Literature Review: A Critical Analysis of Ethical Problems Related to Algorithmic Governance

AI and Ethics (CT5142)

Michael Rice (20347541)

James Conroy (24230670)

Jonas Michel (24238749)

26th November 2024

1 Introduction

In a world where technology shapes nearly every facet of human life, algorithms and their developers have become the silent architects of societal decision-making. Lessig (2009) introduced the notion that "code is law", a perspective that increasingly resonates with today's ever-evolving digital landscape [1].

Algorithmic Governance, as defined by Just and Latzer (2017), is the idea of a socio-technical system where norms and rules are defined and enforced by "algorithms", on both the "horizontal and vertical extensions of traditional government" [2]. The horizontal extension involves looking beyond public entities into the interactions of governing bodies with private actors, while the vertical extension is concerned with the strata that combine to form such an entity [2]. Within these two dimensions algorithms can shape human behaviours, perceptions and social orderings.

In the context of Algorithmic Governance, algorithms are defined, by social scientists, not only as the algorithms themselves, but also encompasses those who design them, those who they effect, the data on which they act and the bodies that allow for their operation [3].

In the following, a comprehensive literature review of Algorithmic Governance as a concept will be conducted while critically evaluating its impact on society today. In this exploration, particular attention will be given to significant sub-topics, including bias, transparency, and control in relation to algorithms, with a consistent emphasis on real-world examples to illuminate these issues.

2 Algorithmic Governance - Code is Law

Any critical examination of algorithmic governance necessitates firstly addressing key building blocks. As mentioned, bias, transparency and control emerge as pivotal themes, each with a unique footprint on the conversation.

Bias, in an algorithmic context, refers to any systematic distortion or unfair outcomes arising from data characteristics or design choices [4]. Transparency denotes the extent to which algorithms and their internal processes and decisions are accessible and understandable to all stakeholders [5]. Finally, control, in relation to algorithms and software, pertains to mechanisms through which human oversight is maintained over these systems [6]. Ensuring a balance between the three of these is crucial for the innovation of fair and ethically aligned algorithms, whilst keeping the relevant stakeholders accountable [7].

The case studies we analyse below show that algorithmic governance is not a singular clearly structured concept, but takes different forms in different contexts and is shaped by varying degrees of transparency.

2.1 Assessing Recidivism Risk: COMPAS

An illustrative case where this delicate balance was disrupted is the COMPAS system, an algorithm developed by Northpointe at the request of the U.S. government to assess criminal risk in areas such as pretrial release, general recidivism, and violent recidivism [8, 9]. Widely considered as a symbol of algorithmic injustice, COMPAS has faced mass scrutiny for its evident bias and lack of transparency [8].

To critically analyse COMPAS, it is essential to delve into its flawed design principles and the resulting outcomes. Initially, the very base principles of the algorithm's design were erroneous [9]. In a 2016 ProPublica investigation of the algorithm it was found that those part of racial minorities were labelled 1.9 times more likely to re-offend than their white counterparts [8, 9, 10]. This bias, through some difficulty due to Northpointe's refusal to divulge the principles of their algorithm's operation, can eventually be attributed to several key design choices [9, 10].

Firstly, a policy that treated false negatives 2.6 times more punishing than false positives [11] was built on the idea that it is "much more dangerous to release Darth Vader than it is to incarcerate Luke Skywalker" [12]. Secondly, the data used to train the algorithm featured zip codes, which became an effective proxy for an individual's race [13]. A final factor that undermined the system's ethical legitimacy was the evident misalignment between its intended task and the training data used in its development. Northpointe's algorithm was trained exclusively on data reflecting repeat arrests rather than convictions, and on reported crimes rather than verified occurrences [11].

Moreover, the unjust outcomes generated by COMPAS were exacerbated by a widespread lack of understanding regarding its underlying operations, a deficiency that extended even to the U.S. government, meaning the algorithm was not controlled or known to those who deployed it [14]. This was perceived by many, particularly those affected by unjust decisions, as an abdication of responsibility and a glaring absence of accountability on the part of the governing bodies [14]. One individual adversely impacted by such a decision was Eric Loomis, who received a six-year prison sentence in Wisconsin after being found driving a car involved in a shooting, a crime not typically warranting imprisonment [14]. Loomis contended that the use of proprietary software in his sentencing violated his right to due process, as he was denied the opportunity to challenge the accuracy of the algorithm's assessment [15].

2.2 Predictive Policing: HunchLab

HunchLab, now owned by ShotSpotter, is a predictive policing software used in the United States. In contrast to COMPAS, HunchLab has a record of being transparent with its internal practices [16]. Using machine learning algorithms, the system predicts where and when crimes will occur, allowing police departments to effi-

ciently focus their policing efforts on places that need it most [16]. We analysed the ethical concerns relating to algorithmic governance based on HunchLab's 2017 citizen's guide [16] and relevant literature around that time.

In addition to basic crime-related information such as the specific location it occurred, type of crime and date/time of reporting, HunchLab's algorithms are trained on extensive data such as known offender's locations, locations of likely targets of crime, weather, seasonal cycles of crime and socioeconomic factors [16]. The processing of such large amounts of data raises significant privacy concerns. Without adequate data security, leaking of information such as known offenders and potential targets could result in catastrophic consequences [17].

Of the algorithms studied by Brauneis and Goodman (2018), HunchLab's is the most dynamic [11]. They do a modelling run with the latest information every few weeks to re-calibrate the model, creating a new algorithm each time. Their citizen's guide discloses use of "gradient boosting decision trees", which are more black-box in nature compared to regular decision trees [18]. Black-box algorithms typically raise transparency concerns, but even more so in this context given that the black-box is ever-changing [19]. This opaqueness can be understandable, however, if the objective is to prevent manipulation or gaming of the algorithms [11]. Randomness is also introduced to frustrate efforts by criminals to predict patrolling plans and to ensure officers aren not monotonously assigned to same routes each day [16, 17].

Even if predictive policing can accurately predict where the crime "hot spots" are, police officers cannot know how the algorithm's decision making relates to their own knowledge and judgement [11]. If officers arrive too late or in the wrong location to prevent a murder taking place, is accountability now delegated to the algorithmic powers that be? Whilst algorithms cannot tell the police department that rival gangs are about to engage in a violent confrontation, it is possible that a local resident could [20]. The over-reliance on technology-driven policing risks staff losing their autonomy, and disenfranchising communities who have historically played a key role in providing valuable information about what is happening on the ground [7].

A key bias in HunchLab's algorithms is the fact that they are based on where crimes are reported, and not necessarily where all crimes occur [16]. Studies have shown that crimes are underreported in disadvantaged and immigrant neighbourhoods [21, 22]. Thus, an algorithm trained on reported crimes may end up directing police away from the most vulnerable in society [11]. HunchLab openly discuss this in their Citizen's Guide, believing that "if reporting biases are due to distrust of the police, then we believe that letting the bias exist within the data is appropriate" [16]. This acknowledgment raises fundamental questions about whether predictive policing can ever be ethically implemented when its core functionality depends on perpetuating existing societal inequalities.

3 Proposed Frameworks and Guidelines

Algorithms have the potential to be unbiased if they are designed correctly [3]. However, achieving a fair algorithm is challenging, even with humans in the loop, because biases often exist from the outset [23]. Tsamados et al. (2021) suggest fairness can be promoted by excluding protected characteristics such as race and gender from decision making and ensuring false positives and false negatives are equal across protected groups. However, it is not universally agreed upon that fairness in this regard is achievable, as Mitchell et al. (2021) notes, "there can be no harmony among definitions of fairness in a world where inequality and imperfect prediction are the reality"[9].

Green (2022) proposes a strategy for promoting substantive algorithmic fairness, focusing on addressing discrimination and inequality through relational and structural reforms [25]. The framework focuses on whether inequalities in outcomes are caused by or reinforce social hierarchies, like racism or sexism, or result from biased systems. It aims to reduce these inequalities by fixing the causes and improving decision-making processes. Algorithms are instead seen as tools to support these changes, not as the main solution [25].

Another significant challenge is avoiding black-box behaviour in algorithms. Yeung (2018) argues that algorithmic governance based on black-box systems is inherently undemocratic. In constitutional democracies, legitimate sovereign rule is grounded in two key principles: first, that people live under rules of their own making; and second, that these rules can be contested through transparent, adversarial procedures [3]. Brauneis and Goodman (2018) propose that governments use their contracting powers to insist that software vendors produce the following documentation to promote transparency: accurately describing the predictive goal and application, relevant and excluded data (and why), time and place limitations of data, specific predictive criteria (training data should match predicted outcomes), analytical/development techniques, validation studies and plain language explanations of the algorithm and outputs [11].

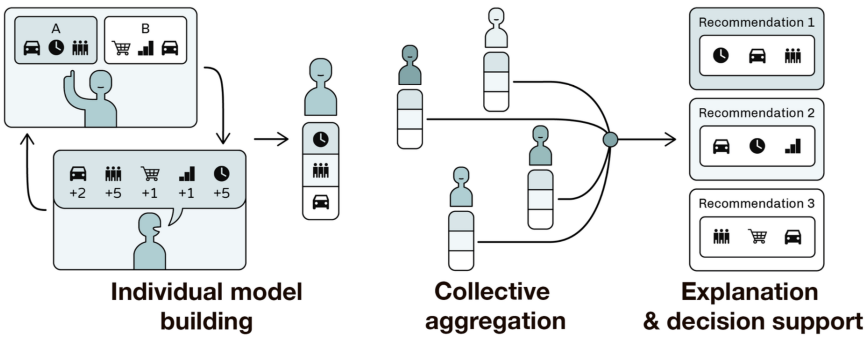


Figure 1: WeBuildAI Framework - Highlevel (taken from Lee et al. 2019)

For systems designed from scratch, Lee et al. (2019) proposes another framework,

illustrated in Figure 1. This approach emphasizes transparency by addressing three key questions [23]: What values should be embedded in the system? How should these values be prioritized, and how should fairness be measured? The implementation begins with building individual models that reflect stakeholder values, which are weighted differently. These models are then aggregated using predefined metrics, producing final recommendations. Crucially, these recommendations should always be explained to stakeholders to ensure transparency and avoid undesirable black-box behaviour [23]. By taking multiple stakeholders' values into account, this framework is promising. However, it relies heavily on the accuracy and relevance of predefined metrics, and stakeholders' values can be subjective [23].

4 Conclusion

Algorithmic governance presents challenges in assessing transparency, bias, and fairness. Transparency, a critical component, is difficult to quantify, as it requires not just access to information but comprehension of the underlying logic, decision-making significance, and consequences. Bias and fairness also pose measurement difficulties, as definitions often varying with context, leaving room for interpretation and inconsistency.

Fortunately, jurisdictions such as the European Union have implemented laws to mitigate these issues. The GDPR provides a "right not to be subject to automated decision-making" and a "right to explanation", necessitating that organizations inform individuals about the logic, significance, and consequences of such systems [26]. The AI Act complements this by introducing stricter regulations for high-risk AI systems, ensuring ethical safeguards and accountability, including a human-in-the-loop for legally significant decisions [6].

Evaluating fairness in algorithmic governance requires balancing data needs, ethical considerations, and compliance. While leveraging personal data can enhance fairness, it raises ethical and privacy concerns. Effective frameworks must address these issues by defining fairness, ensuring meaningful evaluation metrics, and implementing robust regulatory oversight.

Lastly, partnerships with private entities demand detailed contractual frameworks to prevent mistrust, emphasizing accountability and transparency. Critical evaluation of existing frameworks will drive ethical and effective algorithmic governance.

References

[1] Lawrence Lessig. *Code: And Other Laws of Cyberspace*. ReadHowYouWant.com, 2009.

[2] Natascha Just and Michael Latzer. "Governance by Algorithms: Reality Construction by Algorithmic Selection on the Internet". In: *Media, culture & society* 39.2 (2017), pp. 238–258.

[3] Karen Yeung. "Algorithmic Regulation: A Critical Interrogation". In: *Regulation & Governance* 12.4 (2018), pp. 505–523. DOI: 10.1111/rego.12158. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/rego.12158.

[4] Alexandra Jonker and Julie Rogers. *Algorithmic Bias: Understanding and Addressing Bias in AI*. 2024.

[5] Qiaochu Wang et al. "Algorithmic Transparency with Strategic Users". In: *Management Science* 69.4 (2023), pp. 2297–2317.

[6] European Commission. *AI Act, Article 14: Human Oversight*. 2024.

[7] Christian Katzenbach and Lena Ulbricht. "Algorithmic Governance". In: *Internet Policy Review* 8.4 (2019), pp. 1–18.

[8] Julia Angwin et al. *Machine Bias: Risk Assessments in Criminal Sentencing*. 2016.

[9] Shira Mitchell et al. "Algorithmic Fairness: Choices, Assumptions, and Definitions". In: *Annual Review of Statistics and Its Application* 8.Volume 8, 2021 (2021), pp. 141–163. ISSN: 2326-831X. DOI: 10.1146/annurev-statistics-042720-125902.

[10] Julia Angwin et al. "Machine Bias". In: *Ethics of Data and Analytics*. Auerbach Publications, 2022, pp. 254–264.

[11] Robert Brauneis and Ellen P Goodman. "Algorithmic Transparency for the Smart City". In: *Yale JL & Tech.* 20 (2018), p. 103.

[12] Joshua Brustein. *This Guy Trains Computers to Find Future Criminals*. 2016.

[13] Brianna Lifshitz. "Racism Is Systemic in Artificial Intelligence Systems, Too". In: *Georgetown Security Studies Review* (2021).

[14] Ellora Thadaney Israni. "When an Algorithm Helps Send You to Prison". In: *New York Times* 26 (2017).

[15] Harvard Law Review. "State v. Loomis: Wisconsin Supreme Court Requires Warning before Use of Algorithmic Risk Assessments in Sentencing". In: *Harvard Law Review* 130 (2017), pp. 1530–1537.

[16] HunchLab. *A Citizen's Guide to HunchLab*. 2017.

[17] Fei Yang. "Predictive Policing". In: *Oxford Research Encyclopedia of Criminology and Criminal Justice*. Oxford University Press Oxford, UK, 2019.

[18] Ángel Delgado-Panadero et al. "Implementing Local-Explainability in Gradient Boosting Trees: Feature Contribution". In: *Information Sciences* 589 (2022), pp. 199–212. ISSN: 0020-0255. DOI: 10.1016/j.ins.2021.12.111.

[19] Cary Coglianese and David Lehr. "Transparency and Algorithmic Governance". In: *Administrative Law Review* 71.1 (2019), pp. 1–56. ISSN: 00018368, 23269154. JSTOR: 27170531. (Visited on 11/23/2024).

[20] Kami Chavis Simmons. "Police Technology Shouldn't Replace Community Resources". In: *New York Times* (2015).

[21] Eric P Baumer. "Neighborhood Disadvantage and Police Notification by Victims of Violence". In: *Criminology : an interdisciplinary journal* 40.3 (2002), pp. 579–616.

[22] Carmen M Gutierrez and David S Kirk. "Silence Speaks: The Relationship between Immigration and the Underreporting of Crime". In: *Crime & Delinquency* 63.8 (2017), pp. 926–950.

[23] Min Kyung Lee et al. "WeBuildAI: Participatory Framework for Algorithmic Governance". In: *Proceedings of the ACM on human-computer interaction* 3.CSCW (2019), pp. 1–35.

[24] Andreas Tsamados et al. "The Ethics of Algorithms: Key Problems and Solutions". In: *Ethics, governance, and policies in artificial intelligence* (2021), pp. 97–123.

[25] Ben Green. "Escaping the Impossibility of Fairness: From Formal to Substantive Algorithmic Fairness". In: *Philosophy & Technology* 35.4 (Oct. 2022), p. 90. ISSN: 2210-5441. DOI: 10.1007/s13347-022-00584-6.

[26] Protection Regulation. "Regulation (EU) 2016/679 of the European Parliament and of the Council". In: *Regulation (eu)* 679 (2016), p. 2016.

Declaration of AI Usage:

In general the whole text was written by us based on the literature review we conducted. Some part of the text were given to ChatGPT after writing to correct the grammar and style of the text or refine it a bit more. This was achieved with the simple prompt "Please correct grammar and style mistakes: [TEXT]". Other than that there was no generative AI used. We tried using ChatGPT to find sources as a starting point, but in this case, generative AI was not very effective at identifying proper sources for our needs. In the case of finding the correct specific sources within the GDPR and the EU AI Act texts, AI use was actually useful but we made sure to cross-reference with the actual sources afterwards.

For the poster we only used AI for generating the one picture (red AI computer science person). The prompt for creating the image was: "A human silhouette filled with glowing digital patterns, emphasizing the influence of algorithmic governance on individuals". The rest of the poster was created by us.

The Ethics of Algorithmic Governance

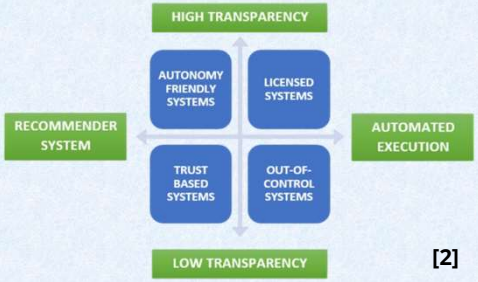
Code of Conduct? Or Just Code?

CONCEPTUAL INTRODUCTION

Algorithmic Governance refers to the use of algorithms and automated systems by governing bodies in or achieving decisions, enforce policies etc. This administrative approach has gained popularity in recent years due to significant improvements in the improved capabilities of such systems.

An **'algorithm'** , in turn, denotes " a set of commands that a computer follows in order to perform calculations or other problem-solving operations." [1] in this context, this could mean social benefit allocation systems, public health management systems etc.

In theory, algorithms offer governing bodies potential benefits, including objectivity, efficiency, and scalability. However, reaching these benefits has proven challenging.



[2]

This difficulty can be attributed to several factors, including, but not limited to;

- The reinforcement of existing societal ***biases*** in algorithmic design.
- The need for clear and logical ***transparency*** concerning decision making and results.
- Challenges in assigning ***accountability*** for algorithmic decisions as well as their outcomes.

METHODOLOGY

An exact set of steps were followed to produce this poster, it's corresponding literature review and our communication piece, a podcast script.

- First, we gathered papers in and around the area outlined by our chosen question.
- Following this, we categorized these findings into subcategories and assigned each one to a group member.
- After the studying of our respective documents, we discussed the topic further to ensure well-rounded knowledge on all aspects of the question.
- We then collaboratively created our literature review, poster and podcast script, with a separate emphasis on how to present the gathered information, depending on the medium.

REAL WORLD EXAMPLES

COMPAS, an algorithmic risk-assessment tool developed by Northpointe for use by the US Government in areas such as pretrial release and recidivism.

NORTHPOINTE suite

HunchLab, a predictive policing software utilized in the United States to forecast the timing and location of potential crimes, helping guide law enforcement deployment.

LITERATURE REVIEW

A selection of the papers we reviewed to enhance our understanding of this topic included:

- ***Fairness Choices, Assumptions and Definitions*** - (Mitchell, 2021)
- ***Algorithmic Transparency for the Smart City*** - (Brauneis & Goodman, Yale Law Journal, 2018)
- ***Algorithmic Authority: the Ethics, Politics, and Economics of Algorithms that Interpret, Decide, and Manage*** - (Lustig, Pine, 2017)
- ***Escaping the Impossibility of Fairness: From Formal to Substantive Algorithmic Fairness*** - (Green, 2022)
- ***Algorithmic regulation: A critical interrogation*** - (Yeung, 2018)

MAIN FINDINGS

- The importance of the balance between bias, transparency and control.
- The significance of a two-channel regulative approach, both technologically and lawfully.
- The decisiveness of training data and design choices.
- A focus on the stakeholder involvement and documentation.

CONCLUSION

Based on our learnings from the literature and our analysis of the topic, our main conclusion is concerned with the mandating of more robust ethical frameworks for the design and development of such Automated Decision-Making systems.

These frameworks must steer both the creation and documentation processes, ensuring that ethical considerations are integral at every stage of development.

While we recognize the challenges involved, we believe that dedicating as much attention to regulation as is given to development could significantly help address the existing issues.

REFERENCES

[1] S. Upadhyay, "What Is An Algorithm? Characteristics, Types and How to write it | Simplilearn," Simplilearn.com, Nov. 18, 2022.

[2] R. Sethi, V. Ratan Vatsa, and P. Chhaparwal, "IDENTIFICATION AND MITIGATION OF ALGORITHMIC BIAS THROUGH POLICY INSTRUMENTS," International Journal of Advanced Research, vol. 8, no. 7, pp. 1515–1522, Jul. 2020

ACKNOWLEDGEMENTS

We would like to thank our lecturer, Heike Felzmann, for providing a semester of intriguing and thought-provoking lectures.

Podcast-Script: A Critical Analysis of Ethical Problems Related to Algorithmic Governance

AI and Ethics (CT5142)

Michael Rice (20347541)

James Conroy (24230670)

Jonas Michel (24238749)

26th November 2024

Podcast Script

Mike (Host):
Hello and welcome to a new episode of "AI Today" the podcast where we unpack the ethical dilemmas shaping our algorithm-driven world. I'm Mike McChip, your host. Today, we're diving into the challenges of algorithmic governance, with specific attention given to transparency, fairness and control. Today we welcome two guests: Jaccard, an AI Ethics and Policy expert, and Jenkins, a data scientist and industry consultant specialising in algorithmic design. Thank you both for joining us.

Jaccard (Guest 1):
Thank you, Mike. It's a pleasure to be here.

Jenkins (Guest 2):
Thanks for having me. Looking forward to the discussion.

Mike (Host):
Let's start with transparency. But before we dive in, let's make sure our listeners understand the examples we'll be discussing. Jaccard, can you explain what COMPAS is?

Jaccard (Guest 1):
Of course. COMPAS stands for Correctional Offender Management Profiling for Alternative Sanctions. It's an algorithm used in the U.S. justice system to predict the likelihood of someone reoffending, aiming to assist judges in making decisions about sentencing, bail, or parole. However, its methodology is proprietary, meaning its internal calculations are hidden. This has sparked concerns about its fairness and accuracy. For instance, studies have shown that COMPAS disproportionately labels Black defendants as high-risk compared to white defendants, even when their likelihood of reoffending is similar. This lack of transparency makes it difficult to identify or address biases in the system.

Mike (Host):
Jaccard, you've been vocal about the need for greater transparency in algorithms like COMPAS. Why is this so critical?

Jaccard (Guest 1):
Transparency is essential because it's the foundation of accountability. With COMPAS Judges rely on its recommendations without understanding how those decisions are made. This creates significant risks, especially when biases go unchecked. If we want fair systems, we need clear explanations of how they work. Transparency allows us to scrutinize, challenge, and also to improve these algorithms. It's about ensuring that the decisions they make are justifiable and ethical.

Mike (Host):
Jenkins, as a data scientist, how do you balance the need for transparency with the protection of intellectual property and trade secrets?

Jenkins (Guest 2):
It's a delicate balance. While companies have a legitimate interest in protecting their proprietary algorithms, they also have a responsibility to ensure that their systems are fair, accountable and transparent. But complete transparency isn't always practical. There are other tools like HunchLab, where revealing too much would be against the public interest.

Mike (Host):
Thanks for mentioning HunchLab here. Could you to explain our listeners what HunchLab is and how to evaluate transparency in this specific case?

Jenkins (Guest 2):
Yes of course. HunchLab is a predictive policing tool that analyzes data like past crime reports, socioeconomic factors, and even weather patterns to predict where crimes are likely to occur. For example, if the algorithm identifies a specific area as a potential crime hotspot, it might direct more police patrols there. Police departments use this information to allocate resources and plan patrols. A certain degree of opacity is necessary to maintain the effectiveness of these predictive policing tools. For example, revealing too much about the algorithm's decision-making process could allow bad actors, such as criminals, to exploit patterns in patrol assignments.

Mike (Host):
Jaccard what do you think about this? How can we ensure that algorithms like HunchLab are transparent enough to be accountable while also protecting sensitive information? Can transparency and security coexist?

Jaccard (Guest 1):
It's a challenging question. Transparency and security are often seen as opposing forces, but they don't have to be. We can achieve a balance by implementing measures like independent audits, explainability tools, and data protection protocols. For example, HunchLab could provide high-level explanations of its predictions without revealing the specifics of its algorithms. This way, the public can understand the rationale behind the decisions without compromising the system's security. It's about finding the right level of transparency that ensures accountability without jeopardizing the effectiveness of the algorithm. In fact companies should be held accountable by law to ensure a certain level of transparency.

Mike (Host):
Interesting. Let's shift our focus to fairness. Jenkins, you've worked on algorithmic design in various industries. How do you define fairness in the context of algorithms, and what challenges do you face in achieving it?

Jenkins (Guest 2):
Fairness is about ensuring algorithms treat everyone equitably, but that's easier said than done. One approach is to exclude sensitive variables like race or gender, but even then, biases can creep in through proxy variables like zip codes. It could be

also possible to get a gender bias through the data, by some given correlations, even if we drop gender related data. So, it's a complex challenge. Fairness needs to be context-specific, and perfect fairness might not even be achievable. The algorithms are only as good as the data they're trained on, and historical biases can be deeply ingrained in that data.

Mike (Host):
So you are saying that even if we try to exclude sensitive variables, biases can still be present in the data. Jaccard, do you see it the same way?

Jaccard (Guest 1):
Not entirely. I believe fairness goes beyond individual treatment to addressing systemic inequalities. Algorithms like HunchLab, for example, rely on crime related data that's often biased against disadvantaged communities. If we're not addressing these structural issues, fairness remains a pipe dream.

Mike (Host):
That is a good point. You are bringing the structural bias into the discussion. Jenkins, how can we mitigate these biases in algorithmic systems, especially when the data itself is biased?

Jenkins (Guest 2):
It's a tough nut to crack. First of all if we as a society are biased towards certain groups, it's hard to blame developers for not handling this issue. This is a societal problem. However, there are ways, one is through iterative testing and feedback loops. We can monitor how algorithms perform in real-world settings and adjust them to minimize unintended consequences. For that we need to have a diverse team of developers and ethicists to identify and address potential biases.

Mike (Host):
Jaccard, you were talking about the need for governance mechanisms to ensure fairness. The EU's AI Act emphasizes human oversight for high-risk systems. What does meaningful oversight look like to you?

Jaccard (Guest 1):
I think meaningful oversight has to be proactive and informed. It's not enough to have humans rubber-stamp algorithmic decisions. Decision-makers need to understand how the systems work and have the authority to override them when necessary. Otherwise, it's just an illusion of control. Oversight should also involve regular audits, transparency reports, and mechanisms for redress when things go wrong. It's about holding developers and users accountable for the impact of these systems.

Mike (Host):
Jenkins, does that align with your experience in the industry? Is that feasible in practice?

Jenkins (Guest 2):
First of all I totally agree with Jaccard here. It's a good idea to have a human oversight. However, in practice, it's not always easy to implement. Companies are often focused on efficiency and profit, and adding layers of oversight can slow down decision-making processes. But the risks of not having oversight are too great to ignore. We need to find ways to balance accountability with innovation. It's a challenge, but it's a necessary one. There also might arrise a trade-off between micromanaging every decision and high-level validation and governance.

Mike (Host):
Let's wrap up by discussing potential solutions. Are there any best practices or emerging technologies that give you hope for the future of algorithmic governance?

Jaccard (Guest 1):
I'm optimistic about the growing interest in explainable AI and fairness-aware algorithms. There are frameworks aiming to make algorithms more interpretable and accountable. These tools can help us understand how decisions are made and identify biases before they cause harm. I'm also encouraged by the push for stronger regulations like the EU's AI Act. By setting clear standards and requirements, we can ensure that algorithms are developed and used responsibly. But ultimately, it's up to all of us to demand accountability and transparency in algorithmic systems.

Mike (Host):
Jenkins, do you have any final thoughts on this?

Jenkins (Guest 2):
I agree with Jaccard. The future of algorithmic governance depends on collaboration between policymakers, technologists, and the public. We need to work together to create systems that are fair, transparent, and accountable. It's a complex challenge, but it's one we can't afford to ignore. Speaking for the developers out there I can say that we are aware of the responsibility, but we also need the support in form of relevant frameworks.

Mike (Host):
Thank you both for such an insightful discussion. Transparency, fairness, and control are clearly interwoven challenges, but conversations like these bring us closer to ethical solutions. And thank you to our listeners for tuning in to "AI Today". Join us next time as we explore another pressing issue at the intersection of technology and society. Until then, stay informed and as always question the code.

Explanation of Focus and Examples:

Our first focus was the differing degrees of transparency that are sufficient in algorithmic governance, depending on the context. Our two case studies allowed us to illustrate how opaqueness can sometimes be necessary for algorithms to be safe (predictive policing), but when significant legal decisions affecting someone's life are being made, transparency is paramount. Our second main focus was fairness, illustrated by the two guests giving their opinions on what is defined by fairness in algorithms. One opinion was more technical in nature, whilst the other provided a more rounded idea about how fairness cannot be achieved if the data used to train algorithms is just re-enforcing systemic inequalities already existing in society. We assume the audience has little existing knowledge about algorithmic governance, so we explained related concepts where appropriate to help expand their knowledge.

Development Process and AI Usage:

The podcast's content was developed using our literature review as inspiration. Using this in-depth knowledge of transparency, fairness and control, we condensed these topics down into an accessible format for a general audience, using the two case studies for illustration. We also highlighted possible solutions and frameworks to help combat the issues highlighted in the case studies. We decided on three roles: one host and two guests. Both guests are very knowledgeable, with one being more technical and the other having a bias to ethical considerations. This provided the audience with contrasting viewpoints, with the host moderating and seeking clarity where required. We wrote the script and then used AI to help refine it, requesting that content is more accessible to a general audience. We then did the final edits to ensure that the correct tone was aligned with the podcast's purpose.

Limitations and Further Improvements:

The podcast content is light on explaining proposed frameworks and policies that can help address the issues raised in algorithmic governance. Potentially, a third guest who is a policymaker could help elaborate on this topic. The case studies are also quite US-centric, so there could be scope for using examples that are more varied geographically.