The Algorithmic Transparency Playbook: A Stakeholder-first Approach to Creating Transparency for Your Organization's Algorithms

Andrew Bell
alb9742@nyu.edu
New York University
New York, United States

Oded Nov
New York University
New York, United States
onov@nyu.edu

Julia Stoyanovich
New York University
New York, United States
stoyanovich@nyu.edu

ABSTRACT

Welcome to 2033, the year when AI, while not yet sentient, can finally be considered responsible. Only systems that work well, improve efficiency, are fair, law abiding, and transparent are in use today. It's AI nirvana. You ask yourself: "How did we get here?" You may have played a major role! As more organizations use algorithmic systems, there is a need for practitioners, industry leaders, managers, and executives to take part in making AI responsible. In this course, we provide for influencing positive change and implementing algorithmic transparency into your organization's algorithmic systems.

CCS CONCEPTS

• **Human-centered computing**; • **Social and professional topics** → **Socio-technical systems**; *Computing / technology policy*;

KEYWORDS

course, fundamentals of HCI, algorithmic transparency, responsible AI

ACM Reference Format:
Andrew Bell, Oded Nov, and Julia Stoyanovich. 2023. The Algorithmic Transparency Playbook: A Stakeholder-first Approach to Creating Transparency for Your Organization's Algorithms. In *CHI Conference on Human Factors in Computing Systems (CHI '23), April 23-28, 2023, Hamburg, Germany*. ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3544549.3574169

1 BENEFITS

The primary learning objective of this course it to teach participants **how transparency can be implemented for algorithmic decision systems within an organization**. This course is intended for beginners in algorithmic transparency and algorithmic decision-making. It consists of four modules, in which participants will learn how to:

(1) **Inventory** and identify the algorithmic systems within their organization that require transparency.
(2) **Plan & Design** transparency for their algorithmic systems.
(3) **Implement** transparency for their algorithmic systems.
(4) **Maintain** algorithmic transparency into the future.

There are seven learning outcomes for this course. By the time participants complete the course, they will be able to answer the following questions about algorithmic transparency:

(1) What is algorithmic transparency and how is it defined?
(2) Who are the stakeholders of algorithmic transparency?
(3) What are the goals of algorithmic transparency?
(4) What are the existing methods for algorithmic transparency?
(5) What are the best practices for designing and implementing algorithmic transparency into existing and future algorithmic systems?
(6) How can algorithmic transparency be maintained into the future?
(7) How can I help influence a culture shift in my organization towards adopting algorithmic transparency?

2 INTENDED AUDIENCE(S)

This course is intended for the following audience members:

- **Industry leaders, C-suite executives, managers, and practitioners.** In organizations using algorithmic decision-making systems, it's critical that leadership understands how important algorithmic transparency is, from both a business and an ethical perspective. By taking this course, executives and managers will learn what algorithmic transparency is, why it is useful for different stakeholders, and how it can be implemented within their organization.
- **Auditors, policymakers, and regulators.** In recent years, governments around the world have begun to regulate algorithmic systems, often requiring some degree of transparency. Unfortunately, all existing and emerging laws and regulations on algorithmic transparency share a common weakness: they focus on *what* to do (or what not to do), but leave the brunt of the work to data scientists to figure out *how* [3]. In part, this is due to an information asymmetry between those building algorithmic systems, and those creating policy. By taking this course, policymakers and regulators will bridge this knowledge gap and come to better understand concrete ideas about algorithmic transparency, which will ergo improve their understanding of how to oversee algorithmic systems.

Furthermore, while they are not the primary audience, **data scientists and data/software engineers will also enjoy this course.** They will learn about state-of-the-art and industry-standard practices for algorithmic transparency, become familiar with new

CHI '23, April 29-May 5, 2023, Hamburg, Germany

ACM ISBN 978-1-4503-9422-2/23/04.
https://doi.org/10.1145/3544549.3574169

transparency methods, and gain an understanding of how to implement transparency effectively in their organization. We include a technical appendix to this course for this audience.

3 PREREQUISITES

There are no prerequisites for this course. It is designed for **absolute beginners** in algorithmic decision-making and algorithmic transparency. The course material is designed for an interested *non-technical* audience, and will be appealing to anyone learning what algorithmic transparency is and how it can be incorporated into their organization's algorithmic governance.

4 CONTENT

This course will be taught in a single 75-minute sessions. The content is divided into an introduction to algorithmic transparency (Part 1, Section 4.1) and hands-on experience with case studies (Part 2, Section 4.2), described below.

4.1 Part 1: All About Algorithmic Transparency (45 mins)

The core material of this course comes from the Algorithmic Transparency Playbook, which was written by the authors of this course submission and will be distributed by the Center for Responsible AI at NYU. All course materials are open-access and open-source. Part 1 of this course is divided into two modules, described next.

4.1.1 Module 1: Background on Algorithmic Transparency (30 mins). The course begins with a level-setting activity, further subdivided into 4 sub-modules:

(1) **The *why, what* and *who* of algorithmic transparency**, as discussed in the Abstract, Benefits, and Intended Audience(s) sections of this document.
(2) **Defining algorithmic transparency**, and a nuanced discussion of what it means for an algorithm to be "transparent." We also discuss related definitions like "explinability" and "interpretability."
(3) **The fundamentals of algorithmic transparency**, presented in the style of a *frequently asked questions (FAQ)* sequence. We answer the questions "Why should one care about algorithmic transparency?" and "Who is algorithmic transparency for?" We also include a discussion on what algorithmic transparency *is not.*
(4) **Organizational barriers to algorithmic transparency**. We present five fundamental barriers to transparency in both the public and the private sectors, and discuss how they can be overcome. For example, organizations often believe that using more transparent algorithms as opposed to complex "black-box" systems makes them less accurate—but this is simply not true [2]! Other barriers include risks to an organization's intellectual property or proprietary data, concerns over privacy, risks related to the strategic manipulation (or gaming) of their systems, and the costs of transparency. This section is important for audience members who want to influence a culture shift within their organization towards algorithmic transparency.

This section will be accompanied by an interactive exercise: *before* diving into the course material, we will invite audience members to write down their own questions about algorithmic transparency and identify barriers to its implementation. After covering the content of this introductory module, we will have a 5-10 minute discussion about any unanswered questions. It is also our intention to use this audience participation to grow our materials for future iterations of this course.

4.1.2 Module 2: The Algorithmic Transparency Process (15 mins). This section is accompanied by the Algorithmic Transparency Checklist, which will be distributed to all the participants at the beginning of the course. The checklist is a distillation of the content of this course into actionable steps that participants can take with them. By following the checklist, participants can ensure that they have properly created transparency for algorithms in their organization.

Like with Module 1 (Section 4.1.1), the material of this module is contained in 4 sub-modules:

(1) **Inventory**. Participants will learn how to inventory their organization for algorithmic decision-making systems of all types, and understand which algorithms need transparency. We also discuss how to assess the relative risk of each system as a way to triage the importance of implementing transparency for them.
(2) **Plan & Design**. In this central section of the course, we discuss the stakeholders of an algorithmic system, their goals, and ideas for how these goals can be met in practice. The 5 main stakeholder categories are practitioners, managers, affected persons, humans-in-the-loop, and compliance officers. The 6 possible goals are validity, trust, learning and support, recourse, fairness, and privacy. Participants will be presented with several examples of these stakeholders and their goals for different systems.
(3) **Implement**. Participants will learn-by-example how to implement what was learned in the Plan & Design section into practice. We will show several examples of transparency tools like "system cards," "nutritional labels," visualizations, and dashboards. This section also defines *participatory design* and explains how it can be used to ensure that transparency tools are designed effectively with input from a diversity of stakeholders.
(4) **Maintain**. Participants will learn why and how algorithmic transparency should be maintained into the future.

4.2 Part 2: Case-Studies and Hand-on Exercises (30 mins)

In the second part of this course, we focus entirely on case-studies and practical work for the material learned in Part 1. Again, this section will be divided into 2 modules. In the first (Module), we will present 2 case studies, one as a positive (strong) example of transparency, and one as a negative (weak) example. In the second part (Module 4), we will ask participants to work through 2 additional case studies, giving them an opportunity to see how transparency concepts translate across domains. All case studies will be based on real-world work done by the course instructors.

4.2.1 Module 3: Review Case-Studies (10 minutes). Participants will go through one case-studies included in the course content. The case-study is a **strong** example of how algorithmic transparency can be implemented by a government agency to help curb unemployment. There is a second case-study accompanying the course material that is a fictional example of a **weak** implementation of transparency by a government agency that uses an algorithm to prioritize restaurants for food inspection, that participants will be able to review on their own outside of course time.

4.2.2 Module 4: Case-Studies Game (20 minutes). We will divide the audience up into groups of 5, where participants will be randomly assigned a stakeholder to role-play in one of two scenarios: algorithmic hiring and healthcare. Each group will go through the Algorithmic Transparency Checklist, and practice designing transparency for the system that is responsive to the needs of each stakeholder. If time permits, we will convene for a group discussion of what was learned and what challenges were faced. We will also provide "solutions" to the scenarios that participants will be able to review on their own outside of course time.

5 PRACTICAL WORK

The practical work component of this course is described in Section 4.2.

6 INSTRUCTOR BACKGROUND

The course will be co-taught by Andrew Bell, Julia Stoyanovich, and Oded Nov.

Andrew Bell is a fully-qualified Ph.D. candidate in Computer Science at New York University (NYU). He is co-advised by Julia Stoyanovich and Oded Nov, and is funded by the US National Science Foundation through a Graduate Research Fellowship (NSF GRFP). Andrew has published multiple academic articles related to algorithmic transparency and explainability. Most recently, he was first author on a paper titled *Think About the Stakeholders First! Towards an Algorithmic Transparency Playbook for Regulatory Compliance*, which was accepted into the *Data & Policy Journal* (publication forthcoming). Additional first-author work includes the paper *It's Just Not That Simple: An Empirical Study of the Accuracy-Explainability Trade-off in Machine Learning for Public Policy*, presented at the ACM Fairness, Accountability, and Transparency (FAccT) conference in Seoul, South Korea in 2022 [2], and *Proactive Advising: A Machine Learning Driven Approach to Vaccine Hesitancy*, presented at the International Conference on Healthcare Informatics (ICHI) in 2018 [1]. Other publications related to transparency and explainability include *Machine Learning Informed Decision-Making with Interpreted Model's Outputs: A Field Intervention*, published in the Academy of Management Proceedings [5], and *Algorithmic Long-Term Unemployment Risk Assessment in Use: Counselors' Perceptions and Use Practices* published in the journal Global Perspective [4].

Oded Nov is Professor of Technology Management at New York University's Tandon School of Engineering. His research interests include human-computer interaction, social computing, digital health and the future of work. In collaboration with researchers and practitioners in the medical and business schools at NYU, Oded leads a research program on digital health work, with a focus on the role of AI-based and data-intensive technologies in healthcare work. Oded received his Ph.D. in management information systems at Cambridge University, and he is a recipient of the NSF CAREER award.

Julia Stoyanovich is Institute Associate Professor of Computer Science and Engineering, Associate Professor of Data Science, and Director of the Center for Responsible AI at New York University. Julia's goal is to make "Responsible AI" synonymous with "AI." She works towards this goal by engaging in academic research, education and technology policy, and by speaking about the benefits and harms of AI to practitioners and members of the public. Julia's research interests include AI ethics and legal compliance, data management and AI systems, and computational social choice. She developed and teaches technical courses on responsible data science, co-developed an online course AI Ethics: Global Perspectives, is a co-creator of two comic book series on responsible AI, and a co-designer of a public education course We are AI: Taking Control of Technology. Julia is engaged in technology policy in the US and internationally, having served on the New York City Automated Decision Systems Task Force, by mayoral appointment, among other roles. She received M.S. and Ph.D. degrees in Computer Science from Columbia University, and a B.S. in Computer Science and in Mathematics and Statistics from the University of Massachusetts at Amherst. Julia is a recipient of an NSF CAREER award and a Senior Member of the ACM.

7 RESOURCES

- The full course is open access and open source.
- The Algorithmic Transparency Playbook is a report by the Center of Responsible AI on which this course is based. We plan to have copies of this report for all audiences members to take with them.
- The Algorithmic Transparency Checklist is an actionable distillation of the course content.
- A Python notebook that illustrates how transparency can be implemented, for technical audience members or for managers to present to their technical team.

8 ACCESSIBILITY

All materials that make part of the course will be made available to participants in several formats. These include the course website, designed using Course-in-a-Box, and open-source framework that supports access using different devices, and both printed and a PDF-A versions of the Algorithmic Transparency Playbook and Checklist, tested for screen reader accessibility to meet the needs of the Blind and Low Vision community. In designing the course website, and other electronic and printed materials will use color schemes and fonts that are designed to improve accessibility across a range of visual and learning abilities. We will be grateful to make use of sign language interpreter services provided by the conference.

ACKNOWLEDGMENTS

This research was supported in part by NSF Awards No. 1934464, 1922658, 1916505, 1928614, and 2129076.

REFERENCES

[1] Andrew Bell, Alexander Rich, Melisande Teng, Tin Orešković, Nuno B Bras, Lénia Mestrinho, Srdan Golubovic, Ivan Pristas, and Leid Zejnilovic. 2019. Proactive advising: a machine learning driven approach to vaccine hesitancy. In *2019 IEEE International Conference on Healthcare Informatics (ICHI)*. IEEE, 1–6.

[2] Andrew Bell, Ian Solano-Kamaiko, Oded Nov, and Julia Stoyanovich. 2022. It's Just Not That Simple: An Empirical Study of the Accuracy-Explainability Trade-off in Machine Learning for Public Policy. In *2022 ACM Conference on Fairness, Accountability, and Transparency*. 248–266.

[3] Anna Jobin, Marcello Ienca, and Effy Vayena. 2019. The global landscape of AI ethics guidelines. *Nature Machine Intelligence* 1, 9 (2019), 389–399.

[4] Leid Zejnilović, Susana Lavado, Íñigo Martínez de Rituerto de Troya, Samantha Sim, and Andrew Bell. 2020. Algorithmic Long-Term Unemployment Risk Assessment in Use: Counselors' Perceptions and Use Practices. *Global Perspectives* 1, 1 (2020).

[5] Leid Zejnilovic, Susana Lavado, Carlos Soares, Íñigo Martínez De Rituerto De Troya, Andrew Bell, and Rayid Ghani. 2021. Machine Learning Informed Decision-Making with Interpreted Model's Outputs: A Field Intervention. In *Academy of Management Proceedings*, Vol. 2021. Academy of Management Briarcliff Manor, NY 10510, 15424.