Smart move? A case study on the operationalization of transparency and explainability in an open source career transition algorithm

Chantale Tippett[a]

[a] *Nesta, 58 Victoria Embankment, London, EC4Y 0DS, United Kingdom*

Abstract

Anxiety over jobs being lost as a result of technological change has been rising and falling for centuries, however the most dystopian scenarios of mass unemployment have never fully materialized. The expanding capabilities of powerful new artificial intelligence technologies have prompted some to question whether this time really is different. In light of these concerns, a wide array of smart algorithmic systems are being developed to assist individuals, career advisors, and government departments in navigating the new labor market landscape. Often framed as neutral and apolitical decision assistance tools, some of these systems have nonetheless encountered critiques in the academic and public realms due to their black box nature or ability to entrench systemic inequalities. In this paper, we explore how the concepts of transparency and explainability were operationalized by a team designing an open source career transition algorithm. We find that transparency and explainability are considered to be broadly synonymous with openness, and the translation of these values into action is influenced by a range of factors including team and organizational culture, informal benchmarking against peer projects, as well as goals such as increasing user trust in the algorithm, ensuring widespread access, and hopes that others in the field will build on the project's outputs. The design team viewed both technical (e.g. algorithm design) and non-technical (e.g. stakeholder engagement) activities as important components of ensuring transparency and explainability, and considerations emerged primarily around how these values should be built into project inputs (e.g. the choice of data sources and algorithmic logic) and outputs (e.g. decisions about when to release the algorithm publicly). While the open sourcing of the algorithm is central to advancing the transparency and explainability goals of the project, it also raises questions about longer-term accountability and complicates the *ex-ante* assessment of the impact it will have at the level of the socio-technical system as it has the potential to be taken up and used by many different actors with differing aims. This paper adds to a nascent but much needed literature on the development of career guidance algorithms, and although caution should be exercised when drawing conclusions from a single case study, it proposes several recommendations for other teams working in this field.

Keywords 1

Transparency; Explainability; Machine Learning; Labor Market

1. Introduction

Concerns about widespread job losses caused by technological change (a phenomenon referred to as 'technological unemployment') have been rising and falling since at least the Industrial Revolution, however the most dystopian scenarios have historically proven to be overblown as new jobs and industries rose from the ashes [1]. As increasingly powerful artificial intelligence (AI) technologies start to master the previously automation-immune tasks performed in white collar, well paid

Joint Proceedings of the ACM IUI 2021 Workshop, April 13-17, 2021, College Station, USA
Chantale.Tippett@nesta.org.uk

© 2020 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).
CEUR Workshop Proceedings (CEUR-WS.org)

CEUR-WS.org/Vol-2903/IUI21WS-TExSS-15.pdf

careers such as medicine and law, anxiety has again started to grow with some questioning whether this time really is different. This fear has been compounded by what appears to be an acceleration of trends toward automation during the Covid-19 pandemic [2].

At the same time that concerns are being raised about technology's role in displacing labor, AI has been increasingly adopted as a tool for coping with the new situation. The development of career transition algorithms for individual, private sector and government use has grown rapidly in recent years as people look for ways to navigate the changing landscape. Their proponents have expressed hope that these tools will make career transition advice more efficient, effective, and personalized, ultimately leading to more opportunities for those who are most at risk of displacement. However, they have also been subject to critique in both the academic and public realms. Both the optimistic and critical camps have raised important questions about how transparency and explainability are enacted in career transition algorithms, however little is known about how design teams navigate these decisions in real-world contexts. In this paper, we explore the operationalization of transparency and explainability in a career transition algorithm project. In the vein of an emerging body of ethnographic and evaluative work on algorithm design, this work engages directly with the design team to explore the technical and non-technical aspects of transparency and explainability in the project, as well as the institutional and contextual influences on design choices.

2. Background

2.1. A brief history of automation and technological unemployment

Anxieties around widespread technological unemployment tend to rise and fall. The Industrial Revolution was the first major technological transition that was extensively written about in real time, and from the 18th century onward, economists and others have debated the short- and long-term impacts of technological advancements on levels of employment, wages, and the quality of work [3]–[5]. For instance, in a set of lectures delivered in 1928, economist John Maynard Keynes warned that unemployment due to technological innovation was outpacing our ability to find new uses for labor. Over two decades later, Nobel Prize-winning economist Wassily Leontief cautioned that workers were increasingly being replaced by machines, and that it was not clear that new industries would be able to employ everybody who wants a job [1]. However, concerns about widespread technological unemployment have largely turned out to be overblown. To illustrate this dynamic through one concrete example, fears that automated teller machines (ATMs) would lead to job losses proved false because ATMs allowed banks to operate branch offices at lower cost, prompting them to open many more branches. This expansion ultimately offset the loss in teller jobs such that the number of full-time equivalent bank tellers in the United States grew following the deployment of ATMs in the 1970s [6].

2.2. Contemporary debates on technological unemployment

Although prophecies of mass unemployment have largely failed to materialize, technological advancement has nonetheless had an important impact on the labor market. Earlier waves of automation have primarily affected low- and middle-skill occupations[2], with impacts being felt in manual labor industries such as agriculture and routine tasks such as bookkeeping [7]. Early algorithms were developed in an attempt to directly replicate human decision making, limiting the range of tasks they could automate to those that were routine, repetitive and codifiable. However, advances in techniques such as natural language processing, predictive analytics and image recognition have opened new pathways to automation, leading some to wonder whether this wave of automation risk really is different from those of the past [1], [8]. In 2013, an Oxford University study set off a wave of panic when they reported that 47% of people working in the United States are in jobs

[2] The language of 'low-skill' and 'middle-skill' is common in the literature on this topic, and is not intended here as a normative judgment.

that could be performed by computers or algorithms within the next 10-20 years [9][3]. A great deal of methodological development has since taken place[4], with later work from organizations such as the OECD finding that in fact few jobs have either very high or very low risk of automatability, providing a reprieve from the concerning picture painted by the initial findings from the Oxford study [10]. However, these revised estimates still suggest that the impacts of automation will be felt primarily by lower-skilled, lower-income workers, raising important equity concerns.

2.3. Career transition algorithms: opportunities and controversies

As the debate about the risks of technological unemployment have raged and methodologies for estimating automatability risk have evolved, governments and other organizations have started turning to algorithmic solutions to help navigate the changing labor market, and investment in these technologies has started to grow [11]. For instance, Silicon Valley-based EightfoldAI have raised nearly $180 million USD to develop an 'AI-powered Talent Intelligence Platform' that is described in the company's promotional material as 'the most effective way for organizations to retain top performers, upskill and reskill the workforce, recruit top talent efficiently, and reach diversity goals' [12].

The deployment of algorithmic tools to navigate the changing labor market has also at times been met with critical scrutiny in the domains of public and academic discourse. For instance, Allhutter and colleagues critically interrogate an Austrian job seeker profiling tool which aims to increase the efficiency of government career counselling processes and the effectiveness of active labor market programs [13]. Through an assessment of the tool's technical documentation as well as policy documents (e.g. labor market policy targets), they call into question the system's purported neutrality and reveal the ways in which it helps to enact the framing of unemployment under austerity politics. The authors also found that despite trying to present an image of transparency, significant omissions and under-specifications in the documentation inhibit the achievement of meaningful explanations [13]. Controversies have also arisen in the United Kingdom (UK), where the government's National Career Services tool was met with ridicule and hostility following a series of seemingly illogical career transition recommendations [14], and the re-emergence of a separate ad campaign that appeared to imply that people in creative industries should retrain in areas like cybersecurity [15]. Taken together, these examples demonstrate that career transition advice can be politically charged, and highlights the need to adopt an equally politically aware stance while developing any intervention or tool that seeks to assist people or institutions navigating the labor market.

2.4. Transparency and explainability in smart systems

In a 2019 review on the global landscape of AI ethics published in Nature Machine Learning, transparency was the most commonly cited value, appearing in 73 out of 84 guidelines reviewed [16]. Despite this widespread emphasis, however, the authors of the review report widespread variability in how transparency is interpreted, justified, applied and evaluated in AI ethics guidelines. This finding is consistent with a growing body of taxonomies of different transparency types. For instance, Weller and colleagues [16] identified eight distinct types of transparency, noting that each would require a different sort of explanation and different measures of efficacy:

- Type 1: For a developer, to understand how their system is working, aiming to debug or improve it, to see what is working well or badly, and get a sense for why.
- Type 2: For a user, to provide a sense for what the system is doing and why, to enable prediction of what it might do in unforeseen circumstances and build a sense of trust in the technology.
- Type 3: For society broadly to understand and become comfortable with the strengths and limitations of the system,

[3] Note that the working paper was published in 2013 but the final version of the study was published in 2017.

[4] For instance, switching from occupational to task-level estimates of automation risk produces drastically different results.

overcoming a reasonable fear of the unknown.

- Type 4: For a user to understand why one particular prediction or decision was reached, to allow a check that the system worked appropriately and to enable meaningful challenge.
- Type 5: To provide an expert (e.g. a regulator) the ability to audit a prediction or decision trail in detail, particularly if something goes wrong and requires the assignment of accountability or legal liability.
- Type 6: To facilitate monitoring and testing for safety standards.
- Type 7: To make a user feel comfortable with a prediction or decision so that they keep using the system. This type of transparency primarily benefits the deployer.
- Type 8: To lead a user into some action or behavior such as making a purchase. As with Type 7, this type of transparency primarily benefits the deployer.

Despite the widespread interest in promoting transparency and explainability as well as the broadly held view that this is a worthwhile end in and of itself, recent work has called into question if and how transparency can be operationalized in practice, and whether this is always a desirable goal [17], [18]. For instance, emerging evidence suggests that in certain circumstances, signifiers of transparency such as open sourcing may actually have the unintended effect of producing less – rather than more – critical engagement with algorithmic outputs. Kemper and Kolkman illustrate this point in relation to the UK government's 2050 Calculator, which is an open source energy and emissions model [18]. The developers of this tool found that very few people looked into the documentation, and felt that by open sourcing the model, people were less inclined to contest its outcomes.

3. Case study: transparency and explainability in the development of a career transition algorithm

3.1. Background and case study approach

The algorithmic system that is the focus of this paper aims to identify career transitions that are viable, desirable, and safe in the context of growing concerns around automation. The onset of the Covid-19 pandemic, which has led to widespread job losses and an acceleration of automation trends, has added an additional layer of perceived urgency to the development and deployment process. The algorithm was developed by a non-profit foundation with funding from a second non-profit foundation. It maps similarities between over 1,600 occupations based on the skills and tasks that make up each role, and can be used to identify a set of jobs requiring similar skills and activities. The algorithm can also identify skills that a worker might need to develop in order to move into a new role. 'Desirable' transitions are defined as those that would incur a limited loss of earnings, while 'safe' transitions are those that would also lead to a lower automation risk. The risk of automation is assessed according to estimates of the suitability of tasks for machine learning [19]. At the time of writing, a report detailing the research findings created using the algorithm had been published, however the algorithm and datasets had not yet been publicly released.

Despite the many challenges that exist in studying algorithms (e.g. their black boxed and contingent nature), we adopt one of the methods outlined by Kitchin (2017), which includes interviewing the algorithm designers to understand how their objectives were framed and translated into code, as well as what influences, constraints, and other factors influenced their approach [20]. The findings below are drawn from a series of discussions held with the algorithm's design team in November and December 2020, in which we sought to collectively explore and critically evaluate how transparency and explainability were taken into account in the project. The core team includes two data scientists who were responsible for designing and performing technical validation on the algorithm; a data visualization expert who was responsible for developing a user interface; and two program managers who were responsible for stakeholder engagement and the overall functioning of the project.[5]

3.2. Operationalizing transparency and explainability in the project

The terms 'transparency' and 'explainability' were not explicitly defined in the project, however members of the development team noted that the value of 'openness' was embedded within both the team's core values and the organization's broader charitable objectives, and that this was largely seen to be synonymous with transparency. Team members put forward multiple reasons for wanting to work openly and transparently, including fostering trust in the final product, increasing the likelihood that the outputs would continue to be developed by others, and ensuring that users with fewer resources were still able to work with the tool.

3.3. Transparency and explainability across project inputs, outputs and socio-technical system levels

The project team identified salient questions and actions around transparency and explainability at multiple levels in the project, noting that it may not be possible or desirable to achieve all of them simultaneously.

3.3.1. Transparency and explainability of inputs

Members of the project team suggested that while it was possible to control much of the work around transparency in the project undertaken by the team, questions remained about how to make the work that underpinned the project (e.g. inputs such as the underlying logic, code, etc.) more transparent to end users. On the one hand, the team felt that the use of open data sources that are widely accepted in the field, as well as the use of open source tools (e.g. Python packages), increase the transparency of project inputs. On the other hand, as noted by one of the data scientists, the algorithm relies on an approach developed by a team of academics in the United States, which contains its own sets of assumptions and limitations that may not be easily accessible or comprehensible to an average end user[6]. These limitations nonetheless have implications for how the algorithm can and should be used, and although these are flagged in the published report, the team expressed concern that there may still be assumptions or underlying logics embedded in the algorithm that do not get translated into the way it is ultimately deployed.

3.3.2. Transparency and explainability of outputs

The project team pointed to multiple mechanisms for ensuring the transparency of the outputs, including both technical and non-technical approaches. On the non-technical side, stakeholder engagement over the course of the project played an important role in legitimizing the approach taken and ensuring it was well understood by stakeholders. More broadly, discussions with other organizations working in a similar space were crucial in identifying possible risks and opportunities around transparency and explainability. For instance, one project member noted that engaging with an organization developing a black box career transition algorithm served as encouragement to be more transparent. Similarly, the team reported feeling influenced by public controversies around other career transition algorithms, providing them with further motivation to ensure that pre-release validation processes were robust. As one team member noted: "all it takes is one bad recommendation for people to lose trust".

On the technical side, transparency and explainability were operationalized at the output level primarily through the decision to open source the algorithm following a validation process that included crowdsourcing feedback on the transition recommendations. The lead algorithm designer also emphasized that specific design choices were taken with end

[5] Two additional team members (a data scientist and qualitative researcher) were on parental leave and therefore unable to participate in the discussions.

[6] Although part of the rationale for using the method developed in the US study was that it was itself open (both the code and the data), which appealed to the team from a transparency perspective.

user transparency and explainability in mind, highlighting two specific examples. First, the team used fully interpretable features that allow a clear assessment of which elements contribute most highly to similarities between work activities or work contexts. Second, the team privileged the use of natural language processing methods for comparing job skill sets that allowed for specific skill matches and gaps to be clearly identified, as well as a determination of which skills are contributing the most to the similarity score.

The timing of transparency-enhancing activities was also a factor the team took into account, with one data scientist indicating that releasing the algorithm prior to the necessary validation processes could potentially cause more harm than good. This is why the team preferred to work in a private GitHub repository until the algorithm's release. The team also cautioned against total transparency, noting that decisions about what to show the end user and what to hold back (e.g. in the design of the user interface) are actually a core component of making the tool useful, as showing too much information can also be confusing and lead to decreased interpretability. In this sense, perhaps counterintuitively, the team felt that some information needed to be 'hidden' in an attempt to make the outputs more usable (while noting that this information could still be accessed through other means if desired).

3.3.3. Transparency and explainability within the socio-technical system

Transparency was also discussed by the project team in terms of how the algorithm would interact with the broader socio-technical system in which it's embedded, with one project team member pointing out that "transparency is in the DNA of the project" by virtue of it being aimed at making labor market information more accessible, thereby informing more transparent decision-making by actors within the system (e.g. career counsellors). However, the team agreed that there was some uncertainty about how transparency would be operationalized once the algorithm is made public. For instance, it is unclear to what extent its use can realistically be monitored or whether approaches such as terms of service would provide any real protection against unintended uses. For instance, the team emphasized that the tool is meant to augment the work of career counsellors rather than replace them. They noted that this is especially important given evidence that labor market information alone (without the support from a counsellor) has been shown to have very limited impact on job seeker outcomes [21]. Despite having engaged with career counsellors over the course of the project to ensure this point was clear, the development team acknowledged that there was ultimately no way to guarantee that end users would respect one of the project's core guiding principles that the algorithm be used as a complement to existing tools and processes rather than being used as a replacement. The discussion around the socio-technical level also raised questions about counterfactual scenarios, with team members suggesting that any risks posed by the algorithm should be weighed against the status quo situation.

4. Discussion

Exploring how transparency and explainability were operationalized in a real-world project to develop a career transition algorithm provides an insightful window into the types of questions, considerations and trade-offs at play in practice. The finding that the project team didn't see an important distinction between commonly used AI ethics values such as openness, transparency or explainability is consistent with literature showing that these terms are variably interpreted and applied [16], [17]. The lack of explicit definitions may have allowed for a more flexible and adaptable approach, where actions could evolve organically as needed over the course of the algorithm's development in response to external events such as media critiques of other career transition tools or algorithms, or conversations with other organizations working in a similar space. Indeed, it appears as though informal peer learning and benchmarking played an important role in shaping the project team's perceptions of what efforts were required to foster trust amongst users and other stakeholders, and the ways that this could be achieved through increased transparency. This finding suggests that there is merit in recommendations recently put forward that encourage greater transparency around failures as well as successes in AI development [22].

Despite providing more flexibility, the absence of a clear framework for defining ‘transparency’, ‘explainability’ or ‘openness’ also presents a number of challenges, including the absence of metrics against which success (or failure) can be assessed and a way of tracking trade-offs that were made so that these can be more widely understood and openly interrogated. As described above, the growing literature around transparency types [17], [23], could provide more structure and granularity around what transparency and explainability mean to different audiences or stakeholders in practice. These could be combined with tools such as an algorithmic design history file [24] for tracking design decisions, value alignment, and findings from risk analysis assessments. The development of a set of evaluative metrics to assess not only whether existing approaches to transparency are effective, but also whether they are succeeding in inviting critical feedback, should also be prioritized [18].

The finding that transparency and explainability were most consistently considered and incorporated into the input and output levels of the project, with relatively less focus on socio-technical system considerations, merits further critical analysis. As described in detail elsewhere, evaluations of algorithmic design and logic can only bring us so far in understanding what their effects might be. Gaining a deeper understanding requires us to interrogate how they become embedded in broader sociotechnical systems [20], [25]. However, this is also one of the most challenging tasks to undertake, particularly *ex-ante*, for at least two reasons. The first challenge is practical in nature because in this case, the team does not know exactly who will use the open source algorithm once it’s released, so it is difficult to assess if or how the value of transparency can or should be enacted once it’s taken up by others (and if so, who is responsible for this work). A second challenge which is not unique to this project but rather applies to technology development more broadly, is captured in the Collingridge Dilemma, which highlights the fact that the impact of a given technology is to some extent unknowable until it has been integrated into a given socio-technical system, at which point it is difficult or impossible to change or control it [26]. These challenges, as well as ways of mitigating risks and potential harms to the greatest extent possible while granting that some impacts will inevitably be unknowable a priori, deserve more attention in the transparency and explainability literature.

5. Conclusion

As individuals, companies and governments attempt to navigate changes in the labor market caused by technological innovation and the displacement of workers due to the Covid-19 crisis, it is likely that the use of AI-assisted tools will grow to meet the need for careers information, as well as the demand for reskilling and upskilling advice at an unprecedented scale. It is therefore essential to explore the ways in which algorithm design teams frame, operationalize and measure the success of efforts aimed at increasing the transparency and explainability of these systems.

In this short case study on how transparency and explainability were operationalized in a project to develop a career transition algorithm, we found that these values are considered to be broadly synonymous with openness, and that many considerations are at play when framing the goals, drivers and barriers toward this end. We also found that efforts to advance the aims of transparency and explainability were implemented primarily at the level of the project inputs and outputs, rather than at the level of the socio-technical system.

This paper adds to a sparse but growing body of literature that critically analyses career guidance algorithms [13]. Although we should avoid inferring too much from a single case study, we nonetheless identify three key take-aways from this assessment. The first is that algorithm development teams should agree upon what ‘transparency’ and ‘explainability’ mean at the outset of the design process. As described above, a wide range of new taxonomies at varying levels of granularity and for different users have been developed in recent years. These could be deployed alongside tools such as design history files to strike a balance between conceptual clarity and flexibility. The second key insight is that peer learning mechanisms (broadly interpreted to include cases arising in the media or academic literature as well as discussions with other teams working in a similar space) can prompt helpful reflections on the types of risks and issues that should be taken into account during algorithmic development and deployment. Mechanisms for sharing lessons learned, particularly from failures, should be further developed and encouraged. Finally, the open sourcing of algorithms such as the one developed in this case study creates new opportunities for navigating shocks to the labor market while also creating new risks that it might be used in unintended ways. Further research should focus on elaborating and developing mechanisms for monitoring and accountability in such instances.

6. References

[1] J. Mokyr, C. Vickers, and N. L. Ziebarth, "The History of Technological Anxiety and the Future of Economic Growth: Is This Time Different?," J. Econ. Perspect., vol. 29, no. 3, pp. 31–50, Aug. 2015, doi: 10.1257/jep.29.3.31.

[2] OECD, "OECD Employment Outlook 2020," OECD, 2020. http://www.oecd.org/employment-outlook/2020 (accessed Sep. 21, 2020).

[3] D. H. Autor, "Why Are There Still So Many Jobs? The History and Future of Workplace Automation," J. Econ. Perspect., vol. 29, no. 3, pp. 3–30, Aug. 2015, doi: 10.1257/jep.29.3.3.

[4] R. C. Allen, "Engels' pause: Technical change, capital accumulation, and inequality in the british industrial revolution," Explor. Econ. Hist., vol. 46, no. 4, pp. 418–435, Oct. 2009, doi: 10.1016/j.eeh.2009.04.004.

[5] J. Clifton, A. Glasmeier, and M. Gray, "When machines think for us: the consequences for work and place," Camb. J. Reg. Econ. Soc., vol. 13, no. 1, pp. 3–23, May 2020, doi: 10.1093/cjres/rsaa004.

[6] J. E. Bessen, "How Computer Automation Affects Occupations: Technology, Jobs, and Skills," SSRN Electron. J., 2015, doi: 10.2139/ssrn.2690435.

[7] D. H. Autor, "Work of the Past, Work of the Future," AEA Pap. Proc., vol. 109, pp. 1–32, May 2019, doi: 10.1257/pandp.20191110.

[8] D. Acemoglu and P. Restrepo, "The wrong kind of AI? Artificial intelligence and the future of labour demand," Camb. J. Reg. Econ. Soc., vol. 13, no. 1, pp. 25–35, May 2020, doi: 10.1093/cjres/rsz022.

[9] C. B. Frey and M. A. Osborne, "The future of employment: How susceptible are jobs to computerisation?," Technol. Forecast. Soc. Change, vol. 114, pp. 254–280, Jan. 2017, doi: 10.1016/j.techfore.2016.08.019.

[10] M. Arntz, T. Gregory, and U. Zierahn, "The Risk of Automation for Jobs in OECD Countries: A Comparative Analysis," OECD Social, Employment and Migration Working Papers 189, May 2016. doi: 10.1787/5jlz9h56dvq7-en.

[11] B. Dellot, R. Mason, and F. Wallace-Stephens, "The four futures of work," The RSA Action and Research Centre, London, Mar. 2019. Accessed: Dec. 21, 2020. [Online]. Available: https://www.thersa.org/globalassets/pdfs/reports/rsa_four-futures-of-work.pdf.

[12] Crunchbase, "Eightfold - Crunchbase Company Profile & Funding," Crunchbase, Dec. 21, 2020. https://www.crunchbase.com/organization/eightfold (accessed Dec. 21, 2020).

[13] D. Allhutter, F. Cech, F. Fischer, G. Grill, and A. Mager, "Algorithmic Profiling of Job Seekers in Austria: How Austerity Politics Are Made Effective," Front. Big Data, vol. 3, p. 5, Feb. 2020, doi: 10.3389/fdata.2020.00005.

[14] M. Weaver, "Boxer or lock-keeper? Government careers quiz scorned by users," the Guardian, Oct. 09, 2020.

[15] D. Jordan, "Downing Street joins criticism of 'crass' job ad," BBC News, Oct. 12, 2020.

[16] A. Jobin, M. Ienca, and E. Vayena, "The global landscape of AI ethics guidelines," Nat. Mach. Intell., vol. 1, no. 9, Art. no. 9, Sep. 2019, doi: 10.1038/s42256-019-0088-2.

[17] A. Weller, "Transparency: Motivations and Challenges," ArXiv170801870 Cs, Aug. 2019, Accessed: Dec. 21, 2020. [Online]. Available: http://arxiv.org/abs/1708.01870.

[18] J. Kemper and D. Kolkman, "Transparent to whom? No algorithmic accountability without a critical audience," Inf. Commun. Soc., vol. 22, no. 14, pp. 2081–2096, Dec. 2019, doi: 10.1080/1369118X.2018.1477967.

[19] E. Brynjolfsson, T. Mitchell, and D. Rock, "What Can Machines Learn and What Does It Mean for Occupations and the Economy?," AEA Pap. Proc., vol. 108, pp. 43–47, May 2018, doi: 10.1257/pandp.20181019.

[20] R. Kitchin, "Thinking critically about and researching algorithms," Inf. Commun. Soc., vol. 20, no. 1, pp. 14–29, Jan. 2017, doi: 10.1080/1369118X.2016.1154087.

[21] R. Alexander, G. McCabe, and M. De Backer, Careers and Labour Market Information: An International Review of the Evidence. Education Development Trust, 2019.

[22] E. Moss and J. Metcalf, "Ethics Owners: A New Model of Organizational Responsibility in Data-Driven Technology Companies," Data Soc., p. 74, 2020.

[23] J. Cowls, T. King, M. Taddeo, and L. Floridi, "Designing AI for Social Good: Seven Essential Factors," Social Science Research Network, Rochester, NY, SSRN Scholarly Paper ID 3388669, May 2019. doi: 10.2139/ssrn.3388669.

[24] I. D. Raji et al., "Closing the AI Accountability Gap: Defining an End-to-End Framework for Internal Algorithmic Auditing," ArXiv200100973 Cs, Jan. 2020, Accessed: Jul. 04, 2020. [Online]. Available: http://arxiv.org/abs/2001.00973.

[25] M. Sendak et al., "'The Human Body is a Black Box': Supporting Clinical Decision-Making with Deep Learning," ArXiv191108089 Cs, Dec. 2019, Accessed: Jan. 04, 2021. [Online]. Available: http://arxiv.org/abs/1911.08089.

[26] B. Ribeiro et al., "Introducing the dilemma of societal alignment for inclusive and responsible research and innovation," J. Responsible Innov., vol. 5, no. 3, pp. 316–331, Sep. 2018, doi: 10.1080/23299460.2018.1495033.