 2021.
- [173] Michelle N Meyer. Two cheers for corporate experimentation: The a/b illusion and the virtues of data-driven innovation. *Colo. Tech. LJ*, 13:273, 2015.
- [174] Margaret Mitchell, Simone Wu, Andrew Zaldivar, Parker Barnes, Lucy Vasserman, Ben Hutchinson, Elena Spitzer, Inioluwa Deborah Raji, and Timnit Gebru. Model cards for model reporting. In *Proceedings of the conference on fairness, accountability, and transparency*, pages 220–229, 2019.

- [175] Alex Mitrevski, Santosh Thoduka, Argentina Ortega Sáinz, Maximilian Schöbel, Patrick Nagel, Paul G Plöger, and Erwin Prassler. Deploying robots in everyday environments: Towards dependable and practical robotic systems. *arXiv preprint arXiv:2206.12719*, 2022.
- [176] Brent Mittelstadt, Chris Russell, and Sandra Wachter. Explaining explanations in ai. In *Proceedings of the conference on fairness, accountability, and transparency*, pages 279–288, 2019.
- [177] Alexandra S Mueller, Ian J Reagan, and Jessica B Cicchino. Addressing driver disengagement and proper system use: Human factors recommendations for level 2 driving automation design. *Journal of Cognitive Engineering and Decision Making*, 15(1):3–27, 2021.
- [178] Simone Natale. To believe in Siri: A critical analysis of AI voice assistants. Technical Report March, University of Bremen, 2020.
- [179] Tien T Nguyen, Pik-Mai Hui, F Maxwell Harper, Loren Terveen, and Joseph A Konstan. Exploring the filter bubble: the effect of using recommender systems on content diversity. In *Proceedings of the 23rd international conference on World wide web*, pages 677–686, 2014.
- [180] Amit Kumar Pandey and Rodolphe Gelin. A mass-produced sociable humanoid robot: Pepper: The first machine of its kind. *IEEE Robotics & Automation Magazine*, 25(3):40–48, 2018.
- [181] Hae Won Park, Rinat Rosenberg-Kima, Maor Rosenberg, Goren Gordon, and Cynthia Breazeal. Growing growth mindset with a social robot peer. In *Proceedings of the 2017 ACM/IEEE international conference on human-robot interaction*, pages 137–145, 2017.
- [182] Rik Peeters. The agency of algorithms: Understanding human-algorithm interaction in administrative decision-making. *Information Polity*, 25(4):507–522, 2020.
- [183] Ola Pettersson. Execution monitoring in robotics: A survey. *Robotics and Autonomous Systems*, 53(2):73–88, 2005.
- [184] Yuriy Nevmyvaka, Yi Feng, and Michael Kearns. Reinforcement learning for optimized trade execution. In *Proceedings of the 23rd international conference on Machine learning*, pages 673–680, 2006.
- [185] Antonio Pérez, M Isabel García, Manuel Nieto, José L Pedraza, Santiago Rodríguez, and Juan Zamorano. Argos: An advanced in-vehicle data recorder on a massively sensorized vehicle for car driver behavior experimentation. *IEEE Transactions on Intelligent Transportation Systems*, 11(2):463–473, 2010.
- [186] Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. "Why Should I Trust You?" Explaining the Predictions of Any Classifier. *NAACL-HLT 2016 - 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Proceedings of the Demonstrations Session*, pages 97–101, 2016.
- [187] Angel Rivas-Casado, Rafael Martínez-Tomás, and Antonio Fernández-Caballero. Multi-agent system for knowledge-based event recognition and composition. *Expert Systems*, 28(5):488–501, 2011.
- [188] Florian Rosenberg and Schahram Dustdar. Design and implementation of a service-oriented business rules broker. In *Seventh IEEE International Conference on E-Commerce Technology Workshops*, pages 55–63. IEEE, 2005.
- [189] Michael Rubenstein, Alejandro Cornejo, and Radhika Nagpal. Programmable self-assembly in a thousand-robot swarm. *Science*, 345(6198):795–799, 2014.
- [190] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, et al. Imagenet large scale visual recognition challenge. *International Journal of Computer Vision*, 115(3):211–252, 2015.
- [191] Elliot Salisbury, Ece Kamar, and Meredith Ringel Morris. Toward scalable social alt text: Conversational crowdsourcing as a tool for refining vision-to-language technology for the blind. In *Fifth AAAI Conference on Human Computation and Crowdsourcing*, 2017.
- [192] Matthew J Salganik, Peter Sheridan Dodds, and Duncan J Watts. Experimental study of inequality and unpredictability in an artificial cultural market. *science*, 311(5762):854–856, 2006.
- [193] Filippo Santoni de Sio and Jeroen Van den Hoven. Meaningful human control over autonomous systems: A philosophical account. *Frontiers in Robotics and AI*, page 15, 2018.
- [194] Ben Shneiderman. The limits of speech recognition. *Communications of the ACM*, 43(9):63–65, 2000.
- [195] Mehdi Shanbedi, Saeed Zeinali Heris, Ahmad Amiri, Sadegh Adyani, Mohsen Alizadeh, and Majid Baniadam. Optimization of the thermal efficiency of a two-phase closed thermosyphon using active learning on the human algorithm interaction. *Numerical Heat Transfer; Part A: Applications*, 66(8):947–962, 2014.

- [196] Ben Sheehan, Hyun Seung Jin, and Udo Gottlieb. Customer service chatbots: Anthropomorphism and adoption. *Journal of Business Research*, 115:14–24, 2020.
- [197] Donghee Shin. How do users interact with algorithm recommender systems? The interaction of users, algorithms, and performance. *Computers in Human Behavior*, 109(May 2019):106344, 2020.
- [198] Geoffrey E Hinton, Simon Osindero, and Yee-Whye Teh. A fast learning algorithm for deep belief nets. *Neural computation*, 18(7):1527–1554, 2006.
- [199] Hirokazu Shirado and Nicholas A Christakis. Locally noisy autonomous agents improve global human coordination in network experiments. *Nature*, 545(7654):370–374, 2017.
- [200] David Silver, Julian Schrittwieser, Karen Simonyan, Ioannis Antonoglou, Aja Huang, Arthur Guez, Thomas Hubert, Lucas Baker, Matthew Lai, Adrian Bolton, et al. Mastering the game of go without human knowledge. *Nature*, 550(7676):354–359, 2017.
- [201] David Silver, Aja Huang, Chris J Maddison, Arthur Guez, Laurent Sifre, George Van Den Driessche, Julian Schrittwieser, Ioannis Antonoglou, Veda Panneershelvam, Marc Lanctot, et al. Mastering the game of go with deep neural networks and tree search. *nature*, 529(7587):484–489, 2016.
- [202] Wilko Schwarting, Javier Alonso-Mora, Liam Pauli, Sertac Karaman, and Daniela Rus. Parallel autonomy in automated vehicles: Safe motion generation with minimal intervention. In *2017 IEEE International Conference on Robotics and Automation (ICRA)*, pages 1928–1935. IEEE, 2017.
- [203] Michel Taïx, David Flavigné, and Etienne Ferré. Human interaction with motion planning algorithm. *Journal of Intelligent and Robotic Systems: Theory and Applications*, 67(3-4):285–306, 2012.
- [204] Andreas Theodorou, Robert H Wortham, and Joanna J Bryson. Designing and implementing transparency for real time inspection of autonomous robots. *Connection Science*, 29(3):230–241, 2017.
- [205] Andrea L Thomaz and Cynthia Breazeal. Teachable robots: Understanding human teaching behavior to build more effective robot learners. *Artificial Intelligence*, 172(6-7):716–737, 2008.
- [206] Dayong Wang, Aditya Khosla, Rishab Gargeya, Humayun Irshad, and Andrew H Beck. Deep learning for identifying metastatic breast cancer. *arXiv preprint arXiv:1606.05718*, 2016.
- [207] David Watson. The rhetoric and reality of anthropomorphism in artificial intelligence. *Minds and Machines*, 29(3):417–440, 2019.
- [208] Andreas Wagner. *Robustness and evolvability in living systems*. Princeton university press, 2013.
- [209] Jacqueline M Kory Westlund, Hae Won Park, Randi Williams, and Cynthia Breazeal. Measuring young children’s long-term relationships with social robots. In *Proceedings of the 17th ACM conference on interaction design and children*, pages 207–218, 2018.
- [210] David H Wolpert. The lack of a priori distinctions between learning algorithms. *Neural computation*, 8(7):1341–1390, 1996.
- [211] Simon Haykin. *Neural networks: a comprehensive foundation*. Prentice Hall PTR, 1998.
- [212] Alon Halevy, Peter Norvig, and Fernando Pereira. The unreasonable effectiveness of data. *IEEE Intelligent Systems*, 24(2):8–12, 2009.
- [213] Martin Erwig. *Once upon an algorithm: How stories explain computing*. MIT Press, 2017.
- [214] Sangseok You and Lionel Robert. Emotional attachment, performance, and viability in teams collaborating with embodied physical action (epa) robots. *You, S. and Robert, LP (2018). Emotional Attachment, Performance, and Viability in Teams Collaborating with Embodied Physical Action (EPA) Robots, Journal of the Association for Information Systems*, 19(5):377–407, 2017.
- [215] Rich Zemel, Yu Wu, Kevin Swersky, Toni Pitassi, and Cynthia Dwork. Learning fair representations. In *International conference on machine learning*, pages 325–333. PMLR, 2013.
- [216] Xiaohua Zeng, Abraham O Fapojuwo, and Robert J Davies. Design and performance evaluation of voice activated wireless home devices. *IEEE Transactions on Consumer Electronics*, 52(3):983–989, 2006.
- [217] Jie M Zhang, Mark Harman, Lei Ma, and Yang Liu. Machine learning testing: Survey, landscapes and horizons. *IEEE Transactions on Software Engineering*, 2020.
- [218] Eric Horvitz. Principles of mixed-initiative user interfaces. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, pages 159–166, 1999.
- [219] Vinod Nair and Geoffrey E Hinton. Rectified linear units improve restricted boltzmann machines. In *Proceedings of the 27th international conference on machine learning (ICML-10)*, pages 807–814, 2010.

- [220] Richard O Duda, Peter E Hart, and David G Stork. *Pattern classification*. John Wiley & Sons, 2012.
- [221] Sébastien Bubeck et al. Convex optimization: Algorithms and complexity. *Foundations and Trends® in Machine Learning*, 8(3-4):231–357, 2015.
- [222] Aurélien Géron. *Hands-On Machine Learning with Scikit-Learning with Keras & TensorFlow*. O'Reilly Media, Inc., Canada, 2 edition, 2019.
- [223] Fatima N Al-Aswadi, Huah Yong Chan, and Keng Hoon Gan. Automatic ontology construction from text: a review from shallow to deep learning trend. *Artificial Intelligence Review*, 53:3901–3928, 2020.
- [224] John Wenskovitch, Michelle Zhou, Christopher Collins, Remco Chang, Michelle Dowling, Alex Endert, and Kai Xu. Putting the “i” in interaction: Interactive interfaces personalized to individuals. *IEEE Computer Graphics and Applications*, 40(3):73–82, 2020.
- [225] Percy Liang, Rishi Bommasani, Tony Lee, Dimitris Tsipras, Dilara Soylu, Michihiro Yasunaga, Yian Zhang, Deepak Narayanan, Yuhuai Wu, Ananya Kumar, et al. Holistic evaluation of language models. *arXiv preprint arXiv:2211.09110*, 2022.
- [226] Lalit Mohan Sanagavarapu, Vivek Iyer, and Raghu Reddy. A deep learning approach for ontology enrichment from unstructured text. *arXiv preprint arXiv:2112.08554*, 2021.
- [227] Waqas Haider Bangyal, Najeeb Ur Rehman, Asma Nawaz, Kashif Nisar, Ag Asri Ag Ibrahim, Rabia Shakir, and Danda B Rawat. Constructing domain ontology for alzheimer disease using deep learning based approach. *Electronics*, 11(12):1890, 2022.
- [228] Nithin Buduma, Nikhil Buduma, and Joe Papa. *Fundamentals of deep learning*. " O'Reilly Media, Inc.", 2022.
- [229] Jens Albrecht, Sidharth Ramachandran, and Christian Winkler. *Blueprints for Text Analytics Using Python*. O'Reilly, 2021.
- [230] Stefano Angieri, Alberto García-Martínez, Bingyang Liu, Zhiwei Yan, Chuang Wang, and Marcelo Bagnulo. An experiment in distributed internet address management using blockchains. *arXiv preprint arXiv:1807.10528*, 2018.
- [231] Juliao Braga, Joao Nuno Silva, Patricia Takako Endo, Jessica Ribas, and Nizam Omar. Blockchain to improve security, knowledge and collaboration inter-agent communication over restrict domains of the internet infrastructure, with human interaction. *Brazilian Journal of Development*, 5(7):9013–9029, july 2019. DOI:10.34117/bjdv5n7-103, ISSN 2525-8761.
- [232] Igor M Coelho and Vitor N Coelho. Neocompiler eco: experimentação de consenso em blockchain e contratos inteligentes. In *Anais do VI Workshop do testbed FIBRE*, pages 57–67. SBC, 2021.
- [233] Lara Bonemer Rocha Floriani. *Smart contracts nos contratos empresariais: um estudo sobre possibilidade e viabilidade econômica de sua utilização*. Editora Dialética, 2021.
- [234] Alan E. Kazdin. *The token economy: A Review and Evaluation*. Plenum Press, 2012.
- [235] Antony Lewis. *The Basics of Bitcoins and Blockchains: An Introduction to Cryptocurrencies and the Technology that Powers Them*. Group, Mango Publishing, Coral Gables, FL, 1 edition, 2018.
- [236] Alex Murray, Dennie Kim, and Jordan Combs. The promise of a decentralized internet: What is web 3.0 and how can firms prepare? *Business Horizons*, 65:565–570, 2022.
- [237] Ankita Saxena. Workforce Diversity: A Key to Improve Productivity. *Procedia Economics and Finance*, 11:76–85, 2014.
- [238] Steven D. Travers. Distributed Autonomous Organization: A Blockchain Organizational Archetype. *Strategic Management Society 37th Annual Conference*, 2017.
- [239] Shermin Voshmgir. *Economia dos Tokens: Como a Web3 está reinventando a internet e a relação entre os agentes econômicos*. Token Kitchen, 2 edition, 2020.
- [240] Aries Wanlin Wang. *Crypto Economy: How Blockchain, Cryptocurrency and Token-Economy are Disrupting the Financial World*. Skyhorse Publishing, 2018.
- [241] Guy R Vishnia and Gareth W Peters. Auditchain: A trading audit platform over blockchain. *Frontiers in Blockchain*, 3:9, 2020.
- [242] Denys Virovets and Sergiy Obushnyi. Decentralized autonomous organizations as the new form of economic cooperation in digital world. *The USV Annals of Economics and Public Administration*, 20(2 (32)):41–52, 2021.
- [243] Wikipédia. Decentralized autonomous organization, 2021. [Online; accessed 12-junho-2023].

- [244] Yan Chen and Cristiano Bellavitis. Blockchain disruption and decentralized finance: The rise of decentralized business models. *Journal of Business Venturing Insights*, 13:e00151, 2020.
- [245] Cristiano Bellavitis, Christian Fisch, and Paul P Momtaz. The rise of decentralized autonomous organizations (daos): a first empirical glimpse. *Venture Capital*, pages 1–17, 2022.
- [246] Gail Weinstein, Steven Lofchie, Jason Schwartz, Frank Fried, Shriver Harris, and Jacobson LLP. A Primer on DAOs, 2022.
- [247] Brynly Llyr, Aiden Slavin, and Kevin Werbach. Decentralized Autonomous Organization Toolkit. Technical report, World Economic Forum, 2023.
- [248] Fredrik Ygge and Hans Akkermans. Decentralized markets versus central control: A comparative study. *Journal of artificial intelligence research*, 11:301–333, 1999.
- [249] W Brian Arthur. Complexity and the economy. In *Handbook of Research on Complexity*. Edward Elgar Publishing, 2009.
- [250] Nicolas Bouleau. On excessive mathematization, symptoms, diagnosis and philosophical bases for real world knowledge. *Real World Economics*, 57:90–105, 2011.
- [251] Kyle Croman, Christian Decker B, Ittay Eyal, Adem Efe Gencer, Ari Juels, Ahmed Kosba, Andrew Miller, Prateek Saxena, Dawn Song, and Roger Wattenhofer. On Scaling Decentralized Blockchains (A Position Paper). *Lecture Notes in Computer Science*, 9604:106–125, 2016.
- [252] Eric Budish, Peter Cramton, and John Shim. The high-frequency trading arms race: Frequent batch auctions as a market design response. *The Quarterly Journal of Economics*, 130(4):1547–1621, 2015.
- [253] John Cartledge, Marco De Luca, Charlotte Szostek, and Dave Cliff. Too fast too furious: faster financial-market trading agents can give less efficient markets. In *ICAART-2012: 4th International Conference on Agents and Artificial Intelligence*, pages 126–135. SciTePress, 2012.
- [254] Le Chen and Christo Wilson. Observing algorithmic marketplaces in-the-wild. *ACM SIGecom Exchanges*, 15(2):34–39, 2017.
- [255] Le Chen, Alan Mislove, and Christo Wilson. An empirical analysis of algorithmic pricing on amazon marketplace. In *Proceedings of the 25th international conference on World Wide Web*, pages 1339–1349, 2016.
- [256] J Doyne Farmer and Spyros Skouras. An ecological perspective on the future of computer trading. *Quantitative Finance*, 13:325–346, 2013.
- [257] Emilio Ferrara, Onur Varol, Clayton Davis, Filippo Menczer, and Alessandro Flammini. The rise of social bots. *Communications of the ACM*, 59:96–104, 2016.
- [258] Michael Kearns, Alex Kulesza, and Yuriy Nevmyvaka. Empirical limitations on high-frequency trading profitability. *The Journal of Trading*, 5:50–62, 2010.
- [259] Andrei A Kirilenko and Andrew W Lo. Moore’s law versus murphy’s law: Algorithmic trading and its discontents. *Journal of Economic Perspectives*, 27(2):51–72, 2013.
- [260] Jon Kleinberg, Sendhil Mullainathan, and Manish Raghavan. Inherent trade-offs in the fair determination of risk scores. *Leibniz International Proceedings in Informatics, LIPIcs*, 67:1–23, 2017.
- [261] Jon Kleinberg and Sigal Oren. Mechanisms for (mis) allocating scientific credit. In *Proceedings of the forty-third annual ACM symposium on Theory of computing*, pages 529–538, 2011.
- [262] Farshad Kooti, Mihajlo Grbovic, Luca Maria Aiello, Nemanja Djuric, Vladan Radosavljevic, and Kristina Lerman. Analyzing uber’s ride-sharing economy. In *Proceedings of the 26th International Conference on World Wide Web Companion*, pages 574–582, 2017.
- [263] Jaron Lanier. *You are not a gadget*. Vintage, 2010.
- [264] Michael Latzer, Katharina Hollnbuchner, Natascha Just, and Florian Saurwein. The economics of algorithmic selection on the internet. In *Handbook on the Economics of the Internet*, pages 395–425. University of Zurich, 2014.
- [265] Albert J Menkveld. The economics of high-frequency trading. *Annual Review of Financial Economics*, 8:1–24, 2016.
- [266] Miriam Naigembe. *Bank lending policy, credit scoring and the survival of Loans: A case study of banks X and Y*. PhD thesis, Makerere University, 2010.
- [267] Tommy Lundgren. A microeconomic model of corporate social responsibility. *Metroeconomica*, 62(1):69–95, 2011.

- [268] David C Parkes and Michael P Wellman. Economic reasoning and artificial intelligence. *Science*, 349(6245):267–272, 2015.
- [269] Frank Pasquale. *The Black Box Society: The Secret Algorithms that Control Money and Information*. Harvard University Press, 2015.
- [270] Kasper Roszbach. Bank lending policy, credit scoring, and the survival of loans. *Review of Economics and Statistics*, 86(4):946–958, 2004.
- [271] Jonathan JJM Seddon and Wendy L Currie. A model for unpacking big data analytics in high-frequency trading. *Journal of Business Research*, 70:300–307, 2017.
- [272] Robert J Shiller. *Narrative economics: How stories go viral and drive major economic events*. Princeton University Press, 2020.
- [273] Chih-Fong Tsai and Jhen-Wei Wu. Using neural network ensembles for bankruptcy prediction and credit scoring. *Expert systems with applications*, 34(4):2639–2649, 2008.
- [274] Michael P. Wellman, Peter R. Wurman, Kevin O’Malley, Roshan Bangera, Daniel Reeves, William E Walsh, et al. Designing the market game for a trading agent competition. *IEEE Internet Computing*, 5(2):43–51, 2001.
- [275] Bernard Dafflon. The assignment of functions to decentralized government: from theory to practice. In *Handbook of multilevel finance*, pages 163–199. Edward Elgar Publishing, 2015.
- [276] Shuai Wang, Wenwen Ding, Juanjuan Li, Yong Yuan, Liwei Ouyang, and Fei-Yue Wang. Decentralized autonomous organizations: Concept, model, and applications. *IEEE Transactions on Computational Social Systems*, 6(5):870–878, 2019.
- [277] Alexandre Aronne, Aureliano Bressan, and Haroldo Guimarães Brasil. *Mensuração e Gerenciamento de Riscos Corporativos: Aplicações de Cash Flow at Risk e Real Options*. Saint Paul, 2021.
- [278] Rogério Silva Nacif. *Operações Eficientes, Empresas Rentáveis: Melhorando os Resultados Financeiros por Meio da Gestão de Operações*. Aquila, 2021.
- [279] Wenwen Ding, Juanjuan Li, Rui Qin, Robert Kozma, and Fei-Yue Wang. A new architecture and mechanism for decentralized science metamarkets. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 2023.
- [280] Carlo Appugliese, Paco Nathan, and William S Roberts. *Agile AI: A Practical Guide to Building AI Applications and Teams*. O’Reilly, 2020.
- [281] Kenneth Appel, Wolfgang Haken, and John Koch. Every planar map is four colorable. part ii: Reducibility. *Illinois Journal of Mathematics*, 21(3):491–567, 1977.
- [282] Kenneth Appel and Wolfgang Haken. Every planar map is four colorable. *Bulletin of the American mathematical Society*, 82(5):711–712, 1976.
- [283] Per Bak, Kan Chen, and Michael Creutz. Self-organized criticality in the game of life. *Nature*, 342(6251):780–782, 1989.
- [284] Marc G Bellemare, Yavar Naddaf, Joel Veness, and Michael Bowling. The arcade learning environment: An evaluation platform for general agents. *Journal of Artificial Intelligence Research*, 47:253–279, 2013.
- [285] Roger Bemelmans, Gert Jan Gelderblom, Pieter Jonker, and Luc De Witte. Socially assistive robots in elderly care: a systematic review into effects and effectiveness. *Journal of the American Medical Directors Association*, 13(2):114–120, 2012.
- [286] Andrew Berdahl, Colin J Torney, Christos C Ioannou, Jolyon J Faria, and Iain D Couzin. Emergent sensing of complex environments by mobile animal groups. *Science*, 339(6119):574–576, 2013.
- [287] Ana Berdasco, Gustavo López, Ignacio Diaz, Luis Quesada, and Luis A Guerrero. User experience comparison of intelligent personal assistants: Alexa, google assistant, siri and cortana. *Multidisciplinary Digital Publishing Institute Proceedings*, 31(1):51, 2019.
- [288] Alessandro Bessi and Emilio Ferrara. Social bots distort the 2016 us presidential election online discussion. *First monday*, 21(11-7), 2016.
- [289] Michael Bowling, Neil Burch, Michael Johanson, and Oskari Tammelin. Heads-up limit hold’em poker is solved. *Communications of the ACM*, 60(11):81–88, 2017.
- [290] Murray Campbell, A Joseph Hoane Jr, and Feng-hsiung Hsu. Deep blue. *Artificial intelligence*, 134(1-2):57–83, 2002.
- [291] Juan Miguel Carrascosa, Jakub Mikians, Ruben Cuevas, Vijay Erramilli, and Nikolaos Laoutaris. I always feel like somebody’s watching me: measuring online behavioural advertising. In *Proceedings of the 11th ACM Conference on Emerging Networking Experiments and Technologies*, pages 1–13, 2015.

- [292] Maurice Chiodo and Toby Clifton. The importance of ethics in mathematics. *European Mathematical Society Magazine*, 114:34–37, 2019.
- [293] BACEN. LIFT Challenge, 2022. Accessed in 03/09/2022.
- [294] Wasifa Chowdhury. *Employing neural hierarchical model with pointer generator networks for abstractive text summarization*. PhD thesis, Simon Frazer University: School of Computing Science, 2019.
- [295] Battista Biggio, Igino Corona, Davide Maiorca, Blaine Nelson, Nedim Šrđić, Pavel Laskov, Giorgio Giacinto, and Fabio Roli. Evasion attacks against machine learning at test time. In *Joint European conference on machine learning and knowledge discovery in databases*, pages 387–402. Springer, 2013.
- [296] Jack Clark and Ray Perrault. Introduction to the AI index report 2022. Technical report, Stanford University, 2022.
- [297] Amit Datta, Michael Carl Tschantz, and Anupam Datta. Automated experiments on ad privacy settings: A tale of opacity, choice, and discrimination. *arXiv preprint arXiv:1408.6491*, 2014.
- [298] Yue Deng, Feng Bao, Youyong Kong, Zhiqian Ren, and Qionghai Dai. Deep direct reinforcement learning for financial signal representation and trading. *IEEE Transactions on Neural Networks and Learning Systems*, 28(3):653–664, 2016.
- [299] Pedro Domingos. *The master algorithm: How the quest for the ultimate learning machine will remake our world*. Basic Books, 2015.
- [300] Finale Doshi-Velez and Been Kim. Towards a rigorous science of interpretable machine learning. *arXiv preprint arXiv:1702.08608*, 2017.
- [301] Sebastian Elbaum and John C Munson. Software black box: an alternative mechanism for failure analysis. In *Proceedings 11th International Symposium on Software Reliability Engineering. ISSRE 2000*, pages 365–376. IEEE, 2000.
- [302] Giuliano Da Empoli. *Os engenheiros do caos*. Vestígio, Belo Horizonte, 1 edition, 2019.
- [303] Amir Globerson and Sam Roweis. Nightmare at test time: robust learning by feature deletion. In *Proceedings of the 23rd international conference on Machine learning*, pages 353–360, 2006.
- [304] Dennis R Grossi. Aviation Recorder Overview. In *International Symposium On Transportation Recorders*, page 12, 2006.
- [305] Jose Hernandez-Orallo. Beyond the turing test. *Journal of Logic, Language and Information*, 9(4):447–466, 2000.
- [306] John Kay and Mervyn King. *Radical uncertainty: Decision-making beyond the numbers*. WW Norton & Company, 2020.
- [307] Himabindu Lakkaraju, Ece Kamar, Rich Caruana, and Eric Horvitz. Identifying unknown unknowns in the open world: Representations and policies for guided exploration. In *Thirty-first aaai conference on artificial intelligence*, 2017.
- [308] Tian-Shyug Lee and I-Fei Chen. A two-stage hybrid credit scoring model using artificial neural networks and multivariate adaptive regression splines. *Expert Systems with applications*, 28(4):743–752, 2005.
- [309] Joel Z Leibo, Cyprien de Masson d’Autume, Daniel Zoran, David Amos, Charles Beattie, Keith Anderson, Antonio García Castañeda, Manuel Sanchez, Simon Green, Audrunas Gruslys, et al. Psychlab: a psychology laboratory for deep reinforcement learning agents. *arXiv preprint arXiv:1801.08116*, 2018.
- [310] Nancy G. Leveson. *Engineering a Safer World: Systems Thinking Applied to Safety*. The MIT Press, 2011.
- [311] C Dianne Martin. The myth of the awesome thinking machine. *Communications of the ACM*, 36(4):120–133, 1993.
- [312] Simone Natale et al. *Deceitful media: Artificial intelligence and social life after the Turing test*. Oxford University Press, USA, 2021.
- [313] Olfa Nasraoui and Patrick Shafto. Human-Algorithm Interaction Biases in the Big Data Cycle: A Markov Chain Iterated Learning Framework. *arXiv*, 2016.
- [314] Randolph M Nesse. Tinbergen’s four questions, organized: a response to bateson and laland. *Trends in Ecology & Evolution*, 28(12):681–82, 2013.
- [315] Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. Bleu: a method for automatic evaluation of machine translation. In *Proceedings of the 40th annual meeting of the Association for Computational Linguistics*, pages 311–318, 2002.

- [316] Judea Pearl and Dana Mackenzie. *The Book of Why: The New Science of Cause and Effect*. Basic Books, New York, first edition, 2018.
- [317] David L. Poole and Alan K. Mackworth. *Artificial Intelligence: foundations of computational agents*. Cambridge University Press, second edition, 2017.
- [318] David L Poole and Alan K Mackworth. *Artificial Intelligence: foundations of computational agents*. Cambridge University Press, 2010.
- [319] Stuart Russel and Peter Norvig. *Artificial Intelligence*. Prentice Hall, New York, 3 edition, 2010.
- [320] Jonathan Schaeffer, Neil Burch, Yngvi Bjornsson, Akihiro Kishimoto, Martin Muller, Robert Lake, Paul Lu, and Steve Sutphen. Checkers is solved. *science*, 317(5844):1518–1522, 2007.
- [321] Greg Siegel. *Forensic media: Reconstructing accidents in accelerated modernity*. Duke University Press, 2014.
- [322] Robert Skidelsky. *Information retrieval and hypertext*. Yale University Press, 2020.
- [323] Daniel Smilkov, Nikhil Thorat, Been Kim, Fernanda Viégas, and Martin Wattenberg. Smoothgrad: removing noise by adding noise. *arXiv preprint arXiv:1706.03825*, 2017.
- [324] Megha Srivastava, Hoda Heidari, and Andreas Krause. Mathematical notions vs. human perception of fairness: A descriptive approach to fairness for machine learning. In *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*, pages 2459–2468, 2019.
- [325] Venkatramanan S Subrahmanian, Amos Azaria, Skylar Durst, Vadim Kagan, Aram Galstyan, Kristina Lerman, Linhong Zhu, Emilio Ferrara, Alessandro Flammini, and Filippo Menczer. The darpa twitter bot challenge. *Computer*, 49(6):38–46, 2016.
- [326] Christian Szegedy, Wojciech Zaremba, Ilya Sutskever, Joan Bruna, Dumitru Erhan, Ian Goodfellow, and Rob Fergus. Intriguing properties of neural networks. *2nd International Conference on Learning Representations, ICLR 2014 - Conference Track Proceedings*, pages 1–10, 2014.
- [327] Florian Tramèr, Alexey Kurakin, Nicolas Papernot, Ian Goodfellow, Dan Boneh, and Patrick McDaniel. Ensemble adversarial training: Attacks and defenses. In *6th International Conference on Learning Representations, ICLR 2018 - Conference Track Proceedings*, pages 1–22, 2018.
- [328] Milena Tsvetkova, Taha Yasseri, Eric T. Meyer, J. Brian Pickering, Vegard Engen, Paul Walland, Marika Lüders, Asbjørn Følstad, and George Bravos. Understanding Human-Machine Networks. *ACM Computing Surveys*, 50(1):1–35, 2018.
- [329] Milena Tsvetkova, Ruth García-Gavilanes, Luciano Floridi, and Taha Yasseri. Even good bots fight: The case of wikipedia. *PloS one*, 12(2):e0171774, 2017.
- [330] Alan M Turing. Computing machinery and intelligence. *Mind*, LIX(236):433–460, 1950.
- [331] Koen Van De Sande, Theo Gevers, and Cees Snoek. Evaluating color descriptors for object and scene recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(9):1582–1596, 2009.
- [332] Carissa Véliz. *Privacy is power*. Melville House, 2021.
- [333] Xingyu Xing, Wei Meng, Dan Doozan, Alex C Snoeren, Nick Feamster, and Wenke Lee. Take this personally: Pollution attacks on personalized services. In *22nd USENIX Security Symposium (USENIX Security 13)*, pages 671–686, 2013.
- [334] Yu Yao and Ella Atkins. The smart black box: A value-driven high-bandwidth automotive event data recorder. *IEEE Transactions on Intelligent Transportation Systems*, 22(3):1484–1496, 2020.
- [335] Quan-shi Zhang and Song-Chun Zhu. Visual interpretability for deep learning: a survey. *Frontiers of Information Technology & Electronic Engineering*, 19(1):27–39, 2018.
- [336] Zhiming Zhou, Han Cai, Shu Rong, Yuxuan Song, Kan Ren, Weinan Zhang, Yong Yu, and Jun Wang. Activation maximization generative adversarial nets. *arXiv preprint arXiv:1703.02000*, 2017.
- [337] Susmit Jha, Tuhin Sahai, Vasumathi Raman, Alessandro Pinto, and Michael Francis. Explaining ai decisions using efficient methods for learning sparse boolean formulae. *Journal of Automated Reasoning*, 63(4):1055–1075, 2019.
- [338] Douwe Kiela, Max Bartolo, Yixin Nie, Divyansh Kaushik, Atticus Geiger, Zhengxuan Wu, Bertie Vidgen, Grusha Prasad, Amanpreet Singh, Pratik Ringshia, et al. Dynabench: Rethinking benchmarking in nlp. *arXiv preprint arXiv:2104.14337*, 2021.
- [339] Peter Marc Deisenroth, A. Aldo Faisal, and Cheng Soom Ong. *MATHEMATICS FOR MACHINE LEARNING*. Cambridge University Press, 1 edition, 2020.

- [340] Patrícia Gomes Rêgo de Almeida. Regulação da Inteligência Artificial: Ação Coletiva que Requer Governança. *Computação Brasil*, 7:23–26, 2022.
- [341] Margaret Boden, Joanna Bryson, Darwin Caldwell, Kerstin Dautenhahn, Lilian Edwards, Sarah Kember, Paul Newman, Vivienne Parry, Geoff Pegman, Tom Rodden, Tom Sorrell, Mick Wallis, Blay Whitby, and Alan Winfield. Principles of robotics: regulating robots in the real world. *Connection Science*, 29(2):124–129, 2017.
- [342] Kevin Bonsor and Nathan Chandler. How black boxes work. *HowStuffWorks*, June, 13, 2001.
- [343] Fernanda Braganca and Renata Braga. Os Desafios da Regulamentação Jurídica da Inteligência Artificial no Brasil. *Computação Brasil*, 7:19–22, 2022.
- [344] Lindell Bromham, Russell Dinnage, and Xia Hua. Interdisciplinary research has consistently lower funding success. *Nature*, 534(7609):684–687, 2016.
- [345] Joanna J Bryson, Mihailis E Diamantis, and Thomas D Grant. Of, for, and by the people: the legal lacuna of synthetic persons. *Artificial Intelligence and Law*, 25(3):273–291, 2017.
- [346] Roger Clarke. Regulatory alternatives for ai. *Computer Law & Security Review*, 35(4):398–409, 2019.
- [347] Europe Commission. Coordinated Plan on Artificial Intelligence 2021 Review. Technical report, European Commission, Brussels, 2021.
- [348] Sam Corbett-Davies, Emma Pierson, Avi Feller, Sharad Goel, and Aziz Huq. Algorithmic decision making and the cost of fairness. In *Proceedings of the 23rd acm sigkdd international conference on knowledge discovery and data mining*, pages 797–806, 2017.
- [349] Iain D Couzin, Christos C Ioannou, Güven Demirel, Thilo Gross, Colin J Torney, Andrew Hartnett, Larissa Conradt, Simon A Levin, and Naomi E Leonard. Uninformed individuals promote democratic consensus in animal groups. *science*, 334(6062):1578–1580, 2011.
- [350] Kate Crawford, Meredith Whittaker, Madeleine Clare Elish, Solon Barocas, Aaron Plasek, and Kadija Ferryman. The ai now report. *The Social and Economic Implications of Artificial Intelligence Technologies in the Near-Term*, 2016.
- [351] Kenneth Cukier, Viktor Mayer-Schönberger, and Francis de Véricourt. *Framers: Human advantage in an age of technology and turmoil*. Penguin, 2022.
- [352] Harrison Edwards and Amos Storkey. Censoring representations with an adversary. *4th International Conference on Learning Representations, ICLR 2016 - Conference Track Proceedings*, pages 1–14, 2016.
- [353] EPI. Algorithmic transparency: End secret profiling. Technical report, Electronic Privacy Information Center, 2015.
- [354] Virginia Eubanks. *Automating inequality: How high-tech tools profile, police, and punish the poor*. St. Martin’s Press, 2018.
- [355] Gregory Falco, Ben Shneiderman, Julia Badger, Ryan Carrier, Anton Dahbura, David Danks, Martin Eling, Alwyn Goodloe, Jerry Gupta, Christopher Hart, et al. Governing ai safety through independent audits. *Nature Machine Intelligence*, 3(7):566–571, 2021.
- [356] Hanming Fang and Andrea Moro. Theories of statistical discrimination and affirmative action: A survey. In *Handbook of Social Economics*, volume 1, pages 133–200. Elsevier B.V., 2011.
- [357] Jessica Fjeld, Nele Achten, Hannah Hilligoss, Adam Nagy, and Madhulika Srikumar. Principled artificial intelligence: Mapping consensus in ethical and rights-based approaches to principles for ai. *Berkman Klein Center Research Publication*, 2020-1, 2020.
- [358] Ana Frazão. Discriminação Algoritmica: a relação entre homens e máquinas. Coluna Jota, junho 2021. Trabalho dividido em treze partes.
- [359] Future of Live Institute. Autonomous Weapons: An Open Letter from AI and Robotics Reserarchers. <https://futureoflife.org/2016/02/09/open-letter-autonomous-weapons-ai-robotics/?cn-reloaded=1&cn-reloaded=1>, July 2018.
- [360] IEEE. Ethically Aligned Design: Version 2 - For Public Discussion. *IEEE Standards*, pages 1–263, 2017.
- [361] CEI. Our vision. Technical report, Council on ExtendedIntelligence, 2022.
- [362] Lucas D. Introna and Helen Nissenbaum. Shaping the web: Why the politics of search engines matters. *Computer Ethics*, pages 157–173, 2017.
- [363] Joosr. *A Joosr Guide to... Weapons of Math Destruction by Cathy O’Neil: How Big Data Increases Inequality and Threatens Democracy*. Broadway books, 2016.

- [364] Daniel Kahneman, Olivier Sibony, and CR Sunstein. *Noise*. HarperCollins UK, 2022.
- [365] Daniel Kahneman, AM Rosenfield, L Gandhi, and T Blaser. Noise: How to overcome the high. *Havard Business Review*, 2016. Available in <https://hbr.org/2016/10/noise>.
- [366] Daniel Kahneman. *Thinking, fast and slow*. Macmillan, 2011.
- [367] Pratyusha Kalluri et al. Don’t ask if artificial intelligence is good or fair, ask how it shifts power. *Nature*, 583:169–169, 2020.
- [368] Krishna M. Kavi. Beyond the Black Box. *IEE Spectrum*, 47(8):46—51, 2021.
- [369] Pranav Khadpe, Ranjay Krishna, Li Fei-Fei, Jeffrey T Hancock, and Michael S Bernstein. Conceptual metaphors impact perceptions of human-ai collaboration. *Proceedings of the ACM on Human-Computer Interaction*, 4(CSCW2):1–26, 2020.
- [370] Mervyn King and John Kay. *Radical Uncertainty: Decision-making for an unknowable future*. Hachette UK, 2020.
- [371] Nicole C Krämer, Astrid von der Pütten, and Sabrina Eimler. Human-agent and human-robot interaction theory: Similarities to and differences from human-human interaction. In *Human-computer interaction: The agency perspective*, pages 215–240. Springer, 2012.
- [372] Samantha Krening and Karen M Feigh. Interaction Algorithm Effect on Human Experience. *ACMTrans. Human-Robot Interact*, 7(2):22, 2018.
- [373] Armin Krishnan. *Killer robots: legality and ethicality of autonomous weapons*. Routledge, 2016.
- [374] David Lazer, Ryan Kennedy, Gary King, and Alessandro Vespignani. The parable of google flu: traps in big data analysis. *Science*, 343(6176):1203–1205, 2014.
- [375] Heidi Ledford. Team science. *Nature*, 525(7569):308–311, 2015.
- [376] Joseph CR Licklider. Man-computer symbiosis. *IRE transactions on human factors in electronics*, pages 4–11, 1960.
- [377] Claudia Bauzer Medeiros. Dados, Algoritmos, Máquinas e Pessoas. *Computação Brasil*, 7:11–14, 2022.
- [378] David A Mindell. *Our robots, ourselves: Robotics and the myths of autonomy*. Viking, 2015.
- [379] Sendhil Mullainathan. Biased algorithms are easier to fix than biased people. *The New York Times*, 2019.
- [380] NSCAI. National Security Commission on Artificial Intelligence - Interim Report. *National Security Commission on Artificial Intelligence Report*, pages 1–101, 2019.
- [381] Cathy O’neil. *Weapons of math destruction: How big data increases inequality and threatens democracy*. Broadway books, 2016.
- [382] Ziad Obermeyer, Brian Powers, Christine Vogeli, and Sendhil Mullainathan. Dissecting racial bias in an algorithm used to manage the health of populations. *Science*, 366(6464):447–453, 2019.
- [383] Cristina Godoy Bernardo de Oliveira, João Paulo Cândia Veiga, and Fabio G. Cozman. Regulação da Inteligência Artificial: Qual o Modelo Adotar. *Computação Brasil*, 7:28–31, 2022.
- [384] Raja Parasuraman, Thomas B Sheridan, and Christopher D Wickens. A model for types and levels of human interaction with automation. *IEEE Transactions on systems, man, and cybernetics-Part A: Systems and Humans*, 30(3):286–297, 2000.
- [385] Frank Pasquale. *New laws of robotics: defending human expertise in the age of AI*. Belknap Press, 2020.
- [386] Komal Patel. Testing the Limits of the First Amendment: How a CFAA Prohibition on Online Antidiscrimination Testing Infringes on Protected Speech Activity. *SSRN Electronic Journal*, pages 1–46, 2017.
- [387] Walt L Perry. *Predictive policing: The role of crime forecasting in law enforcement operations*. Rand Corporation, 2013.
- [388] Sundar Pichai. AI at Google: our principles. 2018-07-07, page 1, 2018.
- [389] Tony J Prescott and Julie M Robillard. Are friends electric? the benefits and risks of human-robot relationships. *Iscience*, 24(1):101993, 2021.
- [390] Iyad Rahwan, Manuel Cebrian, Nick Obradovich, Josh Bongard, Jean-François Bonnefon, Cynthia Breazeal, Jacob W Crandall, Nicholas A Christakis, Iain D Couzin, Matthew O Jackson, et al. Machine behaviour. *Nature*, 568(7753):477–486, 2019.
- [391] Byron Reeves and Clifford Nass. How people treat computers, television, and new media like real people and places, 1996.

- [392] Tahira Reid and James Gibert. Inclusion in human-machine interactions. *Science*, 375(6577):149–150, 2022.
- [393] Lionel P Robert. The Growing Problem of Humanizing Robots. *International Robotics & Automation Journal*, 3(1):1–2, 2017.
- [394] Margaret E Roberts. Censored. In *Censored*. Princeton University Press, 2018.
- [395] W Teed Rockwell. Algorithms and stories. *Human Affairs*, 23(4):633–644, 2013.
- [396] Heather M Roff. The strategic robot problem: Lethal autonomous weapons in war. *Journal of Military Ethics*, 13(3):211–227, 2014.
- [397] Jário Santos, Ig Bittencourt, Marcelo Reis, Geiser Chalco, and Seiji Isotani. Two billion registered students affected by stereotyped educational environments: an analysis of gender-based color bias. *HUMANITIES AND SOCIAL SCIENCES COMMUNICATIONS* 1, 9(249):1—16, 2022.
- [398] Daniel Schiff, Aladdin Ayesh, Laura Musikanski, and John C Havens. Ieee 7010: A new standard for assessing the well-being implications of artificial intelligence. In *2020 IEEE international conference on systems, man, and cybernetics (SMC)*, pages 2746–2753. IEEE, 2020.
- [399] Helen Sharp, Yvonne Rogers, and Jennifer Preece. *Interaction Design: Beyond Human-Computer Interaction*. John Wiley & Sons Inc, fifth edition, 2019.
- [400] Ben Shneiderman. Design lessons from ai’s two grand goals: Human emulation and useful applications. *IEEE Transactions on Technology and Society*, 1(2):73–82, 2020.
- [401] Ben Shneiderman. Human-centered artificial intelligence: Reliable, safe & trustworthy. *International Journal of Human-Computer Interaction*, 36(6):495–504, 2020.
- [402] Ben Shneiderman. The dangers of faulty, biased, or malicious algorithms requires independent oversight. *Proceedings of the National Academy of Sciences*, 113(48):13538–13540, 2016.
- [403] Ben Shneiderman, Catherine Plaisant, Maxine S Cohen, Steven Jacobs, Niklas Elmqvist, and Nicholas Diakopoulos. *Designing the user interface: strategies for effective human-computer interaction*. Pearson, 2016.
- [404] Ben Shneiderman. Human responsibility for autonomous agents. *IEEE intelligent systems*, 22(2):60–61, 2007.
- [405] Ben Shneiderman and Pattie Maes. Direct manipulation vs. interface agents. *interactions*, 4(6):42–61, 1997.
- [406] Ben Shneiderman. The future of interactive systems and the emergence of direct manipulation. *Behaviour & Information Technology*, 1(3):237–256, 1982.
- [407] Ben Shneiderman. Direct manipulation: A step beyond programming languages. In *Proceedings of the Joint Conference on Easier and More Productive Use of Computer Systems.(Part-II): Human Interface and the User Interface-Volume 1981*, page 143, 1981.
- [408] Marlo Souza. Tecnologias da Linguagem, Ética em IA e Regulamentação. *Computação Brasil*, 7:32–35, 2022.
- [409] Peter Stone, Rodney Brooks, Erik Brynjolfsson, Ryan Calo, Oren Etzioni, Greg Hager, Julia Hirschberg, Shivaram Kalyanakrishnan, Ece Kamar, Sarit Kraus, Kevin Leyton-Brown, David Parkes, William Press, AnnaLee (Anno) Saxenian, Julie Shah, Milind Tambe, and Astro Teller. Artificial Intelligence and Life in 2030: the one hundred year study on artificial intelligence. Technical report, Stanford University, September 2016.
- [410] Megan K Strait, Cynthia Aguillon, Virginia Contreras, and Noemi Garcia. The public’s perception of humanlike robots: Online social commentary reflects an appearance-based uncanny valley, a general fear of a “technology takeover”, and the unabashed sexualization of female-gendered robots. In *2017 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, pages 1418–1423. IEEE, 2017.
- [411] Barry Strauch. Ironies of automation: Still unresolved after all these years. *IEEE Transactions on Human-Machine Systems*, 48(5):419–433, 2017.
- [412] Wenlong Sun, Olfa Nasraoui, and Patrick Shafto. *Evolution and impact of bias in human and machine learning algorithm interaction*, volume 15. Public Library of Science San Francisco, CA USA, 2020.
- [413] Latanya Sweeney. Discrimination in online ad delivery. *Communications of the ACM*, 56(5):44–54, 2013.
- [414] Niko Tinbergen. On aims and methods of ethology. *Animal Biology*, 55(4):297–321, 2005.
- [415] Ufuk Topcu, Nadya Bliss, Nancy Cooke, Missy Cummings, Ashley Llorens, Howard Shrobe, and Lenore Zuck. Assured autonomy: Path toward living with autonomous systems we can trust. *arXiv preprint arXiv:2010.14443*, 2020.
- [416] Zeynep Tufekci. Youtube, the great radicalizer. *The New York Times*, 10(3):2018, 2018.

- [417] Zeynep Tufekci. Engineering the public: Big data, surveillance and computational politics. *First Monday*, 2014.
- [418] UN. Report of the Working Group on Internet Governance. Technical report, United Nations, 2005.
- [419] University of Montreal. Montréal Declaration for a Responsible Development of Artificial Intelligence. Technical report, University of Montreal, 2018.
- [420] Soroush Vosoughi, Deb Roy, and Sinan Aral. The spread of true and false news online. *science*, 359(6380):1146–1151, 2018.
- [421] Paul Voosen. The ai detectives, 2017.
- [422] Sara Wachter-Boettcher. *Technically wrong: Sexist apps, biased algorithms, and other threats of toxic tech*. WW Norton & Company, 2017.
- [423] Chathurika S Wickramasinghe, Daniel L Marino, Javier Grandio, and Milos Manic. Trustworthy ai development guidelines for human system interaction. In *2020 13th International Conference on Human System Interaction (HSI)*, pages 130–136. IEEE, 2020.
- [424] Christine T. Wolf and Jeanette L. Blomberg. Evaluating the promise of human-algorithm collaborations in everyday work practices. *Proceedings of the ACM on Human-Computer Interaction*, 3(CSCW):1–23, 2019.
- [425] David D Woods, James Tittle, Magnus Feil, and Axel Roesler. Envisioning human-robot coordination in future operations. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 34(2):210–218, 2004.
- [426] Wei Xu. Toward human-centered ai: a perspective from human-computer interaction. *interactions*, 26(4):42–46, 2019.
- [427] Sergio Amadeu Silveira. Governo dos algoritmos. *Revista de Políticas Públicas*, 21(1):267–281, 2017.
- [428] Can Yavuz. *Machine Bias Artificial Intelligence and Discrimination*. PhD thesis, Lund University, 2019.
- [429] Shunyuan Zhang, Nitin Mehta, Param Vir Singh, and Kannan Srinivasan. Can an ai algorithm mitigate racial economic inequality? an analysis in the context of airbnb. *An Analysis in the Context of Airbnb (January 21, 2021)*. Rotman School of Management Working Paper, 2021.
- [430] Ignas Kalpokas. *Algorithmic Governance*, volume 9. Palgrave Macmillan, 2019.
- [431] Meredith Broussard. *Artificial unintelligence: How computers misunderstand the world*. MIT Press, 2018.
- [432] César A Hidalgo, Diana Orghian, Jordi Albo Canals, Filipa De Almeida, and Natalia Martín. *How humans judge machines*. MIT Press, 2021.
- [433] U.S. Department of Education, Office of Educational Technology,. Artificial Intelligence and Future of Teaching and Learning: Insights and Recommendations. Technical report, U.S. Department of Education, 2023.
- [434] Alexis Lambert, Nahal Norouzi, Gerd Bruder, and Gregory Welch. A systematic review of ten years of research on human interaction with social robots. *International Journal of Human-Computer Interaction*, 36(19):1804–1817, 2020.
- [435] Yi Mou, Changqian Shi, Tianyu Shen, and Kun Xu. A systematic review of the personality of robot: Mapping its conceptualization, operationalization, contextualization and effects. *International Journal of Human-Computer Interaction*, 36(6):591–605, 2020.
- [436] Sarah Sebo, Brett Stoll, Brian Scassellati, and Malte F Jung. Robots in groups and teams: a literature review. *Proceedings of the ACM on Human-Computer Interaction*, 4(CSCW2):1–36, 2020.
- [437] Weiyu Wang and Keng Siau. Artificial intelligence, machine learning, automation, robotics, future of work and future of humanity: A review and research agenda. *Journal of Database Management (JDM)*, 30(1):61–79, 2019.
- [438] Lewis Mumfordt. *Technics and Civilization*. University of Chicago Press, Chicago, 1934.
- [439] Joseph Weizenbaum. *Computer Power and Human Reason: From Judgement to Calculation*. W. H. Freeman and Company, 1976.
- [440] C Dianne Martin. Eniac: press conference that shook the world. *IEEE Technology and Society Magazine*, 14(4):3–10, 1995.
- [441] Thomas S. Kuhn. *A Estrutura das Revoluções Científicas*. Perspectiva, São Paulo, 1 edition, 1996.
- [442] Cigdem BAŞFIRINCI and Zuhall ÇİLİNGİR. Anthropomorphism and advertising effectiveness: Moderating roles of product involvement and the type of consumer need. *Journal of Social and Administrative Sciences*, 2(3):108–131, 2015.

- [443] Tarleton Gillespie, Pablo J. Boczkowski, and Kirsten A. Foot, editors. *Media Technologies: Essays on Communication, Materiality, and Society*. The MIT Press, 2014.
- [444] Mark Granovetter and Roland Soong. Threshold models of diffusion and collective behavior. *Journal of Mathematical sociology*, 9:165–179, 1983.
- [445] Mark Granovetter. Threshold models of collective behavior. *American journal of sociology*, 83:1420–1443, 1978.
- [446] John A Bargh and Katelyn YA McKenna. The internet and social life. *Annu. Rev. Psychol.*, 55:573–590, 2004.
- [447] David Lazer, Alex Pentland, Lada Adamic, Sinan Aral, Albert-László Barabási, Devon Brewer, Nicholas Christakis, Noshir Contractor, James Fowler, Myron Gutmann, et al. Computational social science. *Science*, 323(5915):721–723, 2009.
- [448] Pierre Lévy. *L'intelligence collective: pour une anthropologie du cyberspace*. La découverte, 2013.
- [449] Anikó Hannák, Claudia Wagner, David Garcia, Alan Mislove, Markus Strohmaier, and Christo Wilson. Bias in online freelance marketplaces: Evidence from taskrabbit and fiverr. In *Proceedings of the 2017 ACM conference on computer supported cooperative work and social computing*, pages 1914–1933, 2017.
- [450] Leila Hudson, Colin S Owens, and Matt Flannes. Drone warfare: Blowback from the new american way of war. *Middle East Policy*, 18:122–132, 2011.
- [451] Peter M Krafft, Michael Macy, and Alex" Sandy" Pentland. Bots as virtual confederates: design and ethics. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*, pages 183–190, 2017.
- [452] Bjarke Mønsted, Piotr Sapieżyński, Emilio Ferrara, and Sune Lehmann. Evidence of complex contagion of information in social media: An experiment using twitter bots. *PloS one*, 12(9):e0184148, 2017.
- [453] Yuhua Liang and Seungcheol Austin Lee. Fear of autonomous robots and artificial intelligence: Evidence from national representative data with probability sampling. *International Journal of Social Robotics*, 9(3):379–384, 2017.
- [454] Kathleen Richardson, Mark Coeckelbergh, Kutoma Wakunuma, Erik Billing, Tom Ziemke, Pablo Gomez, Bram Vanderborght, and Tony Belpaeme. Robot enhanced therapy for children with autism (dream): A social model of autism. *IEEE Technology and society magazine*, 37(1):30–39, 2018.
- [455] Michael P Wellman and Uday Rajan. Ethical issues for autonomous trading agents. *Minds and Machines*, 27(4):609–624, 2017.
- [456] Alan FT Winfield and Marina Jirotko. The case for an ethical black box. In *Annual Conference Towards Autonomous Robotic Systems*, pages 262–273. Springer, 2017.
- [457] Alan FT Winfield and Marina Jirotko. Ethical governance is essential to building trust in robotics and artificial intelligence systems. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 376(2133):20180085, 2018.
- [458] Néstor García Canclini. *Ciudadanos reemplazados por algoritmos*. Calas, 2019.
- [459] Neil McBride. Robot enhanced therapy for autistic children: An ethical analysis. *IEEE Technology and Society Magazine*, 39(1):51–60, 2020.
- [460] Marc Wiedermann, E Keith Smith, Jobst Heitzig, and Jonathan F Donges. A network-based microfoundation of granovetter's threshold model for social tipping. *Scientific reports*, 10(1):1–10, 2020.
- [461] Colm Kearns, Gary Sinclair, Jack Black, Mark Doidge, Thomas Fletcher, Daniel Kilvington, Katie Liston, Theo Lynn, and Pierangelo Rosati. A scoping review of research on online hate and sport. *Communication & Sport*, pages 1–29, 2022.
- [462] Adam Farquhar, Richard Fikes, and James Rice. The ontolingua server: A tool for collaborative ontology construction. *International journal of human-computer studies*, 46(6):707–727, 1997.
- [463] Oren Etzioni, Michele Banko, Stephen Soderland, and Daniel S Weld. Open information extraction from the web. *Communications of the ACM*, 51(12):68–74, 2008.
- [464] Oren Etzioni, Anthony Fader, Janara Christensen, Stephen Soderland, and Mausam Mausam. Open information extraction: The second generation. In *IJCAI*, volume 11, pages 3–10, 2011.
- [465] Agnieszka Konys and Zygmunt Drajek. Ontology learning approaches to provide domain-specific knowledge base. *Procedia Computer Science*, 176:3324–3334, 2020.