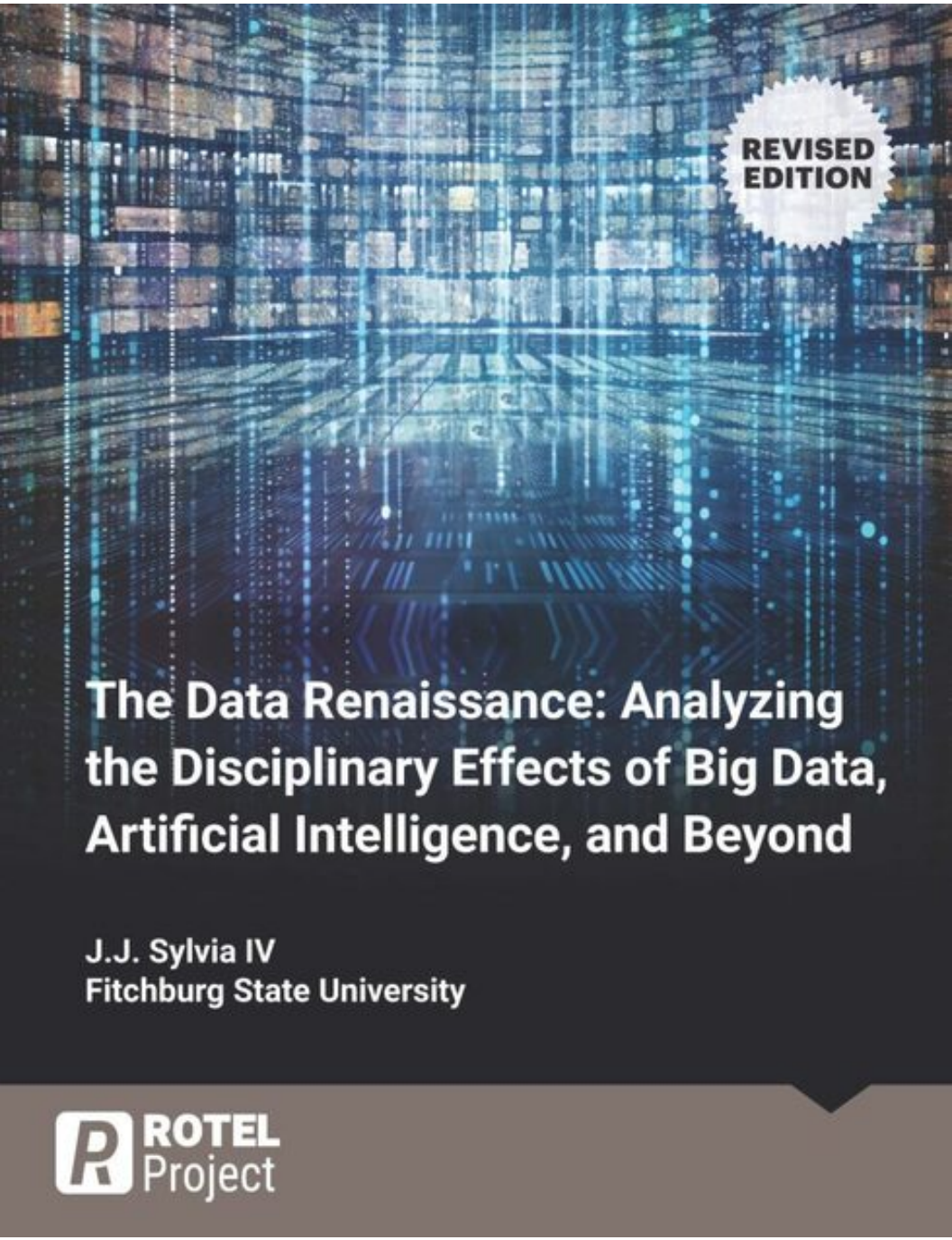


The Data Renaissance: Analyzing the Disciplinary Effects of Big Data, Artificial Intelligence, and Beyond

J.J. Sylvia IV
Fitchburg State University

ROTEL Project

The Data Renaissance: Analyzing the Disciplinary Effects of Big Data, Artificial Intelligence, and Beyond [Revised Edition]

THE DATA RENAISSANCE: ANALYZING THE DISCIPLINARY EFFECTS OF BIG DATA, ARTIFICIAL INTELLIGENCE, AND BEYOND [REVISED EDITION]

J.J. SYLVIA IV

ABOUBACAR CAMARA; ANA'AYA MCGOWAN MOZELL; BRENDAN SMITH; ELISE TAKEHANA; GLENNDALE BARTOLOME; HENRY CHRISTIANSEN; J.J. SYLVIA IV;

KESHAUNI JOHNSON; LEONORA SHELL; AND SOPHIA MOORE

ROTEL (Remixing Open Textbooks with an Equity Lens) Project

Fitchburg, Massachusetts

ISBN 978-1-964276-49-6 (Print)
ISBN 978-1-964276-48-9 (Ebook)

CONTENTS

Part III. Case Study: "It's Perfect, Four Stars!"

Part IV. Media and Data Literacy

Part V. The American Motion Picture Industry and Big Data

Part XII. SOPHIA Discussion Guides

LAND ACKNOWLEDGEMENT STATEMENT FOR THE ROTEL PROJECT

As part of ROTEL Project's mission to support the creation, management, and dissemination of culturally-relevant textbooks, we must acknowledge Indigenous Peoples as the traditional stewards of the land, and the enduring relationship that exists between them and their traditional territories. We acknowledge that the boundaries that created Massachusetts were arbitrary and a product of the settlers. We honor the land on which the Higher Education Institutions of the Commonwealth of Massachusetts are sited as the traditional territory of tribal nations. We acknowledge the painful history of genocide and forced removal from their territory, and other atrocities connected with colonization. We honor and respect the many diverse indigenous people connected to this land on which we gather, and our acknowledgement is one action we can take to correct the stories and practices that erase Indigenous People's history and culture.

Identified Tribes and/or Nations of Massachusetts

Historical Nations

- Mahican
- Mashpee
- Massachuset
- Nauset
- Nipmuc
- Pennacook
- Pocomtuc
- Stockbridge
- Wampanoag

Present-Day Nations and Tribes

- Mashpee Wampanoag Tribe
- Wampanoag Tribe of Gay Head Aquinnah
- Herring Pond Wampanoag Tribe
- Assawompsett-Nemasket Band of Wampanoags
- Pocasset Wampanoag of the Pokanoket Nation
- Pacasset Wampanoag Tribe
- Seaconke Wampanoag Tribe
- Chappaquiddick Tribe of the Wampanoag Indian Nation
- Nipmuc Nation (Bands include the Hassanamisco, Natick)
- Nipmuck Tribal Council of Chaubunagungamaug
- Massachusett Tribe at Ponkapoag

At the time of publication, the links above were all active.

Suggested Readings

Massachusetts Center for Native American Awareness

A guide to Indigenous land acknowledgment

'We are all on Native Land: A conversation about Land Acknowledgements' (YouTube video)

Native-Land.ca | Our home on native land (mapping of native lands)

Beyond territorial acknowledgments – âpihtawikosisân

Your Territorial Acknowledgment Is Not Enough

This land acknowledgement was based on the land acknowledgement of the Digital Commonwealth.

HOW TO USE THIS BOOK

This book embarks on a crucial exploration into how data is wielded across different disciplines, a landscape that is increasingly shaping our modern world. It's essential to note, however, this inquiry is fraught with challenges, particularly because many businesses and organizations guard their data practices as proprietary trade secrets. For instance, platforms like TikTok deliberately shroud their algorithms in mystery, as much of their success hinges on the perception that they've mastered the secret sauce of user engagement.

Our student contributors have diligently navigated these barriers, piecing together an overview of how data is impacting various fields. While we aimed to incorporate DEI perspectives in every chapter, we often found ourselves stymied by the opaque nature of industry practices. The veil of trade secrets not only limits what we can definitively say about data practices in these sectors but also complicates efforts to evaluate these practices from a DEI standpoint.

Despite these challenges, this book fills an essential gap in the current literature. It offers an entry point into the complex interplay of data and industry, providing foundational insights that can spur further inquiry and discussion. However, given

the constraints and limitations, this book should not serve as a standalone course text.

For a more rounded educational experience, we recommend supplementing this book with additional resources that focus explicitly on DEI issues. To help in this regard, the appendix includes several Creative Commons licensed readings. We also provide a suggested readings list, carefully curated to complement the perspectives explored in this text and to broaden the DEI lens through which these issues can be examined.

This book is intended to be a living and continually updated document each time the class is taught, rather than a final product. As such, I'd like to briefly explain how it is laid out and how others might use it for their own courses.

Introduction: The introduction is intended to give an historical overview of the project as well as the class in which it was written. I believe transparency is important and the layout of the class may be helpful to other instructors who are teaching a course and want to adopt this textbook. Because of this, it may be of less interest to students themselves.

Chapter 1: Why Care about Data & Society? When I teach this class, it's important to me that I connect with students about what the major issues are related to data and society, and also why I personally care about them. This chapter attempts to bring those approaches together in writing. This approach is not meant to simply share my own accolades. As I discuss in this chapter, I've found that

discussing data in the abstract can sometimes cause students to tune out, or make it difficult to connect to the subject. By giving an overview of data through the lens of my own career and experiences, I hope students will be able to see my passion and better understand, in a concrete way, why this topic is important.

Chapter 2 Generative AI in the Classroom and Workplace: This chapter includes a lesson with activities and guiding questions that can be used in class to teach about Large Language Models and other generative AI programs, like ChatGPT. It also features important ethical questions and considerations.

Chapter 3: Case Study: "It's Perfect, Four Stars!": This is a case study written from a first-person perspective by a business professional about their experiences related to data and society. While this is currently the only case in the text, I plan to add more of these in the future, sprinkled throughout the text. This particular case study highlights the voice and experience of a woman business owner.

Chapters 4-10: These chapters were written by students on subjects of their own choosing related to their future career interests. Students in future classes will be encouraged to either expand on existing chapters or write new chapters about career paths that aren't currently represented. Instructors may want to assign only relevant chapters to their students, or allow their students to contribute to the text as I do.

Chapter 11: GenAI hype surrounds us on a daily basis,

but so does substantial fear and anxiety. Many worry that such tools will continue to erode critical thinking skills, or remove something that is essentially "human" from the creative process. Others believe that because GenAI tools are trained on the writing and artwork of humans, all use of such tools is a form of intellectual and creative theft. Will the technology continue to improve and eventually achieve sentience? This chapter aims to give an overview of some of these major issues while also demonstrating how to use a variety of GenAI-based tools that might increase the productivity and creativity of professionals.

Chapter 12: SOPHIA Discussion Guides: This chapter features a series of discussion guides that were created by groups of students for public discussions, as part of a partnership with the Society of Philosophers in America. Students in this class could choose to use and modify them to hold their own public discussions, or they could be adapted for in-class discussions.

Chapters 13-14: These two chapters are available via a Creative Commons license and can be assigned to help better fill some of the gaps related to diversity, equity, and inclusion that occur in the above chapters focused on data in specific disciplines. They are written by leading scholars in the field.

Chapter 15: This chapter offers additional recommended reading and viewing, organized by topic. Many of them can be assigned for reading under Fair Use laws or are publicly

available multimedia content such as TV shows, podcasts, and documentaries.

PROJECT RATIONALE

I had the privilege of studying with notable posthuman philosophers Rosi Braidotti and Kate Hayles, whose teachings have been instrumental in shaping my worldview. I am a cis-gendered, heterosexual white man with the advantage of tenure at a New England university. I'm acutely aware that these aspects of my identity come with certain privileges and biases.

My education and mentorship in the field of posthumanism—a philosophy that delves into the complex interplay between technology, identity, and the human experience—inform much of my professional and personal life. In simpler terms, my work is driven by a fascination with how humans and technology interact, particularly in shaping our identities. A commitment to feminist ethics provides another critical layer to my approach, enriching my teaching, research, and overall professional practice.

The concept of Becoming is core to my intellectual endeavors. In layman's terms, Becoming means embracing continual change and acknowledging the complex network of connections that make up our lives. This philosophy encourages us to move beyond societal labels or predefined categories. I realize that my identity and perspective are in

constant flux, shaped by a myriad of interactions and experiences.

This evolving journey—this process of Becoming—shapes my goal of fostering a learning environment that is inclusive, empathetic, and encourages critical engagement.

I am committed to a continual process of learning and unlearning. The lens through which I see the world isn't fixed; it's fluid, continually molded by an ever-changing landscape of ideas and experiences . What you read here is merely a snapshot, a temporary capture of my current understanding, always subject to future transformation.

I decided to create this book because there is very little existing work that explores how data impacts different disciplines, and almost none written with an undergraduate audience in mind. This stands in stark contrast to the vast resources dedicated to the impacts of data across society more broadly.

As I was developing a new course on the topic of Data & Society that would also meet general education requirements for my campus, I wanted students to understand how data is impacting and re-shaping the specific career fields in which they will be working. One of the things that I've discovered about teaching on topics of data and privacy over the years is that it is very important to make the issues personal for students, as this helps drive a connection and interest with the issues being discussed. One of the ways I wanted to do that in this course was to allow students to better understand

how data is being used in their future career fields, while also understanding the ethical challenges associated with those uses.

Unfortunately, there were no resources like this already available, not even in a way that could be cobbled together from multiple different sources. This inspired me to apply for a ROTEL grant that would allow me to collaborate with my students to create this resource as part of our learning experience together. By having students select their own lens through which to write about data, they were able to differentiate their learning and make it personal. One student, in an informal reflection, noted that they enjoyed that they "were allowed to look into what we were interested in through the book chapter."

It was important to me to make this resource available as an OER because it can address an important void in the resources available as classes like this become more common. Further, I believe it's valuable that this book continues to be updated and expanded. My hope is that students in my future classes, as well as other classes that adopt this resource, can see it as an ongoing project that will happily accept the addition of new chapters and updates to those already existing as the field of data studies continues to expand. Ideally, as it continues to grow, students from an even wider variety of fields will be able to find their disciplines represented in these pages and instructors can choose the most relevant sections to assign in their own courses, perhaps in combination with other articles,

podcasts, and case studies. For that reason, I've included an additional suggested reading/viewing/listening list at the end of the text that can provide additional content options.

INTRODUCTION

The Data & Society Class: Process and Collaboration

Course Objectives and Structure

This book was created as part of a Data & Society course taught by Dr. J.J. Sylvia IV at Fitchburg State University in Spring 2023. The course was developed as part of a new interdisciplinary major in Digital Media Innovation. Although the major is hosted in the Communications Media department, its classes span nine different disciplines across campus. The course also has general education designations for Civic Learning and Ethical Reasoning. Because the course is open to majors from across campus, it was also tailored to allow students to explore how data is impacting careers and fields related to their own majors and future plans. The course description is as follows:

> This class explores the uses of data in Communications Media, including tailoring professional communication advertising campaigns, green-lighting film productions, creating profitable micro-transaction mechanisms in video

> games, and more. How is data leveraged to form arguments about society, make decisions, and generate profits? Through hands-on projects, students will analyze ethical challenges related to data visualization, algorithms, privacy, citizen and employee surveillance, and more.

The Data & Society class aims to provide students with a comprehensive understanding of the role and impact of data in various industries and sectors, while emphasizing the ethical, social, and cultural implications of data-driven technologies. The course is structured to encourage collaboration, active engagement, and critical thinking, combining lectures, readings, discussions, and hands-on activities to facilitate a dynamic learning experience.

Perhaps most importantly for the current project, the first iteration of this course was designed around the implementation of a Remixing Open Textbooks through an Equity Lens (ROTEL) grant, which supports faculty in their creation of new open education textbooks for academic courses. My approach to this grant was to bring students into the writing process as part of the course requirements. To do this, I assigned core readings addressing pressing issues in the field of data that were front-loaded toward the beginning of the semester. I then invited students to select topics related to their major, planned career, and/or interests. Students worked on this topic throughout the semester by selecting readings for their classmates on the topic, leading a class presentation

session on the topic, and drafting and revising their chapter for this text multiple times.

The ROTEL grant also facilitated easy access to a wealth of support across campus, including staff members who were able to visit class and offer support to students throughout the process. I'd like to take this opportunity to offer special thanks to Reneé Fratantonio (Head of Instruction and Information Literacy at Fitchburg State Library), Marilyn Billings (Faculty Advisor & Consultant for a Dept. of Education grant with the MA Dept of Higher Education and Framingham State University), Rachel Graddy (Director of Disability Services at Fitchburg State University), and Meagan Martin (Instructional Designer at Fitchburg State University).

The overall goal for this project is to create a first draft of a text that can continue to be revised and extended by the larger academic community. This will be especially important for a field such as this one, where changes to data practices happen quickly. For example, students in future classes may elect to write new chapters, or update and extend existing chapters based on their interests.

One major limitation of this approach which should be noted is that, especially in a small class such as this one, student interests may align closely. Nearly half of the students who contributed chapters to this first collection were involved in the Game Design major in the Communications Media department. For that reason, we all worked together closely to make sure they addressed different ways that data practices

are used broadly within the gaming industry. Ultimately, this means the first iteration of this book is a bit more limited in scope. Nonetheless, I ultimately believe that having students work on a project that they care deeply about is pedagogically more valuable than creating a more topically diverse first draft of this volume.

Adult Learning Connection

One additional element of this grant was to extend opportunities for participation in the project to our local community of adult learners. I did this by teaching a similar course for our Adult Learners of the Fitchburg Area (ALFA) program and offering those students multiple avenues of participation. These courses differ significantly from traditional undergraduate courses in that they do not include grades or traditional assignments, though reading lists are a common element. For this reason, ALFA students were invited to participate in the project in a few different, entirely optional ways, which included submitting a chapter of their own, helping with editing, or mentoring undergraduate students and providing feedback on their work. Ultimately, two of these ALFA students, Kevin and Carol Smith, chose to mentor students in my undergraduate class and visited several times to provide feedback on work-in-progress.

Student Contributions and Chapter Development

Here, I'd like to take a moment to fully outline the process that was used for the development of the chapters included in the text by students, especially in case other courses may like to adopt or amend this process.

Week 3: Students were asked to select a general topic early in the semester, by the end of week three. Their choice here was not yet binding, but meant to provide a guiding framework for their next step, as they developed a larger proposal for their writing. This week featured multiple guests visiting the course. Marilyn Billings gave a presentation that covered the overall goals of the ROTEL grant and why we are creating OER textbooks. She also helped students develop an understanding of the creative commons licenses available and we had a discussion about the type of license we wanted to assign to our project. Reneé Fratantonio also gave a demonstration on how to use library resources to complete research on their topic. Students were given time in class to begin researching their chosen topic and ask for help or guidance from myself and Fratantonio. ALFA mentors Kevin and Carol Smith were also in class on this day and had conversations with all the students in the course. In addition to any feedback the ALFA mentors gave, the challenge of putting their idea into words and talking through it with someone not directly involved in the class was itself valuable to students in the process of selecting their topic.

Week 6: Students were next tasked with writing an approximately 150-300 word proposal for their topic, requiring them to complete additional research on their proposed topic to make sure it was viable. ALFA mentors attended class again on this day to hear the revised and extended proposals and offer feedback. In this class session, students rotated through meetings with both the ALFA mentors and me to workshop these proposals and prepare for the next steps of writing a full chapter. Students also signed up for the day in the semester where they would lead a class session on their chosen topic.

Week 9: The first full draft of the chapter was due during the ninth week of the course, with the expectation that it may still be a bit rough around the edges as students were continuing to learn more about their chosen topic. For this draft, I provided big-picture feedback on the chapters-in-progress. This included suggesting parts of the topic that perhaps were not addressed or needed expansion. I also gave feedback on how students could more deeply address diversity and ethics within their topic. Rachel Graddy and Meagan Martin visited class on this day to discuss accessibility considerations for the writing process.

Week 14: The second draft of the was chapter due this week, and this was intended to be a complete and polished draft that students would consider ready for publication.

Week 15: Between weeks fourteen and fifteen, students all peer reviewed one another's work using Google Drive

commenting and suggesting tools. I also participated in this review process, leaving extensive feedback that included minor issues such as grammar as well as major suggestions for revisions.

Week 17: The final draft of these chapters was due during week seventeen, which was the scheduled final exam period for the class. No exam was given during this period, but students could attend with final questions about the project at this time.

Finally, I should note that some further editing was completed by me on these final drafts. However, in this round of editing I only focused on minor edits aimed at clarity and did not make any structural or thematic changes to the final product created by students. In short, students were provided with significant on-campus support, multiple rounds of iterative feedback, and opportunities to fine tune through three drafts of the chapters. Ultimately, this assignment was challenging for students, as the majority of them had not previously encountered any similar coursework about the implications of data on society, and were therefore exploring an entirely new subject area. The majority of students were also freshman or sophomores.

The Importance of Understanding the Implications of Data on Society

There were important tradeoffs in approaching the course in the manner described above and in letting students help develop the topics addressed in the class. Most significantly, this meant that I was selecting only about half of the overall content of the class in terms of the topics and readings assigned, while the rest was ultimately assigned through the decisions made by students. Therefore, I tried to highlight the major societal issues related to data. Briefly, I explored the following topics during the class:

A Brief History of Information and Big Data: Understanding the historical context of data development and the emergence of big data helps students appreciate the evolution of data-driven technologies and their impact on society.

Artificial Intelligence: Exploring the development and applications of artificial intelligence (AI) provides insights into the ways AI has revolutionized various fields and the ethical considerations that arise from its use.

Data Bias and Algorithms: Examining issues of data bias and algorithmic fairness is essential for understanding how data-driven technologies can unintentionally perpetuate existing biases and discriminatory practices. Students explore a range of readings on this topic, such as works by Jill Walker

Rettberg, Kate Crawford, Catherine D'Ignazio, Lauren F. Klein, and Safiya Noble.

Data Ethics: Delving into the ethical considerations surrounding data collection, use, and dissemination helps students develop a responsible and conscientious approach to data-driven practices. Resources such as "An Introduction to Data Ethics" by Shannon Vallor and William J. Rewak provide valuable guidance.

Quantified Self and Data Visualization: Investigating the quantified self movement and data visualization techniques enables students to explore how data shapes our understanding of ourselves and the world around us. Students engage with readings from Jill Walker Rettberg, Claudo Minca, Maartje Roelofsen, and the Tableau Public Blog.

Analyzing Social Media: Studying social network analysis methods allows students to examine the ways in which data informs our understanding of online interactions and social networks. Resources such as the works of Gruzd, Paulin, and Haythornthwaite, along with Netlytic Video Tutorials, provide a foundation for this exploration.

The Emergence of AI and ChatGPT in the Course

Finally, I believe it would be remiss not to address the historical significance of the rise of OpenAI's ChatGPT and other AI-based tools during the semester this course occurred.

ChatGPT had officially been released in November of 2022, shortly before the course began, therefore I was, to some degree, able to anticipate this change and include related readings on the syllabus. However, the rate at which the tool was updated and the speed with which it was adopted felt overwhelming at times and required hours of attention on a weekly basis to keep up with the ongoing developments.

Every few weeks in class we would check-in on these on-going developments, discussing especially the ethical issues connected with the technology. As part of our assigned lesson, students also spent time in class working with ChatGPT to better understand its affordances and limitations. One theme that emerged from our in-class discussions was that, at least in its current iteration, ChatGPT was very helpful as a brainstorming tool and to edit or explain existing text but was less helpful in generating specialized writing and essays, especially if they required the use of citations. Going further, one assignment in the class that students could choose from among a list was to test how ChatGPT performed on an assignment they had in another course. All students who completed that assignment reported that ChatGPT was not able to do a satisfactory job completing the assignment they chose.

Because this technology was emerging so quickly, the policy I put in place for the use of AI this semester was simply one that required transparency. I asked that students note any time they used generative AI tools along with how they were used.

This is a policy I plan to revisit after reflection on how it went this semester.

In an effort to promote transparency and ethical use of AI, students who incorporated insights or assistance from ChatGPT in their chapters were required to acknowledge its use. This practice ensures that readers are aware of the role of AI in the development of the content and fosters open dialogue about the implications of AI in research and writing. I used ChatGPT for brainstorming and editing purposes in my writing for this text, as a way to further experiment with the tool. However, I did not use it to entirely generate any portions of text.

PART I
WHY CARE ABOUT DATA AND SOCIETY?

Chapter Written by J.J. Sylvia, Ph.D.

"I'm not worried about privacy because I haven't done anything wrong."

– Most People

Learning Objectives

- Explain the ethical challenges and societal

implications of big data, including issues of privacy, trust, and potential for misuse.

- Understand the various approaches to regulating big data and why traditional methods like notice and consent or anonymization are increasingly insufficient.
- Critically discuss the intersection of race, gender, and capitalism in the realm of data science, recognizing how biases can be built into algorithms and data sets

INTRODUCTION

As a tenured professor deeply immersed in the confluence of digital media and posthuman philosophy, my life's work has largely revolved around deciphering the intricate web of technology, identity, and human experiences. Prior to becoming an academic, I worked in the ecommerce sector for two decades, and spent five years working for a nonprofit organization that helped K-12 schools better integrate educational technology. The impetus for this chapter comes from a profoundly personal place—a purposeful sense of self that draws from both my academic background, my professional background, and inherent interests in the subject at hand.

The journey we'll undertake in the following pages isn't just a scholarly expedition; it's also an exploration of my own evolving understanding of how technology can both empower and marginalize, illuminate and obfuscate. In this sense, the chapter serves as a dual lens: one that presents a specific subject matter through the filter of academic rigor, and another that invites you to understand how my own experiences and intellectual journeys have shaped this presentation.

My hope is that the ensuing discussions will not only add to your knowledge base but will also inspire you to consider

your own positionality—your unique vantage point formed by your experiences, background, and education. Just as I have connected my own life story to this area of study, I encourage you to discover your own connections, contradictions, and curiosities as we delve deeper into the complexities of this intriguing subject.

As I've taught about the impacts of big data and artificial intelligence (AI) over the years, I find myself frequently running headfirst into one formulation or another of the above quote, which I've obviously made up and not cited exactly. Or perhaps it's more accurate to say I've been trained on a large set of data that consists of responses to concerns about privacy, processed those through my neural network, and generated some predictive text that looks a lot like what most people say – much like any good large language model (LLM) would do as part of a generative AI process. Either way, developing an approach to teaching about data that cuts through the apathy associated with this quote, or one like it, has become a central focus of my pedagogy. Why exactly should we spend our precious time on this planet thinking about or even *caring* about ideas as abstract and hard to regulate as these?

As it turns out, there are quite a few good reasons. The challenge is these reasons are buried in layers of legal and bureaucratic jargon that, frankly, make it all sound quite boring. Comedian John Oliver described this best when

discussing the intricacies of net neutrality and cable companies on his show *Last Week Tonight:*

> Oh my god! How are you still so dull? And that's the problem. The cable companies have figured out the great truth of America. If you want to do something evil, put it inside something boring. Apple could put the entire text of Mein Kampf inside the iTunes user agreement and you'd just go, "Agree, agree, agree. What? Agree, agree." (Oliver, 2014)

Oliver goes on to distill the issue of net neutrality, explaining it in detail while also making it funny. While I would love to be able to do something like that in every class session that I teach, the amount of content I have to produce every semester while teaching four courses is far greater than the amount of content someone like Oliver produces for his show, and he has an entire team of writers helping him. Nonetheless, I've worked hard over the years to find ways to make the big picture questions about data and society both personal and interesting to students. Let's explore why this matters.

A CONSTITUTIONAL RIGHT TO PRIVACY?

In the United States, the right to privacy moved into the spotlight as part of the controversial 2022 Supreme Court decision in Dobbs v. Jackson Women's Health Organization. This decision removed federal protections for abortion rights, instead deferring the right to legislate abortion to individual states. As monumental and disruptive as that particular decision was, the fallout from legal precedent it overturned to do so is arguably even larger. The syllabus that gives an overview of the case explains:

> As to precedent, citing a broad array of cases, the Court found support for a constitutional "right of personal privacy." Id., at 152. But Roe conflated the right to shield information from disclosure and the right to make and implement important personal decisions without governmental interference. (Dobbs v. Jackson Women's Health Organization, Syllabus, 2022, p. 5)

Let's break down what this means. While the Court did not eliminate the constitutional right of personal privacy, it argued that the Roe v. Wade decision, which originally legalized abortion at the federal level, misconstrued how the right to

privacy actually works. Roe v. Wade made the argument that the right to privacy means that citizens have the right *both* to shield private information from disclosure to authorities *and* use that right to make personal decisions without government interference. In other words, the right to legally obtain an abortion was based on the constitutional right to privacy. The Dobbs v. Jackson Women's Health Organization breaks that link, arguing that the right to shield disclosure of an action does not confer the right to take that action. Said another way, although one has the right not to disclose information about whether they've had an abortion, that right does not make the act of getting an abortion legal.

Although this may appear to be a minor distinction, it potentially disrupts the entire foundation of the right to privacy in the U.S. The majority opinion said this ruling should not affect other cases on which legal rights were tied to the right to privacy. However, in a concurring opinion, Justice Thomas Clarence argued just the opposite:

> For that reason, in future cases, we should reconsider all of this Court's substantive due process precedents, including Griswold, Lawrence, and Obergefell. Because any substantive due process decision is "demonstrably erroneous," Ramos v. Louisiana, 590 U. S. ___, ___ (2020) (THOMAS, J., concurring in judgment) (slip op., at 7), we have a duty to "correct the error" established in those precedents, Gamble v. United States, 587 U. S. ___, ___ (2019) (THOMAS, J., concurring) (slip op., at 9). (Dobbs

> v. Jackson Women's Health Organization, 2022, Thomas, J., concurring, p. 3)

Here, Thomas is specifically arguing that in light of the Court's decision in Dobbs v. Jackson Women's Health Organization, the court should revisit other cases that used the same precedent as Roe v. Wade and correct the error in those cases. What cases does he mention? The 1965 Griswold v. Connecticut case predated Roe v. Wade and established a constitutional right to privacy, recognizing that married couples have the right to use contraceptives. The 2003 Lawrence v. Texas Supreme Court case declared laws criminalizing consensual same-sex sexual activity unconstitutional, affirming the right to privacy and striking down sodomy laws in the United States. The 2015 Obergefell v. Hodges Supreme Court case legalized same-sex marriage nationwide in the United States, recognizing it as a fundamental right protected by the Constitution. In short, Thomas is recommending that the Court revisit the cases that protected the rights to use contraception, to perform consensual same-sex activity, and same-sex marriage and "correct the error" that was made in those decisions.

Let's revisit that quote at the beginning of the chapter in light of this discussion:

> "I'm not worried about privacy because I haven't done anything wrong."
>
> – Most People

Rather than saying you aren't worried about privacy

because you haven't done anything wrong, you should instead understand that in the United States, at least, what counts as right or wrong under the law has long been guided by the constitutional protections of privacy. But we are now living in a shifting landscape where these protections will no longer stand on firm ground. Let's consider one very personal example of data and privacy that has shifted in light of the Dobbs decision.

Many women have long tracked their menstrual cycles for a wide variety of reasons, including, but not limited to better understanding their health, as a form of birth control, as a way to increase the likelihood of conception, for medical reasons, and to look for signs of menopause. A plethora of cell phone apps are available that can help track this information. None of the activities listed above are illegal, so it may be easy to believe there's no need to be concerned about privacy in this case. However, if one lives in a state where abortions are no longer legal post-Dobbs, this data can potentially be collected and used as evidence of abortion if there are irregularities (which can occur naturally) in menstruation cycles. Efforts to protect period-tracking app data specifically have thus far failed (Moomaw, 2023). It's important to note, though, that digital evidence can be collected and used against those who seek abortions from sources far beyond period-tracking apps, including wearable technology, internet-connected household appliances, purchase history, routine data gathering by government agencies, and social media usage (Conti-Cook,

2020). All of these sources of data travel outside of one's home because they travel over the internet, therefore they are not protected by the remaining right to privacy.

Without the legal protection of privacy, many of our previously guaranteed rights, including whom we marry, whom we engage with in sexual activity, and whether or not we have children are either no longer legal already or may not be in the near future. It doesn't get much more personal than that.

LITTLE BROTHER

George Orwell's dystopian novel coined the term "Big Brother" for overly intrusive governments that use surveillance to erode privacy. However, in the era of big data, the use of our data in other areas of society should also raise concerns, as we now live in a culture of algorithms. I have elsewhere called the private companies that use our data, as opposed to the government, Little Brother (Sylvia IV, 2016a). Here, too, I have learned that if one is going to care about how data is being used, the consequences of it need to be felt personally. Let me briefly walk through how this approach has developed as part of my professional work on topics related to big data.

An interactive H5P element has been excluded from this version of the text. You can view it online here: https://rotel.pressbooks.pub/datarenaissance/?p=38#h5p-1

Podcast: Living in a Culture of Algorithms

Episode Summary:

danah boyd weaves together her work on youth, privacy, and data-driven technologies, to examine the complicated social and cultural dynamics underpinning social media, the messiness of "big data," and the problematic implications of using algorithms designed for one problem to address societal issues without accounting for unintended consequences.

Aperveillance

My goal is to make questions about data come alive, using creative and/or artistic practices that allow us to understand the ethical challenges presented by big data in new ways. My first attempt at this was a project titled Aperveillance.

Figure 1. Aperveillance project displayed on the media wall at Hunt Library, North Carolina State University. The image shows a grid of multiple live webcam images overlaid by text noting the crimes committed in Raleigh, NC the previous day.

On my website, I describe the project in this way:

While teaching a special topics course (COM/ENG 395) on Big Data and the Rhetoric of Information, I asked my students to create data visualizations with either Tableau or P5.JS, and I joined them in creating my own project for the assignment. Although much of the concern in my field surrounds issues related to surveillance, I thought it would be interesting to think about the types of watching that are now becoming increasingly possible with open data. Thus, I coined the term

aperveillance for my project, which derives from the Latin "aper" meaning open, and "veiler" meaning to watch.

This project uses webcam images that are publicly available in North Carolina, primarily around Raleigh, but including other areas of the state... It also uses Raleigh's open crime data to randomly include information about the previous day's crimes juxtaposed on top of the webcam images. This is intended to provoke questions about the type of watching we as citizens are able to do with open data on the web. (Sylvia IV, 2016b).

This project was displayed as part of a Code+Art exhibit at the North Carolina State University library and at a local conference. However, on a whim, I made one last-minute tweak to the project, not visible above, which ended up being the most interesting part of the project. For anyone who was viewing the project on a device that had a camera attached or built-in to it, I grabbed an image from that local camera and mixed it randomly into the grid of local webcam photos. This was by far the aspect of the project that generated the most interest while the project was being displayed. To my surprise, audience members posed questions that reflected significant concerns about their privacy once they saw themselves displayed in the data. This served as a moment of inspiration that would lead to my next project.

I should also note that, although this project was viewable live on the web, the images taken from the viewer's webcam were only ever displayed on the local device on which one

was viewing the project. They were never displayed to anyone else via the internet or saved or archived in any way. But, the concerned reactions by viewers helped me better understand that even creative projects like this one, which used local data, would only have the impact I was seeking if the impact was felt in a truly personal way.

BECOMING DATA

This insight led directly to my next interactive project. While teaching at Fitchburg State University, I secured a small grant that allowed me to hire two students to help write and code this project. Much like the Aperveillance project, this project also relies on a bit of trickery, in which the program itself acts as if it's completing a data analysis which does not actually occur. However, it's important to understand that in both of these cases, the trick could actually be implemented for real, but is not done so in order to protect privacy. In other words, the Aperveillance project could have displayed the images from the local webcam live to everyone via the internet, and even saved them. And the Becoming Data project could have actually completed the data analysis that it fakes. However, I'm not personally interested in violating anyone's privacy – I simply want them to experience what such a violation feels like at a personal level that is not possible when discussing data abstractly.

Becoming Data uses a Microsoft Kinect and the Processing programming language to create an augmented reality interface in which users interact with the screen in order to start an analysis of their own data. It includes a fairly annoying terms of service agreement that must be navigated and

accepted. Then the system acts as if it is performing the following tasks, with a percentage completion bar and animations for each:

- Facial Recognition Analysis
- Pinging cell phone
- Sentiment Analysis
- Social Network Analysis
- Accessing Credit Data

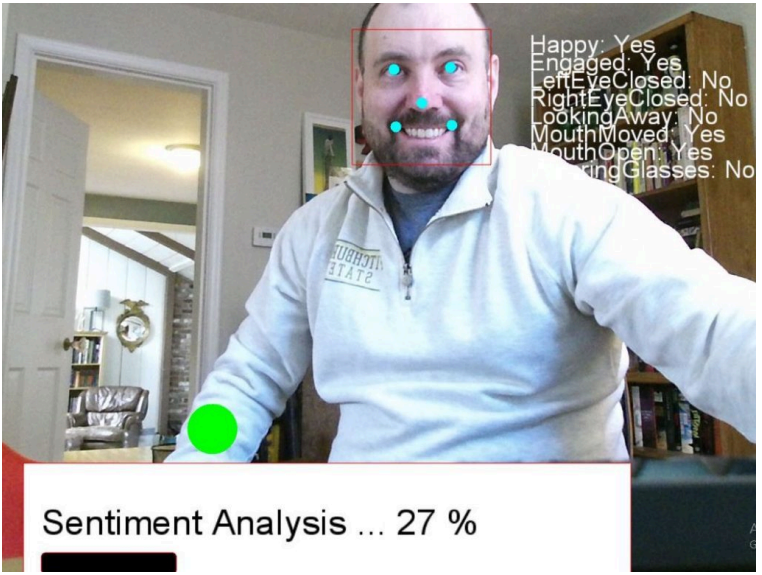


Figure 2. Becoming Data screenshot. The image shows the Microsoft Kinect in the middle of performing a sentiment analysis. The face in the image is surrounded by a red box with cyan dots marking the eyes, nose, and corners of the mouth. Text on the image indicates the man is happy, engaged, has both eyes open, is not looking away from the camera, moved his mouth, has his mouth open and is not wearing glasses.

Next, the program shares the following results screen-by-screen, with a simple random number inserted where each bold number is included below:

1. Total Time You Spent Reading License Agreement: [Actual number recorded] seconds. This agreement was approximately 2,500 words. The average time to read this

amount of text is between 8 and 20 minutes. What did you miss?

2. Analysis of recent food purchases indicates that you may be depressed! Social media channels will now feature ads for meditation apps and online therapy services **28%** more frequently. If you click one of these, you may start receiving ads suggesting you are bi-polar.

3. Location-based data collected from your cell phone indicates that you have visited the gym less frequently than the national average. Your health insurance rates will rise **7%** this year.

4. Based on an analysis of your cell phone battery, ride sharing services will increase your fare by about **33%.**

5. Car dealerships can access your recent search history. Based on an analysis of your recent searches, if you were to visit a dealer today, they might offer you a loan with an interest rate that is **2%** higher.

6. Using a 2015 patent, the creditworthiness of your friends across social media sites has been analyzed. Unfortunately, some of your friends have low scores. As some new credit card companies take this into account, your credit score could drop by as many as **171** points.

7. An analysis of all of your social media posts reveal that

your posts contain **76** bad words and **115** mentions of alcohol or drugs. These may or may not be problematic in context, but they have been archived and will be reported on your next employment background check.

8. The web browser you use most frequently on your phone has been correlated with an increased likelihood to leave a job sooner than other employees. Approximately **21%** of employers will not even consider your application due to the browser you use.

9. Your recent social media feeds show **65**% more ads with negative sentiments recently. This increase may be due to experiments run by the company or an influx of memes by Russian-backed ads. However, this change in your feed means you are **33**% less likely to vote in the next election.

10. A stress analysis of your face indicates that you are **57**% unlikely to be placated by a customer service representative. If you call for customer service now you will be routed to an operator who has been specially trained to handle difficult customers. They will be unwilling to meet your requests.

This project was completed days before the COVID-19 pandemic caused schools across the country to close in early 2020, so it has had limited opportunities for sharing. However,

it has been featured on my campus as part of the Speaker's Series for the Center for Teaching and Learning and as part of classes that I taught once students returned to in-person learning.

Participants and the audience members viewing the interaction have had very strong reactions, which usually first focus on how unfair a particular result is, followed by questioning if the results are real. As I noted above, the actual results presented are fake, but all of the situations shared in the results are based on real-world patents or applications of data use. This is the best way I have found to date to make the effects of Little Brother – the use of your data by corporations – feel real and feel concerning. Our personal data can be collected and used in ways that are detrimental to not only our wallets but also our very livelihood.

Does this make you care?

A/B AND MULTIVARIATE TESTING

I have further argued that even without access to *personal* data, massive amounts of data cause an ethical problem related to manipulation. This manipulation is related to the ethical implications of A/B and multivariate e-commerce optimization testing (Sylvia IV, 2010). These techniques, which allow e-commerce websites to test different versions of a page to improve outcomes like sales or reduce abandoned shopping carts, might seem innocuous. However, I believe there's more to consider.

I've been involved in owning or managing e-commerce websites for 20 years, and I first became aware of the issues surrounding this sort of testing in the first decade of my career. It was around this time that Google's free Web Analytics software launched and was available for free to the general public. This allowed virtually any site that wanted it to run these A/B and multivariate tests and collect data on them. I saw the power of these tools first hand as I integrated them into the site I was managing. It was witnessing this power that first raised my ethical concerns. This is part of what brought me back to school to pursue my master's degree, and later my doctorate.

I examined these practices through various ethical lenses and I've found that they can lead to manipulative site design. The goal is to subtly encourage consumers to spend more. Although another viewpoint might see these practices as aiding consumers—making websites more user-friendly or easier to navigate—I think the reality is more complex. The goal, much like the field of advertising in general, is to create new desires to purchase products you don't actually need. But this iterative process lets websites get really good, really quickly at persuading you in ways that are not at all transparent. How could you possibly imagine that the color of the checkout button on your favorite website makes you more likely to actually complete a purchase unless you've studied web design or communication theory?

The primary issue that I took at the time with these practices was that the *why* didn't matter. Why does a certain size and color button make people spend more? Why does a certain shade of blue make people more likely to click a link? This type of testing cannot answer that question. As we transitioned into the age of big data, that problem has only become more pronounced. Big data is very good at making correlations between things, but not able to explain why those correlations exist.

And this brings us face-to-face with the difficult theoretical questions we must all face in the age of big data.

BIG DATA

The Five V's

What we refer to as big data is typically defined through the five v's definition: volume, velocity, variety, value, veracity. Put as simply as possible, the five v's include a massive amount of different types of data that are being collected with increasing frequency from multiple sources. Outputs are providing great value to the organizations that can make use of it, while presenting significant challenges if one needs to determine the accuracy or truth of the content represented by such data.

Where does all of this data come from?

Early on, most of it was generated by human actions, through the data we leave as we browse the internet and use devices with sensors built into them, from our cell phones and smart watches to the thermostats and doorbells in our houses. But the low cost and huge amounts of data generated by sensors has led to their implementation into smart cities, shipping processes, and beyond in ways that allow them to collect data

on the world that goes beyond the human. For example, most international shipping now uses RFID tags to collect information and monitor shipments. Just how cheap are all of these censors? According to DuBravac (2015), a typical smartphone in 2015 could have all of the following sensors for an additional $5.00 in manufacturing costs: proximity, ambient light, accelerometer, gyroscope, magnetometer, ambient sound, barometer, temperature/humidity, and M7 motion. Check out the documentary below for an overview of how big data is being used:

Documentary: The Human Face of Big Data

Documentary: The Human Face of Big Data on Vimeo

But this leads to yet another question: why do we so willingly give up all of this data for free to corporations that use it to manipulate us and increase their profits?

Access to Data: Weapons of Math Destruction

Although he has since fallen into significant controversy

because of his political views, journalist Glenn Greenwald (2014) spoke clearly about this challenge in his TEDGlobal talk. Greenwald was one of the journalists who helped NSA whistleblower Edward Snowden publish his story about the way that the U.S. government was abusing the U.S. Patriot Act to illegally collect information on U.S. citizens. In that speech, Greenwald notes that we do seem to intuitively care about privacy. For example, if someone were to ask us for our email address and password, we very likely wouldn't share that information, even with close friends.

And yet, we give up the contents of our personal email to corporations like Google and the details of our social lives and personal messages to companies like Meta, which owns Messenger, Instagram, and What's App. One possible reason we might feel comfortable sharing this information is because we trust these companies. For many, this was explicitly true when it came to Google, at least for many years. However, not everyone trusts technology firms in the same way:

> When we consider the race of our respondents, white individuals (the baseline/omitted category in our model) are the racial group that is least confident in the three tech companies, save for respondents who identified as multi-racial or as some race other than our main four groupings. Interestingly, there doesn't seem to be a meaningful difference between Asian, Hispanic, or Black respondents. (Kates et al., 2023, para. 17)

In short, Asian, Hispanic, and Black people trust technology

firms such as Google more than white individuals. This means that they are more likely to share personal data and less likely to consider the negative impacts that can stem from that sharing. Further, any education past high school led to a decrease in trust. Gender showed some difference in trust levels, but was relatively small or had a small enough sample size so as to decrease the overall statistical significance of the results:

> ...respondents identifying as female [were] slightly more confident than males in our tech companies, but the substantive magnitude of this difference is quite small. Those identifying as either non-binary or neither male nor female, however, are vastly less confident, though our results only reach significance at the 0.10 level, given the paucity of such respondents in our panel. (Kates et al., 2023, para. 19)

Until they eliminated it in 2018, Google's company motto was "Do No Evil." If you've been paying attention to the world of technology, you can already see where this story is heading. Google has been the subject of antitrust investigations, security vulnerabilities that left personal data accessible, and fears of search-induced filter bubbles that may have helped sway political elections. Many of those who trusted Google with their intimate and personal data in the early 2000s no longer do so. Although people have lost trust in all institutions, their trust in technology companies, in particular, decreased the most drastically between 2018 and 2021. Notably, this was true across every sociodemographic category analyzed (Kates

et al., 2023). Overall, trust in technology companies has decreased for everyone.

Cathy O'Neil describes the use of this data in the form of algorithms, "weapons of math destruction." In the podcast below, she explains how this works and how it magnifies inequality in our society.

Podcast: Weapons of Math Destruction with Cathy O'Neil

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://rotel.pressbooks.pub/datarenaissance/?p=45#h5p-2

Data & Society: Weapons of Math Destruction

Episode Summary:

Tracing her experiences as a mathematician and data scientist working in academia, finance, and advertising, Cathy O'Neil will walk us through what

she has learned about the pervasive, opaque, and unaccountable mathematical models that regulate our lives, micromanage our economy, and shape our behavior. Cathy will examine how statistical models often pose as neutral mathematical tools, lending a veneer of objectivity to decisions that can severely harm people at critical life moments.

Cathy will also share her concerns around how these models are trained, optimized, and operated at scale in ways that she deems to be arbitrary and statistically unsound and can lead to pernicious feedback loops that reinforce and magnify inequality in our society, rather than rooting it out. She will also suggest solutions and possibilities for building mathematical models that could lead to greater fairness and less harm and suffering.

However, even if that's not your personal experience, or even if there is a corporation you trust implicitly, no corporation lasts forever. And when that company is sold or dissolved, its assets are often transferred elsewhere, possibly to much less trustworthy owners. Although we may be aware of that possibility in the abstract, I would like to share a case study about how the implications of this process impacted me.

Livejournal Case Study

This reality became personal for me in 2019, as I was researching Russia's internet policies as part of an article I was writing with a colleague about Russia's interference via social media in the 2016 U.S. presidential election. While doing that research, I discovered that the social media site LiveJournal, which had been popular in the very early 2000s, had not only been sold to Russian oligarchs, but all of their servers were physically moved to Russia. Why did this matter so much to me?

A short history of LiveJournal can make this clearer. Its origin story is somewhat similar to that of Facebook in that it was launched out of the college dorm room of its creator Brad Fitzpatrick in 1999. I had already been blogging for several years by the time the site began to gain popularity. In fact, as best as I can tell, I very likely had one of the first one hundred blogs ever published on the internet when I launched mine as a high school sophomore in 1998. My friends and I competed with one another to release new and more creative features for our blogs. But this interest in the software behind the blog gave way to a more sustained interest in the content of the blogs. Fitzpartick's new site also allowed the creation of friends lists, which meant that rather than taking the time to visit each of our blogs separately, we could all sign up for accounts and have the most recent updates appear in one feed, in chronological

order. This is standard today, but was a huge leap forward when it was created.

This means I was using LiveJournal as I transitioned from high school to college. This can be a highly emotionally turbulent time, as you may be aware. Many of us who used LiveJournal at the time would write very long and very personal entries. Of course, it also had quite advanced security features, meaning you could create customizable lists that determined who could see each specific post. While this particular feature still exists on some platforms today, it has yet to be replicated in such an intricate way as LiveJournal allows. I also wasn't alone in my usage of LiveJournal. It peaked at over 2.6 million active users within a 90-day period in 2005.

These filters, and an implicit trust in Fitzpatrick, gave me confidence to write about very personal things online. Because Fitzpatrick also posted in his own journal, it felt very much as if I knew him personally, though my later study of communication theory would reveal that this was really only a parasocial relationship. As my life continued to evolve, I slowly stopped using the platform, and hadn't thought about it in some years until the day I stumbled across the news of its move to Russia. Why does this all matter to me?

The short version is that LiveJournal was sold a few times over the years before it ultimately ended up in Russia. The key here is that Russia's laws allow the government to access any information on servers located in their country, without the kind of strong protections like the need for a warrant that

are in place in the United States. Does it really matter that the Russian government now has easy access to all of my old private, password protected writing? Probably not. I haven't revisited the volumes of writing I did there in well over a decade, but as far as I remember, there was nothing truly egregious that I ever posted. But at minimum, the detailed musings of myself as a teenager could certainly be embarrassing and almost definitely cringeworthy to the version of me that is now a tenured professor. The types of things people posted about then weren't as curated and glossy as they are today. We would post about things we clearly coach people not to post on the internet today.

As my professional research has progressed into criticisms of Russia and their impacts on democracies around the globe, a small voice in my head can't help but wonder if there's something somewhere in all of that writing that could be used against me, especially if it were taken out of context. Russia is well known to operate blackmail schemes.

And to think, all of that worry because the teenage version of me placed so much trust in Brad Fitzpatrick. And yet, we know that others are at much greater risk. In the 2016 election, Russian troll factories specifically targeted Black and Latinx U.S. voters on social media, actively dissuading them from voting at all as a way to bolster Donald Trump's success in the election. Since then, their methods have gotten even more complex. For example, they have set up fake sites designed to look like they offer help for those struggling with their sexual

identity and how or whether to share it with friends and family. The Russian trolls then use those conversations to blackmail the participants into taking actions that advance Russian goals (Sylvia and Moody, 2019).

Racial Capitalism

As we saw in the last section, everyone is at risk when our personal stories and data become entangled with websites, even those we may initially trust. However, that risk is not evenly dispersed, as marginalized people are almost always the most significantly impacted by the challenges our society faces related to data and algorithms. These challenges have many layers, but they begin at the very beginning of our technology, during the coding process itself. If we're discussing Little Brother, corporations who use data, then connections between capitalism and racism are a necessary piece of the puzzle needed to untangle this story.

Sometimes, these implicit biases emerge because the technology is created predominantly by white people who only test the code on other white people or use data sets that don't reflect diverse people and/or skin tones. Why does this happen? The technology workforce is overwhelmingly white. For example, only 4% of Google's workforce is Black, Black people represent only 1% of tech projects that receive venture funding (Russonello, 2019). The following documentary, *Coded Bias*, explores these challenges:

Documentary: Coded Bias by PBS

PBS: Coded Bias

In an increasingly data-driven, automated world, the question of how to protect individuals' civil liberties in the face of artificial intelligence looms larger by the day. Coded Bias follows M.I.T. Media Lab computer scientist Joy Buolamwini, along with data scientists, mathematicians, and watchdog groups from all over the world, as they fight to expose the discrimination within algorithms now prevalent across all spheres of daily life.

While conducting research on facial recognition technologies at the M.I.T. Media Lab, Buolamwini, a "poet of code," made the startling discovery that some algorithms could not detect dark-skinned faces or classify women with accuracy. This led to the harrowing realization that the very machine-learning algorithms intended to avoid prejudice are only as unbiased as the humans and historical data programming them.

Coded Bias documents the dramatic journey that

follows, from discovery to exposure to activism, as Buolamwini goes public with her findings and undertakes an effort to create a movement toward accountability and transparency, including testifying before Congress to push for the first-ever legislation governing facial recognition in the United States and starting the Algorithmic Justice League.

These problems have most famously been explored by Safiya Noble (2018) in her book *Algorithms of Oppression*. Noble ultimately links these algorithmic problems back to capitalism, because they are created primarily by privately held companies whose main goal is to generate profit. Additionally, U.S. law of the past several decades has allowed many sites to function as monopolies that are able to purchase any potential competitors. A major example of this is Meta's purchases of Instagram and What's App. She explains this in greater deal in the following podcast:

Podcast: Algorithms of Oppression with Safiya Noble

Data & Society: Algorithms of Oppression

Episode Summary:

In “Algorithms of Oppression”, Safiya Umoja Noble challenges the idea that search engines like Google offer an equal playing field for all forms of ideas, identities, and activities. Data discrimination is a real social problem; Noble argues that the combination of private interests in promoting certain sites, along with the monopoly status of a relatively small number of Internet search engines, leads to a biased set of search algorithms that privilege whiteness and discriminate against people of color, specifically women of color.

Through an analysis of textual and media searches as well as extensive research on paid online advertising, Noble exposes a culture of racism and sexism in the way discoverability is created online. As search engines and their related companies grow in importance—operating as a source for email, a major vehicle for primary and secondary school learning, and beyond—understanding and reversing these disquieting trends and discriminatory practices is of utmost importance.

The capitalist imperative for profit is often either at the root of, or exacerbates these challenges. This is due in large part to the way that the internet has evolved and the way that many technology companies rely on advertising for their revenue. When a site relies on advertising to make money, they make more money the longer everyone stays on their site. This creates problematic outcomes, like YouTube's suggested viewing algorithm leading viewers to watch increasingly radicalized content (Sylvia and Moody, 2022). This approach has been dubbed the "Attention Economy," and you can learn more about its promises and perils in the following podcast:

Podcast: Adtech and the Attention Economy

Data & Society: Adtech and the Attention Economy

Episode Summary:

Data & Society Sociotechnical Security Researcher Moira Weigel hosts author Tim Hwang to discuss the way big tech financializes attention. Weigel and Hwang explore how the false promises of adtech are just one example of tech-solutionism's many fictions.

Of course, these problems are not limited to the United States, as they ripple out to the entire Global South. Racial capitalism is deeply ingrained in modern capitalist structures, affecting everything from labor markets to social movements. Exploring these challenges can be difficult. While racial capitalism was initially described as a form of data colonialism, recent scholars have suggested this may oversimplify what's happening. The podcast below, featuring Sareeta Marute and Emiliano Treré, explores the challenges while also highlighting possible avenues of resistance, underscoring the need for a critical examination of how data, race, and capitalism intersect in today's world.

Podcast: Data & Racial Capitalism

Data & Society: Data & Racial Capitalism

Episode Summary:

The conversation between the host and guests Sareeta Amrute and Emiliano Treré delves into complex issues such as digital activism, data colonialism, racial capitalism, and the Global South. Emiliano explores the challenges faced by indigenous and marginalized groups in Mexico, while both

guests discuss the multifaceted nature of the Global South and critique the term "data colonialism." They also explore the pervasive algorithmic condition, the complexities of resistance, and the privilege and impossibility of disconnection. Sareeta's insights into IT workers in Berlin and their relationship with code highlight nuanced forms of resistance. The conversation concludes with an emphasis on everyday "counter conducts" and the importance of recognizing life outside of the algorithmic condition, offering hope for a more equitable and just future.

Additionally, it's important to consider feminist critiques of existing data practices. Data Feminism is an emerging field that intersects data science, feminism, and social justice, aiming to address the limitations of traditional data science methodologies. This approach applies an intersectional feminist lens to scrutinize who is involved in data collection, the purpose behind it, and the potential consequences for various communities. By doing so, it seeks to create a more ethical and inclusive data science practice that is sensitive to power dynamics, systemic inequalities, and context (D'Ignazio & Klein, 2020).

Ethical considerations are paramount in this interdisciplinary field, especially when dealing with big data

collaborations between development organizations and large tech corporations. The concept of the "paradox of exposure" is introduced to question the benefits and risks of being counted in data sets, particularly for marginalized communities. This nuanced approach calls for participatory methods and co-creation to ensure that data collection and interpretation are both ethical and contextually appropriate (D'Ignazio & Klein, 2020).

The definition of what constitutes "data science" is also under scrutiny in this framework. Traditional definitions often marginalize interdisciplinary approaches and specific groups, particularly women and people of color. Data Feminism advocates for a broader, more inclusive definition that values ethical considerations and innovation from marginalized communities. This not only leads to more accurate and robust data science but also contributes to a more equitable and just society (D'Ignazio & Klein, 2020).

You can learn more about this in the following podcast, featuring the authors of the 2020 book, *Data Feminism:*

Podcast: Data Feminism

Data & Society: Data Feminism

Catherine D'Ignazio and Lauren F. Klein discuss their new book "Data Feminism," with Data & Society's Director of Research Sareeta Amrute.

Regulating Data

At this point, you may be wondering why we don't simply create better laws to address these issues with big data, and for example, prevent monopolies or the sale of social networks to foreign countries. While we could perhaps legislate the rules around how companies can be sold, regulating the actual use of big data turns out to be quite complicated. The reason for this goes back to the *why* question we addressed earlier, or rather the lack of the *why* question in the correlations made by big data. Let me explain.

Big data, by its nature, relies on the secondary usage of data, meaning it explores the connections between points of data that weren't understood or weren't the primary reason for collecting that data. An example of the primary use of data would be the collection of web-browser usage to understand how people are accessing a site and the most commonly used browser for which it should be best designed. A secondary usage of part of that data could be used to link browser usage to employment records in order to correlate browser choice

with job performance. Browser usage data was not collected with that potential connection in mind, but a correlation was discovered in the data. Why is that true? My students love to speculate and try to create possible explanations, but the truth is, we simply don't know.

We could ban all secondary uses of data, but this would mean that we miss out on the good things big data can do: predicting outbreaks, preventing fires in New York City, fraud prevention, medical research on how wearables can predict upcoming heart attacks before they happen, etc. The point of big data is function creep. The function is the creep.

I've written elsewhere about potential regulation options that have been explored, but ultimately cannot be successful (Sylvia IV, 2016a). It's worth exploring these in detail to understand the significant challenges.

Notice and Consent

First, historically we have attempted to regulate data usage through notice and consent as part of the terms of service for a site or app. This approach is based on the 1980 Organization for Economic Co-operation and Development (OECD) Guidelines. The guidelines require users to be notified during sign up about what data will be collected and how it will be used. While this has always had limitations, it no longer even makes sense in the age of secondary uses of big data. Notice and consent is supposed to explain how your data will be used

and give you the option to consent to that usage. While this is at least feasible for primary uses of data, we simply cannot know ahead of time what connections secondary uses of data will make. This means notice and consent practices have had to evolve to be so broad they essentially allow any use of the data generated, which more often than not passes through the servers of multiple different companies as part of analytics and ad serving processes. To truly understand how your data would be used, you would also need to read the notice and consent statement for every company through which your data passes.

The ability to read and understand such policies is also impacted by language barriers, especially for global technology companies. Many companies do not publish their terms of service or community guidelines in the languages of all of the people they serve. As of March 2019, Facebook translated their community standards into 41 of the 111 languages offered, Instagram 30 out of 51, WhatsApp 9 of 58, YouTube 40 of 80, Twitter 37 of 47, and Snapchat 13 of 21 (Fick and Dave, 2019). It's important to note users also encompass more languages than those officially supported by the platform. Additionally, Fick and Dave reported that Facebook translates the policies when a critical mass of users speak a specific language, but have no threshold for what they consider a critical mass.

There are additional challenges with this approach. Most sites have adopted a policy that allows only use or non-use of their site depending on whether or not you consent to the use

of your data. If you don't consent, you don't get access to the services. The power dynamic here is tilted entirely in favor of the large corporations. If you're on the job market seeking a new position, how likely are you to opt out of using a service like LinkedIn if you don't fully agree with how they will use your data?

Further, these policies are difficult to read and time-consuming. A few years ago, I explored Facebook's terms of service only as they related to the use of data. An analysis showed that it would take the average person about 15 minutes to read that policy. Perhaps worse, the policy was written at an approximate average grade level of 13, meaning one would need at least some college education to be able to fully understand the policies. This is particularly problematic because 54% of adults in the U.S. are literate below the 6th-grade level (Rothwell, 2020). This puts white individuals, followed closely by Hispanics, at the greatest disadvantage because they have the highest rate of low literacy skills in the U.S. (35% White, 34% Hispanic, and 23% Black) (National Center for Educational Statistics, 2019). Researchers Lorrie Faith Cranor and Aleecia McDonald (2008) found the average length of a privacy policy to be 2,514 words, which would take the average person ten minutes to read. They then figured out that the average person visits between 1,354 and 1,518 websites in a given year. This comes out to requiring twenty-five full days a year, or seventy-six work days to read all of the policies associated with the websites we visit. Using some

further calculations, they determine that if everyone in America read every privacy policy they're supposed to, it would add up to a nationalized total of 53.8 billion hours. This has likely increased quite significantly since this calculation was done in 2008.

We all joke about how no one reads these terms of service. But there's a reason. We couldn't possibly have enough time to actually read them. But most importantly, it's simply not possible to tell users what the secondary uses will be ahead of time.

Anonymization

One suggestion is built on the historically successful model of anonymizing data. However, it has become quite apparent that in the age of big data, the larger the data, and the more data sets that can be combined, the harder it becomes to truly anonymize any data in a way that prevents it from being anonymized by someone determined enough to do so. Many years ago, Chris Whong (2014) was able to access New York City taxicab data through a Freedom of Information Laws request. Although the data had been anonymized before being released, he was able to correlate data with publicly posted photographs to determine particular rides celebrities took, including how much (or how little!) they tipped. He was able to take this a step further by finding clusters of rides that dropped off in the same neighborhoods over time, and tie

this to public records and social media accounts to identify a specific person who was regularly using taxis to visit gentlemen's clubs. This is a relatively straightforward example, but the larger point is that when enough data can be connected and correlated, deanonymization becomes much easier.

Deletion

Viktor Mayer-Schönberger (2009) has argued that we can make technical changes to how data is created and stored in computer systems. This proposed change would essentially allow all data to be given an automated deletion date. For example, all posts made to Twitter might be set to automatically delete after a one-year time period.

While this would certainly work from a technical standpoint, there are several practical challenges associated with this. For example, we would likely want to create the possibility to extend or change the date of deletion, which leaves open the possibility of such extensions happening indefinitely. This makes sense, as we may not want to automatically delete treasured family photographs, for instance. Furthermore, the question of who gets to set the deletion time period will be of utmost importance. If this is left to the corporations collecting data, they may simply extend the time period to be quite long.

Here, though, we have to also remember the deeper dynamics of big data. Even if we created new, incredibly strict

regulations that put the power of choosing the time period for deletion into the hands of individual users rather than corporations, this approach would yet again risk losing some of the positive benefits that big data promises. For example, the heart rate data collected by wearables today might provide the data that an algorithm in 30 years time is able to use to predict and prevent the onset of various degenerative diseases. We might need significant longitudinal data to make exciting new correlational breakthroughs. These types of interventions would be most beneficial to the elderly and those with chronic diseases or cardiovascular risks (Chandrasekaran, 2020). Black adults and American Indians are twice and 1.5 times as likely to suffer from cardiovascular risks as White adults, so such advances could be especially helpful for those populations (Javed et al., 2022).

Regulate Harmful Uses

A Microsoft Global Privacy Summit (n.d.) suggested that regulators should focus on creating laws that prevent harmful uses of data. The discussions at this summit attempted to update the original OECD guidelines that promoted notice and consent. But these ultimately expanded the uses of data available to corporations so long as they weren't deemed harmful by "society," a deeply vague and problematic term. I further analyzed this proposal in this way:

> Rather than truly being guidelines for protecting the privacy of consumers, they are instead guidelines for managing the power wielded by corporations...
>
> Much of the data storage and processing is now done in the cloud, meaning through distributed computing. Big data projects are especially likely to be done this way because individual computers are often not powerful enough to process such large amounts of information, giving rise to services such as Apache's Hadoop, which offers just such distributed computing. This cloud computing, in combination with website services being distributed to so many third-party organizations, means that data flows are frequently crossing many different borders spanning organizations, nations, and most importantly, legal frameworks. Even if the United States were to create strong laws as a dissuasion to using data, it seems likely that data-reliant organizations would find a welcoming home in other countries with less strict laws. This process might, for instance, mirror those transformations in online gambling. Though illegal in the U.S., the servers are hosted in other countries, and still relatively accessible by U.S. citizens. (Sylvia IV, 2016)

Put simply, restrictive laws in one country might cause the servers to be moved to more lenient countries. In the case of online gambling cited above, there has been an increased push by several states to legalize and provide access to such gambling so that the taxes on such activities are not lost to other countries.

Ultimately, the biggest question here is who gets to decide what uses are harmful. The answer to that question moves

out of the realm of privacy and into the realm of power and control.

AN OPEN QUESTION

Due to these challenges, it remains an open question how we might regulate the use of big data in ways that allow for its beneficial uses but prevent the harmful uses, at least in part because of challenges related to who gets to decide what counts as beneficial and harmful. Privacy protections would in theory allow users to decide when and what data of theirs to protect, but as we saw at the beginning of this chapter, privacy protections are in the midst of an erosion in the United States. Further, existing privacy protection only applies to materials located on one's own property, so any data that flows across the internet is not protected in that way.

I hope you can see clearly that these challenges related to big data and privacy apply to all of our daily lives. They are pressing, important, and difficult. But understanding what these challenges are is of utmost importance. The emergence of generative AI into prominence in 2022 and 2023 has made such questions even more pressing. Ethical discussion guides included in this book can be used to help start those conversations.

WRAP UP

Key Takeaways

- The issue of trust in technology companies is complex and varies across different demographic groups, with factors like race, gender, and educational level influencing how much personal data individuals are willing to share.
- Traditional methods for data regulation, such as notice and consent or anonymization, are becoming increasingly inadequate due to the complexities and secondary uses of big data, making it difficult to genuinely protect user privacy.
- The field of data science is grappling with ethical concerns, particularly around biases

that can affect marginalized communities; these biases are often unintentionally built into algorithms due to a lack of diversity among those who create and test technology.

- The regulation of big data faces significant challenges, including jurisdictional issues and the fundamental question of who gets to define what constitutes harmful or beneficial use of data, making it a complex issue of power and control.

Exercises

1. In what ways do you personally trust or distrust technology companies with your data? Do you think your race, gender, or educational level influences your level of trust? Discuss your reasons.

2. Choose one method of data regulation discussed in the material (e.g., notice and consent, anonymization, deletion) and argue its pros and cons. Can you suggest any modifications to make it more effective in the age of big data?

3. Listen to one of the podcasts mentioned in the material and summarize its key points. How does the podcast deepen your understanding of the ethical challenges posed by big data, and what solutions does it offer?

REFERENCES

Chandrasekaran, Ranganathan, Vipanchi Katthula, and Evangelos Moustakas. "Patterns of Use and Key Predictors for the Use of Wearable Health Care Devices by US Adults: Insights from a National Survey." *Journal of Medical Internet Research* 22, no. 10 (October 16, 2020): e22443. https://doi.org/10.2196/22443.

Conti-Cook, Cynthia. "Surveilling the Digital Abortion Diary." SSRN Scholarly Paper. Rochester, NY, October 28, 2020. https://doi.org/10.2139/ssrn.3666305.

D'Ignazio, Catherine, and Lauren F. Klein. *Data Feminism*. Strong Ideas Series. Cambridge, Massachusetts: The MIT Press, 2020.

Dobbs v. Jackson Women's Health Organization, 597 ___ U. S., 2022. https://supreme.justia.com/cases/federal/us/597/19-1392/case.pdf

DuBravac, Shawn. *Digital Destiny: How the New Age of Data Will Transform the Way We Work, Live, and Communicate*. Washington, DC: Regnery Publishing, a Salem Communications Company, 2015.

Greenwald, Glenn. *Why Privacy Matters*. TED Talks. TED, 2014. https://www.ted.com/talks/glenn_greenwald_why_privacy_matters?language=en.

Fick, Maggie, and Paresh Dave. "Facebook's Flood of Languages Leave It Struggling to Monitor Content." *Reuters*, April 23, 2019, sec. Media Industry. https://www.reuters.com/article/us-facebook-languages-insight-idUSKCN1RZ0DW.

Javed, Zulqarnain, Muhammad Haisum Maqsood, Tamer Yahya, Zahir Amin, Isaac Acquah, Javier Valero-Elizondo, Julia Andrieni, et al. "Race, Racism, and Cardiovascular Health: Applying a Social Determinants of Health Framework to Racial/Ethnic Disparities in Cardiovascular Disease." *Circulation: Cardiovascular Quality and Outcomes* 15, no. 1 (January 2022): e007917. https://doi.org/10.1161/CIRCOUTCOMES.121.007917.

Kates, Sean, Jonathan Ladd, and Joshua A. Tucker. "How Americans' Confidence in Technology Firms Has Dropped: Evidence from the Second Wave of the American Institutional Confidence Poll." Brookings, June 14, 2023. https://www.brookings.edu/articles/how-americans-confidence-in-technology-firms-has-dropped-evidence-from-the-second-wave-of-the-american-institutional-confidence-poll/.

Mayer-Schönberger, Viktor. *Delete the Virtue of Forgetting in the Digital Age*. Princeton, NJ: Princeton University Press, 2009.

McDonald, Aleecia M., and Lorie Faith Cranor. "The Cost

of Reading Privacy Policies." *I/S: A Journal of Policy for the Information Society* 4, no. 3 (2008): 543–68.

Microsoft Global Privacy Summit Summary and Report Outcomes, pp. 1-24. Retrieved April 24, 2014, from http://www.vmsweb.net/attachments/pdf/NoticeandConsent.pdf

Moomaw, Graham. "Youngkin Administration Opposes Shielding Menstrual App Data from Search Warrants." NBC 12 News, February 15, 2023. https://www.nbc12.com/2023/02/15/youngkin-administration-opposes-shielding-menstrual-app-data-search-warrants/.

National Center for Educational Statistics. "Adult Literacy in the United States," July 2019. https://nces.ed.gov/pubs2019/2019179/index.asp.

Noble, Safiya Umoja. *Algorithms of Oppression: How Search Engines Reinforce Racism*. New York: New York University Press, 2018.

Oliver, John. *Net Neutrality: Last Week Tonight with John Oliver (HBO)*, (2014). https://www.youtube.com/watch?v=fpbOEoRrHyU.

Rothwell, Jonathan. "Assessing the Economic Gains of Eradicating Illiteracy Nationally and Regionally in the United States." Gallup, September 8, 2020. https://www.barbarabush.org/wp-content/uploads/2020/09/BBFoundation_GainsFromEradicatingIlliteracy_9_8.pdf.

Russonello, Giovanni. "How Big Tech Allows the Racial Wealth Gap to Persist." *The New York Times*, June 21, 2021, sec. U.S. https://www.nytimes.com/2021/06/21/us/politics/big-tech-racial-wealth-gap.html.

Sylvia IV, J.J. "The Ethical Implications of A/B and Multivariate E-Commerce Optimization Testing." In *Ethical Issues in E-Business: Models and Frameworks*. Hershey, NY: Business Science Reference, 2010.

Sylvia IV, J.J. "Little Brother: How Big Data Necessitates an Ethical Shift from Privacy to Power." In *Controversies in Digital Ethics*, edited by Amber Davisson and Paul Booth. Bloomsbury Academic, 2016a. https://doi.org/10.5040/9781501310553.

Sylvia IV, J.J. "Research." *J.J. Sylvia IV* (blog), August 29, 2016b. http://www.jjsylvia.com/research/.

Sylvia IV, J.J., and Kyle Moody. "False Information Narratives: The IRA's 2016 Presidential Election Facebook Campaign." In *Handbook of Research on Deception, Fake News, and Misinformation Online:*, edited by Innocent E. Chiluwa and Sergei A. Samoilenko. Advances in Media, Entertainment, and the Arts. IGI Global, 2019. https://doi.org/10.4018/978-1-5225-8535-0.

Sylvia IV, J.J., and Kyle Moody. "BreadTube Rising: How Modern Creators Use Cultural Formats to Spread Countercultural Ideology." CLCWeb: Comparative Literature and Culture 24, no. 1 (August 15, 2022). https://doi.org/10.7771/1481-4374.4291.

Whong, Chris. “FOILing NYC’s Taxi Trip Data” Chris Whong (blog), March 18, 2014. https://chriswhong.com/open-data/foil_nyc_taxi/.

PART II
GENERATIVE AI IN THE CLASSROOM AND WORKSPACE

Chapter Written by J.J. Sylvia IV and Elise Takehana[1]

Learning Objectives

- Explain the key differences between ChatGPT-3 and ChatGPT-4, including their capabilities and limitations.
- Develop an understanding of how generative AI can be utilized in various career paths, and

1. This guide was designed to be used in class as a way to introduce the topic of generative AI.

be able to critically assess the ethical and practical implications of its use in those fields.

- Acquire practical skills in generating effective prompts for ChatGPT, and will be able to evaluate the AI's outputs for quality, relevance, and potential biases.

GENERATIVE AI PRE-TEST

Please complete the Pre-Test in 15 minutes or less.

HOW GENERATIVE AI WORKS

Warm-up

Word Association

Choose a few of these to discuss as a group:

- The dog chased its [blank].
- I put my homework in my [blank].
- He hit the baseball with a [blank].
- She wore a beautiful red [blank].
- We watched the movie with a bucket of [blank].
- The teacher wrote on the [blank].
- During summer, I love to swim in the [blank].
- I read the entire book but didn't understand the [blank].

- Every morning, she drinks a cup of [blank].
- He listened to his favorite song on the [blank].
- For my birthday, I got a new [blank].
- The astronaut looked out at the [blank].
- She likes to paint with water [blank].
- I play my favorite video game on the [blank].
- The athlete runs fast on the [blank].
- My favorite pizza is topped with [blank].
- The bear in the zoo loves to eat [blank].
- They cheered as their team scored a [blank].

There's also a connection here to cell phone predictive texting.

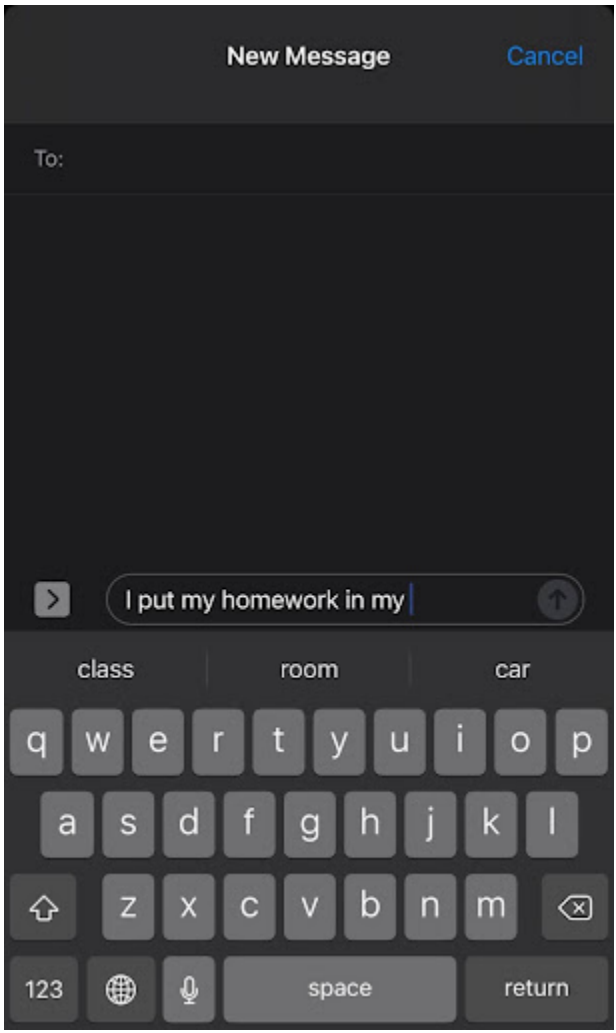


Figure 3: A screenshot of a cell phone text message being typed. The text being typed reads, "I put my homework in my," and the suggested autofills include, class, room, and car.

Markov chains use likelihood as predictors for the next in a

sequence of words. How is this predicted? It's based on the texts that the model was trained on. Spend some time exploring these sources to better understand the process and the training data:

1. Jill Walker Rettberg – exploration of training data for GPT-3
2. What are large language models and how do they work?
3. How Generative AI Really Works
4. What is ChatGPT Doing and Why Does It Work?

DIFFERENCES BETWEEN CHATGPT 3 AND 4

1. GPT-4: How Is It Different from GPT-3.5?
    a. Amount of memory (25k to 3k).
    b. 40% more likely to generate factual information.
    c. More capable of reading emotions in the user.
    d. Performs better on standardized tests.
    e. Paid ChatGPT-4 now has live web browsing and other plugins, otherwise it can't browse the web. Need to change settings to access.

LINKS TO TOOLS

Spend some time exploring a few of the generative AI tools below:

- Open AI: ChatGPT: https://chat.openai.com/
- Google: Gemini: https://bard.google.com/
- Adobe: Firefly: https://firefly.adobe.com/
- DALL·E 2: https://labs.openai.com/

PROMPT-WRITING TIPS

How you write your prompts is a very important aspect of the quality of results that you get. Spend some time iterating the way you write prompts. We've also shared lots of links below that will help develop prompt writing skills.

1. Brainstorming Ideas / Articles:
    a. https://twitter.com/BrianRoemmele/status/1643032326652452864
    b. https://www.nytimes.com/2023/04/21/opinion/chatgpt-journalism.html
    c. https://www.nytimes.com/2023/05/25/technology/ai-chatbot-chatgpt-prompts.html
    d. https://www.oneusefulthing.org/p/a-guide-to-prompting-ai-for-what
    e. https://www.oneusefulthing.org/p/how-to-use-ai-to-do-practical-stuff
    f. https://prompts.chat/
    g. https://twitter.com/MushtaqBilalPhD/status/1621379333943083009
    h. https://twitter.com/MushtaqBilalPhD/status/1637715972705468417
    i. https://twitter.com/thatroblennon/status/

1610316022174683136

j. For Images:
    i. https://artificialcorner.com/youre-using-midjourney-wrong-here-s-how-to-create-better-images-than-99-of-midjourney-users-c876fbe7915e
    ii. https://letsenhance.io/blog/article/ai-text-prompt-guide/

2. Suggested Tips:
    a. It performs better when you provide it info rather than ask it for info. This is a way to avoid hallucination problems. Try "Summarize the following text:" or "Explain the following text at an 8th grade reading level:" Consider using some complex academic article abstracts to test this.
        i. GPT3.5 cannot follow a URL to get information, but it will hallucinate content and appear that it can do so if the URL is descriptive enough. You can't simply paste in a URL, you should paste the content directly in.
    b. Ask it to re-write text in different styles. "Rewrite this in the style of..."
    c. It tends to perform better when you assign it a role and give it a task and format. "Act as... to complete the task of ... and maintain the format of..." Examples here.

i. Act as a professor and write...
    1. Acting as a professor, your task is to design a syllabus for a class on data and society. Include a weekly reading list. Please use only real verifiable sources that are cited. The class should be appropriate for college freshmen.
ii. Act as a life coach...
iii. Act as a senior front-end developer
iv. Act as a nurse...
v. Act as a travel agent...

d. Ask it to generate text-to-image for other AI tools like Midjournay/DALL-E 2
e. Create an interactive choose-your-own adventure game.
    1. "Create an interactive choose-your-own adventure game about Star Trek. I will play the role of the captain. You will prompt me with multiple choice options for what actions I will take, but also allow me to give my own answers that go beyond the choices."
f. Use a temperature setting between 0.1 and 1. 1 is more creative, more likely to hallucinate, more unpredictable. 0.1 is the most stable and confident result.

3. Iterative prompting
    a. We recently drafted a chapter for an edited

collection about incorporating AI, and included a supplemental page with iterative prompting examples: http://www.jjsylvia.com/wicked-ai/

b. If you have a very long prompt generation, you may need to type "continue" to have it finish the text generation.

AI CAREER RESEARCH

Objectives:

1. Develop an understanding of the diverse ways AI is likely to be utilized in future workplaces.
2. Research a specific career path of interest and its potential interaction with AI.
3. Share your findings.

Instructions:

Part 1: Initial Research

1. Choose a specific career path you are interested in.
2. Conduct initial research on the current status of AI in that field. Some potential questions to guide your research might include:

    - How is AI currently being utilized in this field?
    - What specific tasks or roles are being automated or assisted by AI?
    - What are the benefits and potential drawbacks of

this AI integration?

Part 2: Future Forecasting

Based on your research and understanding of AI capabilities, predict how AI might further influence this career path in the next 10-20 years. Consider the following aspects:

1. What additional tasks or roles could be automated or assisted by AI?
2. What new opportunities might arise due to AI integration?
3. What challenges could professionals in this field face due to increased AI use?

Part 3: Presentation Creation

We're going to create a collaborative presentation. Open the following Google Slides deck and then add one slide. Share your findings with text and images.

Articles We Found Across the Disciplines

Libraries: Hennig, Nicole, and Daniel Pfeiffer. "A Tech Librarian Explains How to Build AI Literacy," April 24,

2023. https://www.choice360.org/libtech-insight/a-tech-librarian-explains-how-to-build-ai-literacy/.

Computer Science: Cheng, Michelle. "Coding Instructors Are Adding AI to Their Lessons—before AI Replaces Them," May 8, 2023. https://qz.com/coding-instructors-are-adding-ai-to-their-lessons-befor-1850393865.

Video: *Utilizing AI for Documentary Production – with Basil Shadid and Philip Shane*, 2023**.** https://www.youtube.com/watch?v=9dP3mYZ4FR8.

Nursing: University of Calgary. "From curiosity to care: A mindful integration of AI in nursing education," May 8, 2023. https://nursing.ucalgary.ca/news/curiosity-care-mindful-integration-ai-nursing-education

Sociology: Balmer, Andrew. "A Sociological Conversation with ChatGPT about AI Ethics, Affect and Reflexivity," May 3, 2023. https://journals.sagepub.com/doi/full/10.1177/00380385231169676

Speech: Haynes, James. **"**What ChatGPT and AI can do for speakers,**"** https://thespeakerlab.com/what-chatgpt-and-ai-can-do-for-speakers/

Psychology: Ruiz, Rebecca. "3 things to know before talking to ChatGPT about your mental health" KJanuary 30, 2023, https://mashable.com/article/how-to-chat-with-chatgpt-mental-health-therapy

Education: Heaven, Will Douglas "ChatGPT is going to change education, not destroy it." April 6, 2023.

https://www.technologyreview.com/2023/04/06/1071059/chatgpt-change-not-destroy-education-openai/

Business: Matt Symonds. "More Diversity and Opportunity, Less Trigonometry – The Future of Graduate Management Education and the GMAT." April 27, 2023. https://www.forbes.com/sites/mattsymonds/2023/04/27/more-diversity-and-opportunity-less-trigonometrythe-future-of-graduate-management-education-and-the-gmat/?sh=22ca76914438

Mathematics: Ferlazzo, Larry. "How Teachers Are Using Artificial Intelligence in Classes Today," May 2, 2023. https://www.edweek.org/technology/opinion-how-teachers-are-using-artificial-intelligence-in-classes-today/2023/05

Journalism: Manjoo, Farhad. "ChatGPT Is Already Changing How I Do My Job," NY Times. April 21, 2023 https://www.nytimes.com/2023/04/21/opinion/chatgpt-journalism.html

Journalism: Carlson, Nicholas. "My editor's note to the newsroom on AI: Let's think of it like a 'bicycle of the mind," Business Insider. April 13, 2023 https://www.businessinsider.com/how-insider-newsroom-will-use-ai-2023-4

HANDS-ON PROJECT

Social Media Campaign

For this activity, you're going to create a social media campaign: Social Media Assignment

Essay Reflection

Find an essay or other piece of writing that you've previously created for an assignment in school. Share it with ChatGPT and then ask ChatGPT to comment on the quality of the writing of your essay and paste the essay in the prompt. Think about how much you agree or disagree with its comments and ask several follow-up questions that ask for specific examples from your essay. Try asking it to make suggestions about sources to use or ways to rewrite your work, but remember that ChatGPT isn't there to "fix" your writing. Think critically about what ChatGPT values in writing and what you value.

Example: https://chat.openai.com/share/651420bc-8662-4623-aaa1-02610eeeac63

Accompanying Questions

- Comment on the quality of the writing of this student essay: PASTED THE ESSAY HERE.
- Can you provide specific examples from the student essay where the style was weak and the points were not well supported by evidence?
- Why do you think the student's introduction and conclusion were weak? Provide examples from the essay in your response.
- Can you rewrite some of the most boring sentences in the essay in a more engaging way?
- Can you recommend some sources that this student could use to help them develop their essay? Include links when available.
- What are some counterarguments or oversights a critical reader could have of this student's essay's position?
- What are the strongest points the essay makes?

DISCUSSION OR REFLECTION QUESTIONS

Article for discussion kick-off:

Here are the top skills you will need for an 'A.I.-powered future,' according to new Microsoft data

Discussion:

1. What are the potential advantages and disadvantages of using AI tools, such as generative AI, to assist with school work?
2. If a student uses an AI tool to write an essay or complete a project, who should receive credit for the work – the student, the AI, or both? Why?
3. Can the use of AI tools for academic tasks be considered a form of cheating? Why or why not?
4. How might the use of AI tools for academic work influence a student's learning process and development of critical thinking skills?

5. In what ways might the use of AI tools for academic work affect the teacher-student relationship and academic evaluation processes?
6. Do students have a responsibility to disclose when they've used AI tools for school work? Why or why not?
7. What guidelines or policies could schools implement to govern the use of AI tools in academic work?
8. How does the use of AI tools in school work raise questions about the nature and purpose of education?
9. How might socioeconomic disparities in access to AI tools affect academic fairness and equity?
10. How can the educational sector ensure that the use of AI tools aligns with academic integrity principles and promotes equitable educational outcomes?
11. How do you think AI should be used in the classroom? For assignments? What would you want your teachers and future professors to know about AI?

LANGUAGE, DIVERSITY, INCLUSIVITY, AND CHAPTGPT

Guiding Questions

1. Knowing what ChatGPT is trained on (search engine crawl, ebooks, reddit, and wikipedia), what kinds of cultural concepts or groups might not be included?
    - What about oral languages, since less than 10% of human languages are written?
    - What about non-standard inscription media like the Benin bronzes, Incan quipu, or Maori carvings?
    - Is a translation ever an accurate representation of the original?

2. What languages do you speak and what have you noticed about moving back and forth from those languages?
    - "How To Speak Bad English" (8:20-13:40) podcast episode on Global English and accent reduction. A major point here is that more English speakers are

"nonnative" than "native" so "native" speakers need to adjust their expectations on what "clear communication" is.

- Visualizing the Most Used Languages on the Internet

Discussion Questions:

For the group discussing "ChatGPT threatens language diversity":

1. How does the AI respond to prompts in non-English languages?
2. Does the AI show any bias towards English language or syntax when generating responses?
3. Try typing a sentence with non-English syntax in English. How does the AI respond?

For the group discussing "ChatGPT is multilingual but monocultural":

1. Generate a story set in a non-Western culture. Does the AI accurately and respectfully incorporate elements of

that culture?

2. How does the AI respond to prompts containing cultural idioms, references, or concepts?
3. Look up some common phrases or idioms in less commonly used languages. How does the AI respond to these prompts?

For the group discussing "Proper English and normative grading practices":

1. Try typing sentences in various English dialects or accents (e.g., African American Vernacular English, Singlish, Hinglish). How does the AI respond?
2. Does the AI seem to favor a particular type of English in its responses?
3. How does ChatGPT's answer to your question change as you rephrase the same question (ie using "Black" rather than "African American") Does it perpetuate stereotypes or exhibit biases?

Links to Useful articles on this

Summary points

a. "Unmasking AI Harms and Biases".
b. "ChatGPT threatens language diversity in the age of AI"
    i. With white male voices authoring the majority of the training material, the default voice replicates those language patterns.
c. "ChatGPT is multilingual but monocultural, and it's learning your values"
    i. Diversity is not just in languages and dialects used but in the cultural beliefs and ideologies embedded in the training material.
d. "ChatGPT & Writing in the Secondary ELA Classroom"
    i. Our normative grading practices around "proper English" encourage students to mask language diversity.
e. "OpenAI's Linguistic Diversity Initiatives in AI Language Testing"
    i. OpenAI's proposed solutions focus more on including less common languages but say much less about how to addressing race and gender stereotypes in language use.

POST-TEST AND SURVEY

Please complete the following Post-Test and Survey in 30 minutes or less.

An interactive H5P element has been excluded from this version of the text. You can view it online here: https://rotel.pressbooks.pub/datarenaissance/?p=74#h5p-3

WRAP UP

Key Takeaways

- Generative AI models like ChatGPT-4 have evolved significantly in their capabilities, including better factual accuracy and emotional understanding, but they come with limitations such as the inability to browse the web unless specific settings are enabled.
- The use of AI tools in academic and professional settings raises important ethical questions around authorship, academic integrity, and the potential for perpetuating biases or excluding certain cultural perspectives.
- Prompt-writing techniques greatly influence the quality of outputs from generative AI;

understanding how to effectively structure prompts can lead to more accurate and useful responses.

- The role of AI in various career paths is not just an imminent future but a present reality, necessitating research and ethical considerations about how AI will shape and be shaped by professional practices.

Exercises

1. How do you think the use of AI tools like ChatGPT could affect language diversity and inclusivity? Consider ChatGPT's training data and its implications for representing various cultures and languages.
2. Using the generative AI tools listed in the "Links to Tools" section, each student is

tasked with creating a piece of content (it could be text, art, or any form of digital media). Afterward, discuss as a class the ethical considerations you had to make while using these tools. Were you concerned about the originality of your work, the biases in the AI, or other ethical issues?

3. Write a prompt for ChatGPT that aims to generate a summary of a complex academic article. Evaluate the accuracy and clarity of the AI-generated summary. What does this exercise reveal about the strengths and limitations of using AI for academic purposes?

FURTHER READING

"AI Text Generators: Sources to Stimulate Discussion among Teachers." https://docs.google.com/document/u/0/d/1V1drRG1XlWTBrEwgGqd-cCySUB12JrcoamB5i16-Ezw/mobilebasic.https://www.theatlantic.com/technology/archive/2023/02/chatgpt-ai-detector-machine-learning-technology-bureaucracy/672927/

Fyfe, Paul. "How to Cheat on Your Final Paper: Assigning AI for Student Writing." *AI & SOCIETY*, March 10, 2022. https://doi.org/10.1007/s00146-022-01397-z.

Bogost, Ian. "ChatGPT Is About to Dump More Work on Everyone." The Atlantic, February 2, 2023. https://www.theatlantic.com/technology/archive/2023/02/chatgpt-ai-detector-machine-learning-technology-bureaucracy/672927/.

Hoffman, Reid, and ChatGPT. *Impromptu*. 2023. https://www.impromptubook.com/wp-content/uploads/2023/03/impromptu-rh.pdf.

Knight, Will. "These ChatGPT Rivals Are Designed to Play with Your Emotions." May 4, 2023. https://www.wired.com/story/fast-forward-chatgpt-rivals-emotions/

Ray, Augie. "What Does It Mean If AI Can Convey More Empathy Than Humans?," 2023. https://www.linkedin.com/pulse/what-does-mean-ai-can-convey-more-empathy-than-humans-augie-ray.

Sentient Syllabus Project. "Syllabus Resources." Google Docs, 2023. https://docs.google.com/document/d/1O1_uUvF8OYbleru5QyjjuNP_On7h5vaVQC2GaSQ315U/edit?usp=sharing&usp=embed_facebook.

"Sparks of Artificial General Intelligence: Early Experiments with GPT-4," https://arxiv.org/pdf/2303.12712.pdf

The New York Times. "Hard Fork." https://www.nytimes.com/column/hard-fork.

UNESCO. "ChatGPT and Artificial Intelligence in Higher Education: Quick Start Guide," 2023. https://www.iesalc.unesco.org/wp-content/uploads/2023/04/ChatGPT-and-Artificial-Intelligence-in-higher-education-Quick-Start-guide_EN_FINAL.pdf.

Wong, Matteo. "ChatGPT is Already Obsolete," May 19, 2023. https://www.theatlantic.com/technology/archive/2023/05/ai-advancements-multimodal-models/674113/.

PART III
CASE STUDY: "IT'S PERFECT, FOUR STARS!"

Chapter Written by Leonora Shell, M.A.T.[1]

Editor's Note: This chapter is written from a first-person perspective by a woman who owns a business that sells products on Etsy. It is intended to highlight her personal experience as a woman.

Learning Objectives

1. This chapter was written based on a guest lecture that was given to both the undergraduate and ALFA sections of the Data & Society course. In the spirit of full disclosure, Leonora Shell is the wife of volume editor J.J. Sylvia IV. To avoid conflict of interest, the shop referenced has not been disclosed, either here or in the guest lectures.

- Critically analyze the impact of review systems on e-commerce platforms, particularly how they affect small business owners and vulnerable populations.
- Gain an understanding of how algorithms and automated systems can both support and hinder the operation of online marketplaces, influencing the livelihoods of the sellers involved.
- Develop the ability to evaluate the ethical implications of review and rating systems in digital commerce.
- Analyze the effects of review systems on power dynamics and societal inequalities through a feminist lens.

INTRODUCTION

Black Mirror's 2016 episode, "Nosedive" explores a dystopian future in which the protagonist's life is ruined as she accidentally lowers her overall personal rating over the course of a very bad day. The idea of a personal rating is already starting to take shape in the form of China's social credit system, but even in the U.S., we have an analogous system of rating and ranking, even if it hasn't been centralized in one consolidated place.

The effects of starred ratings and reviews for consumer products have been heralded as a way to create an objective, quantifiable method for assessing the quality of a product or service (Gunasekaran, 2019). On the surface, this seems to be true, a way to summarize a consumer experience using a simple five starred approach, ranging from five stars meaning you loved it, to one star being "disappointed." More often than not, however, these ratings are not about the particular good or service, but more about the mismanagement of expectations by the consumer (Peak Performance Digital, n.d.). Furthermore, negative ratings are often unaccompanied by any sort of relevant commentary or a way for a company or individual seller to improve. As more women enter the space of e-commerce and business, the reviews have taken on more

sexist and harmful tones as well as the introduction of AI or automated bots that crawl sites and take down a seller's listings without warning or an effective way to counter the decision that didn't involve a human's judgment at all.

HUMAN COMMERCE

After over ten years of selling handmade products on an e-commerce site specifically designed for handmade goods, one that touts the importance of keeping commerce human, while continually and methodically removing any empathy from reviews or oversight from actual human beings, it is clear that a linchpin moment in the change to a less qualitative and human experience in the marketplace was the transition to starred reviews by customers in late 2013. With this, there was also the removal of the option to rate or review customers by a seller.

As a business owner writing this overview of the impact of starred reviews, I wanted to share my insights as an individual shop being rated on an ever-changing platform in an economy that not only demands constant growth but also perfection. I am a professional, and negative reviews are generally a place to learn and grow; they certainly don't bother me like they used to — as of this writing I have over 23,000 sales, over 3,300 reviews and over 10,000 followers on social media, a community built organically over a decade around our brand. As a result of our success as a brand, we have never had to take on outside funding, loans or debts. We have been extremely lucky and grateful for the tools that we have used to get us where we are today. So much of what we do, sell, and market

has success or failure based on algorithms created by other publicly-traded companies. Since this is my current full-time employment, I will not be using the name of the company in this chapter, as it is against their current user policies to do so in a negative (albeit truthful) way.

LIVING AND DYING BY THE ALGORITHMS

So, how does the algorithm work? This is the great question with e-commerce and online marketing across various social media, search, and commerce platforms. Since the handmade e-commerce site in question went public in 2015, the priorities of the company change after each quarterly board meeting. There is a cycle and culture of pushing for ways to endlessly improve and be tinkered with — AI and bots crawl the site and can remove and shut down shops with little to no warning. Trends can be decided upon and implemented based on keyword searches of a couple hundred individuals, the lists of celebrity influencers, a feature in an advertisement or a selection by an employee of the company.

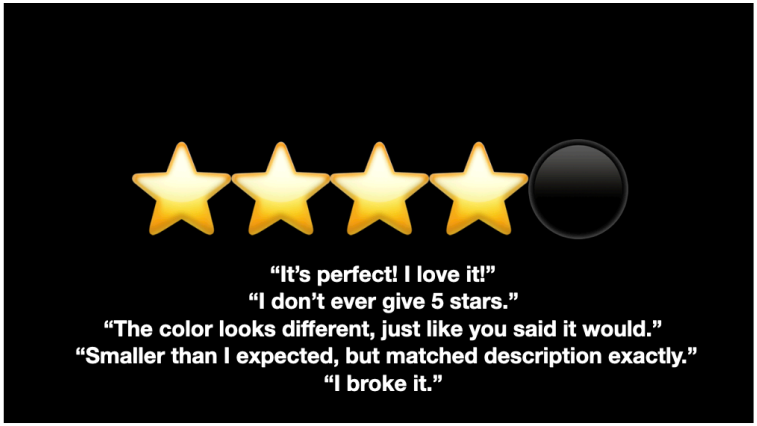
We are here to focus on the reviews, however, as no marketplace that exists is perfect for consumers and business owners. In essence, the higher a given shop's reviews, the more frequently their products appear in search results, thus resulting in more sales (Collinger and Malthouse, 2015). This is good, right? Because as a consumer, you would want only the best products shown to you when you search for them rather than the worst or least popular — but what if that popularity was artificially downgraded for smaller sellers and

falsely upgraded for drop-shippers or sellers that weren't hand-making their products? The site attempted to remedy any customer dissatisfaction with a mathematical formula, the details of which were hidden from both sellers and buyers, called the the Order Dissatisfaction Rating (ODR) (Glassenberg, 2020). This ODR included the amount of customer complaint cases brought against a seller and shop ratings, among other metrics. If your ODR rose to an amount higher than was defined as acceptable (at the time less than 1% of reviews could be 1 or 2 stars, over the course of 90 days), your shop would be warned, closed and/or you would be suspended from using the site. Unfortunately, only about 10% of purchasers review their orders, and this metric only took into account the reviews that were made, not the overall percentage of orders fulfilled with satisfied customers.

The ODR approach was ended in 2020 because so many shop dissatisfaction ratings increased after customers were unhappy with the disruptions to shipping and the overall logistics of the planet's supply chain due to the COVID-19 pandemic. Those customers that felt out of control took it out on sellers in the form of low reviews. In 2021, a program was created called the "Star Seller Program," which was a way to display the ODR to consumers and sellers. Previously, this ODR metric was only accessed with a non-navigable link which was made available on a Reddit post or discussion boards hosted by other organizations under a now-deleted section of the site called "customer service performance." The

link is now broken. This new program began in 2021 and required that 95% of a shop's reviews over 90 days be five star reviews, meaning that for every 4 star review a shop needed 19 five star reviews to maintain their standing as a star seller. If a shop only receives reviews on 10% of their sales, a single 4-star review would require 190 additional sales with exclusively five star reviews to recover their star seller standing. In 2022, after considerable seller feedback, this policy was remedied so that a shop had to maintain a 4.8 average star rating over 90 days, significantly increasing the ability for shops to be a part of the program that sets them apart from other sellers.

THE "FAULT" IN OUR STARS



Some favorite four star reviews over the last ten years.

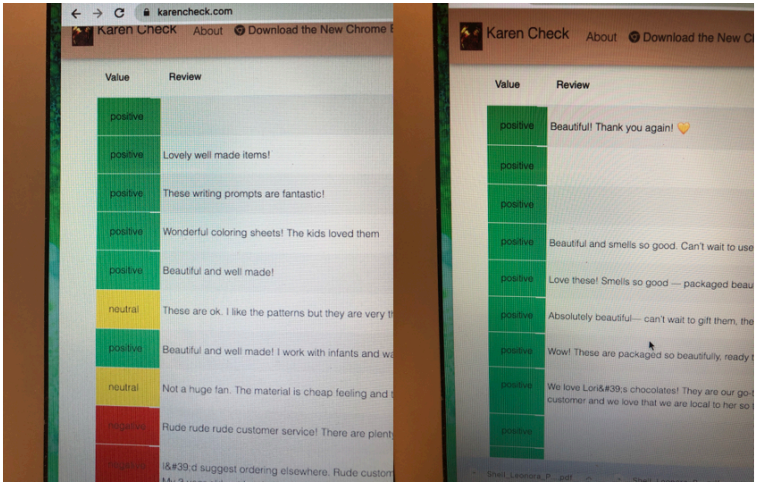
Figure 1. The image features four stars and several quotes from user reviews. These read: "It's perfect! I love it!", "I don't ever give 5 stars.", The color looks different, just like you said it would.", "Smaller than expected, but matched description exactly.", and "I broke it."

Let's take a brief aside to evaluate four star ratings. Some of my most frustrating reviews over the past decade have been four star reviews. After a brief analysis of these four star reviews, almost half of those our shop received contained the words "perfect" and "love." When kindly questioned if there was

anything we could have done to help us get a perfect score, several customers stated that they simply don't give five star reviews to any products ever. I have stopped asking this question as some consumers retaliate and end up lowering their initial review, citing that all we cared about was a review. Other customers state that the color is different than pictured, just like it was clearly stated in the listing and/or as the customer requested a custom color or variation... and so it did, indeed, vary. Maybe the item was smaller than they expected, however it matched the description exactly, and they loved it. Or, more maddeningly, they broke it and didn't let us know or refused to take a replacement despite our robust policies stating that if they break an item we will replace it, no questions asked. There is no policy, kindness, or gesture that can remedy these situations, and that's just how business works sometimes.

WHO RATES THE RATER?

Historically, sellers on the handmade e-commerce site were able to leave tiered reviews for customers that were negative, neutral or positive; however, with the establishment of starred reviews, this ability was removed by the site and later replaced with only the ability to block buyers rather than provide any feedback. A site that was helpful when making decisions about how to engage with a customer was called "KarenCheck.com" — this site has now been hobbled, as it has been blocked from accessing reviews connected with usernames on the e-commerce site in question. However, for a time it was useful to see the history and types of reviews left by a customer to check their expectations and ways of communication they were most likely to positively respond.



Screenshots of KarenCheck.com before it was effectively banned by the marketplace.

Figure 2. Screenshots of KarenCheck.com before it was effectively banned by the marketplace. The image shows one reviewer with a mix of positive, neutral, and negative reviews, and another reviewer with all positive reviews.

Additionally, in 2021, social media sites such as TikTok encouraged customers to scam sellers to get free items by leaving negative reviews and demanding returns, only to return trash from their homes or something else to the seller. Our small shop personally received a box of hair and another customer used glittered paper to fill their returned package, causing us to have to do an additional round of completely sterilizing and cleaning our workshop.

THE SHIFT TO "OBJECTIVE" STARS

So, where did all of this start? In early 2013, the reviews that customers could leave featured three options: "negative," "neutral," or "positive." In late 2013, these values were switched to a starred approach, with negative reviews being translated to one or two stars, one star meaning "disappointed" and two stars meaning "Not a fan." Neutral translated to three and four star ratings meaning "It's okay." and "Like it," respectively. Lastly, positive reviews were translated to five stars, meaning "love it." These words pop up when a user hovers over the number of stars to select which one best describes their shopping experience. The resulting "experiences" for the customer shifted from a qualitative, subjective, generic feedback of experience approach for leaving product reviews to a quantitative, objective, specific and rating/grade approach. The shift from seller or overall shop review to the review of an individual item, which varies, and by the nature of the marketplace, should be handmade and individually created for each customer. This method is not only dehumanizing for the seller, but gives the buyer a significant amount of leverage in the future of a shop or item in a given marketplace because of the way it impacts the

algorithms that determine which items are featured in searches.

Slide from presentation given by the author on this topic.

Figure 3. Slide from presentation given by author. It features the differences in the rating system from early 2013 to late 2013, noting changes from qualitative to quantitative, subjective to object, generic to specific, and feedback/experience to rating/grade.

All of this in itself could be considered fine; it is a way to create a seemingly equitable marketplace for sellers with a transparent review system allowing freedom of expression of contentment or discontentment with a particular item or service. This could absolutely be said about a marketplace wherein each seller and buyer represent the same demographics, but not one where the sellers and buyers represent populations with historically very different societal power and autonomy.

WHO ARE WE RATING?

So, who are we rating anyway? Why do we care about stars versus negative/neutral/positive experience metrics? Of course, we need some way to assess sellers and their products, but I argue that this metric unequally impacts the most vulnerable business owners. As of the most recent report of this U.S. marketplace at the time of this writing, 86% of sellers on this platform identified as female (Drah, 2021). These are employment opportunities for women to be business and micro-business owners that are self-made. They are not involved in multi-level marketing schemes or employed by someone else; these are ways women can make their own hours and own their own businesses. For many years, this platform encouraged its users to quit their day jobs and go full time with their craft. These sellers are also twice as likely to be under the age of 35, with a median age of 39. This represents a lot of work-at-home parents, those between more stable employment, or part-time workers. Aligned with the national income average, 17% of sellers make less than $25K per household. Additionally, 97% of these shops are run from home. On the other side of things, 70% of buyers identify as female. However, the site has announced that they are

encouraging and focusing on bringing more men to the platform in 2023 (Ryan, 2022).

According to Pew research (Smith and Anderson, 2016), men are more likely to read and leave reviews and younger consumers are more likely to leave reviews. Anecdotally, those men have been more critical, citing that they never leave five star reviews, are unwilling to change a review based on new information, or even accept a refund. This gives me pause, as this means that the site is bringing a potentially more critical population to rate and review predominantly young women, who are historically underemployed and underrepresented in business ownership opportunities and spaces (Lake, 2023). And, despite women nearly reaching gender parity in 2023 in business education programs, they still aren't compensated or funded equally — in most cases making half of their male counterparts (Arora, 2020).

CONCLUSION

So, what are we to do in a culture driven by capitalism? According to Cory Doctorow (2023), the concept of "enshittification" explains the market forces that encourage a platform to cater to different strata of the population over time. The platform is successful in keeping each group happy and dependent on the platform until they shift methods and gears to bring in another group in order to grow, ultimately frustrating everyone who was already there, leaving those initial users who were early adopters of the platform and brought it to existence and initial success, in the dust. This enshittification applies to the handmade marketplace in question, leaving its original handmade sellers and early devotees in the throes of quirky algorithms and shifting priorities after almost each quarterly shareholder meeting. This cautionary tale leaves everyone to consider, and reconsider, their place in a business world where various platforms vie for their and others' time, attention, and most importantly, money. Each year, an additional small percentage is added to the amount of each sale that is claimed by the platform, with less and less value-add. This extra cost burden gets passed on to the consumer, or more likely, the seller to stay competitive.

Now, if I get a negative review, I strive to respond in a more productive way. I leave good reviews for places I love – small businesses that are making it work despite so many pressures to close up shop and go work for someone else, so many pressures to abandon their own good ideas and lives and devote them instead to making more millionaires become billionaires. Five stars go to the small restaurants in our neighborhood that I want to keep open. When people feel out of control of their lives, they receive bad news, or something sad and out of their control happens, they attempt to take some control in some way, and sometimes that's leaving a negative review. In my case, it's to leave a positive one. It's a very human thing to do, it turns out. A way to keep commerce human.

I remind myself that this one e-commerce platform is just that: it's just one place that I choose to spend my time, ideas, energy, and money. Over the last decade, I've created several supportive environments outside of the e-commerce site that allow connection with customers and positive experiences with families using our handmade products. I won't let the design of the platform or the whims of the shareholders detract from my positive experience. To do that is to lose what remaining joy I have and bring to what I make and share with others. And no bad review can take that away from me. A review is not my identity, it is not me. I am the only one who can manage my own expectations and I encourage you to do the same the next time you are asked to leave a review.

WRAP UP

Key Takeaways

- The shift from qualitative to quantitative review systems in e-commerce platforms has had a profound impact on small business owners, often amplifying the power imbalance between sellers and consumers.
- Algorithms and automated systems, while designed to improve marketplace efficiency, can inadvertently penalize sellers through non-transparent metrics and sudden policy changes.
- The demographic makeup of a platform's user base can influence the nature and tone of reviews, with evidence suggesting that

women and younger business owners might be disproportionately affected.

- The concept of "enshittification" encapsulates the risk that platforms, in their quest for growth, can alienate their original user base, undermining the very communities that contributed to their initial success.

Exercises

1. Consider the ethical implications of relying solely on algorithms to manage reviews and seller standings in an e-commerce platform. How can these systems be improved to account for the human element in commerce?

2. Explore the concept of "enshittification" in the context of other online platforms or services you have used. Can you identify any instances

where a platform's changes have alienated its original user base? Discuss the long-term sustainability of such strategies.

3. Given the gender and age demographics outlined in the text, analyze how these might interact with the review and rating systems to create a potentially biased marketplace. What steps could platforms take to mitigate these biases?

REFERENCES

Arora, Rohit. "Why Male Entrepreneurs in the US Make Double Their Female Counterparts." CNBC, March 10, 2020. https://www.cnbc.com/2020/03/10/why-male-entrepreneurs-in-the-us-make-double-their-female-counterparts.html.

Collinger, Tom, and Edward C. Malthouse. "From Reviews to Revenue Volume 1: How Star Ratings and Review Content Influence Purchase." PowerReviews. Northwestern University, 2015. https://www.powerreviews.com/wp-content/uploads/2019/02/From-Reviews-to-Revenue-Northwestern-Report-Volume-1.pdf.

Doctorow, Cory. "Pluralistic: Tiktok's Enshittification (21 Jan 2023) – Pluralistic: Daily Links from Cory Doctorow." *Pluralistic: Daily Links from Cory Doctorow* (blog), February 27, 2023. https://pluralistic.net/2023/01/21/potemkin-ai/.

Drah, Hermina. "19 Etsy Statistics for a Profitable 2023." *Why Does Everything Suck?* (blog), June 10, 2021. http://whydoeseverythingsuck.net/blog/etsy-statistics/.

Glassenberg, Abby. "ODR on Etsy: What Sellers Need to Know." Craft Industry Alliance, February 18, 2020.

https://craftindustryalliance.org/odr-on-etsy-what-sellers-need-to-know/.

Gunasekaran, Ajay. "Why Star Ratings Don't Make Sense for All Product & Services." Medium, 2019. https://uxdesign.cc/why-star-ratings-dont-make-sense-for-all-product-services-908e3b3901db.

Lake, Sydney. "MBA Programs Are Nearly Reaching Gender Parity with More than 41% Women Enrollment." Fortune, January 27, 2023. https://fortune.com/education/articles/mba-programs-are-nearly-reaching-gender-parity-with-more-than-41-women-enrollment/.

"Nosedive." Streaming. *Black Mirror*. Netflix, 2016.

Peak Performance Digital. "5 Stars or Bust: Why the Star Rating System for Reviews Is Broken – Peak Performance Digital," n.d. https://peakperformancedigital.com/5-stars-or-bust-why-the-star-rating-system-for-reviews-is-broken/.

Ryan, Tom. "Has Etsy Been Missing the Boat with Men?" *RetailWire* (blog), May 9, 2022. https://retailwire.com/discussion/has-etsy-been-missing-the-boat-with-men/.

Smith, Aaron, and Monica Anderson. "Online Shopping and E-Commerce: 2. Online Reviews." *Pew Research Center* (blog), December 19, 2016. https://www.pewresearch.org/internet/2016/12/19/online-reviews/.

PART IV
MEDIA AND DATA LITERACY

Figure 1. Media & Data Literacy. Media and data is being shown with pencils and paper.

Chapter Written by Henry Christiansen

Learning Objectives

- Critically evaluate various forms of media and data, identifying biases and ethical implications inherent in them.
- Gain a comprehensive understanding of both media literacy and data literacy, and how these two fields intersect in various aspects like race, gender, and social class.
- Understand the challenges and opportunities in media and data literacy education, and be equipped with strategies for incorporating these literacies into different learning environments and career paths.

INTRODUCTION

Currently, media literacy and data literacy are still relatively new and evolving fields. It's important to acknowledge that access to these forms of literacy is not uniform across all communities. Systemic barriers, such as socio-economic status, educational opportunities, and cultural factors, often create a gap in media and data literacy skills. It requires a lot of work to ensure that people of all ages and backgrounds have the skills and knowledge they need to navigate the complex world of media and data effectively. In this chapter, I explain the similarities and differences between these approaches and explore how they are currently being taught.

MEDIA LITERACY

Media literacy is accessing, analyzing, evaluating, and creating media messages in various forms, including print, audio, video, and digital content (*Media Literacy Defined,* n.d.). Media literacy is more important than ever in today's digital age, where information is readily available through various media channels. It's crucial to note that the media often portray different races, genders, and social classes in stereotypical or biased ways. Media literacy equips individuals with the tools to critically analyze these portrayals, questioning their origins and implications. Additionally, language can be a significant barrier in understanding and interpreting media content. Multilingual media literacy programs can help bridge this gap, making media literacy more accessible to people who speak different languages. Furthermore, it's essential for media literacy education to be culturally sensitive and inclusive. Educators should take into account the diverse backgrounds of learners to ensure that media literacy is not just a skill but a tool for social inclusion.

By being media literate, individuals can become more informed and responsible media consumers, able to examine and assess media messages and their sources critically. Media literacy can empower individuals in many ways. For instance,

it can help them distinguish between fact and fiction, identify bias and propaganda, and recognize manipulative techniques used in media messages.

It can also enable them to understand better the cultural, social, and political contexts in which media messages are produced and consumed. Furthermore, media literacy can foster creativity and innovation by allowing individuals to express their ideas and perspectives through various media forms, such as writing, photography, video production, and digital media. In short, media literacy is a crucial skill enabling individuals to navigate the complex media landscape and make informed decisions about the media content they consume and create. By developing media literacy skills, individuals can become active and engaged media citizens, capable of participating in the media discourse and shaping the media culture.

CHALLENGES WITH MEDIA LITERACY

Some key challenges that need to be addressed include a lack of awareness. Many people need to be made aware of what media literacy is and why it is essential. There is a need for more education and outreach to help people understand the value of media literacy. There is a digital divide between those with access to technology and those without access (Taylor, 2022). This can limit the ability of some individuals to develop media literacy skills and can exacerbate existing inequalities. With the rise of fake news and misinformation, it has become increasingly challenging for people to distinguish fact from fiction. There is a need for more emphasis on critical thinking and fact-checking skills.

Several steps can be taken to promote media literacy. We need to ensure that media literacy is taught in schools and that students have the opportunity to develop these skills from a young age. Media literacy is not just the responsibility of educators but also of media professionals, policymakers, and parents. Collaboration between these groups can help ensure that everyone has access to accurate information and is equipped to navigate the media landscape. With the rapid pace of technological change, we need to be innovative in our

approach to media literacy. This may involve the use of new technologies, such as virtual and augmented reality, to help people develop media literacy skills engagingly and interactively. In summary, media literacy is an essential skill for navigating the modern media landscape effectively.

While there are still many challenges to overcome, there are also many opportunities to promote media literacy through education, collaboration, and innovation. Media literacy and data literacy are connected in that they both involve critical thinking skills. In order to effectively analyze media content or interpret data, one must be able to ask questions, evaluate sources, and think critically about the information presented. Additionally, both skills require an understanding of how information is created, disseminated, and consumed in today's digital world.

DATA LITERACY

What is data literacy? There are many resources that explain data literacy, but one person who explains data literacy well is Tim Stobierski. Mr. Stobierski is a marketing specialist and contributing writer for Harvard Business School Online, who writes about data literacy. His compelling 2021 article explains: "Data literacy is a term used to describe an individual's ability to read, understand, and utilize data in different ways. It doesn't require an individual to be an expert—as a data scientist or analyst might be considered—but rather to show an understanding of basic concepts, such as different types of data, Common data sources, Types of analysis, Data Hygiene, Tools, Techniques, and Frameworks." (p. 1).

Data Literacy is increasingly important today. It's crucial to discuss the ethical implications of data collection, especially how it can disproportionately affect marginalized communities. Additionally, the potential for bias in data should not be overlooked. A lack of diversity in data science can perpetuate systemic inequalities, making it essential to address this issue in data literacy education. These issues are discussed in Chapter 1 as well as articles included in the appendix. Organizations and individuals are inundated with

vast amounts of data. It empowers individuals to make informed decisions, identify trends, solve problems, and effectively communicate insights derived from data. Data can empower organizations and individuals to share information and collaborate. Data is to be shared and explored by individuals to gain skills and knowledge. ("The 2020 Global State of Enterprise Analytics", 2020)

With data being consumed by many people and platforms, individuals need to understand and analyze the data they encounter. Data literacy is not just about understanding numbers and statistics but also about being able to interpret and communicate the insights derived from the data. It involves understanding the context in which the data was collected, the biases that may be present, and how to use data to make informed decisions. In today's digital age, data is everywhere and being data literate is essential for success in many industries. Data literacy also involves understanding the ethical implications of collecting, analyzing, and using data and the importance of privacy and security. Overall, data literacy is an important skill for anyone who wants to be able to understand, analyze, and communicate insights derived from data.

SIMILARITIES

Media literacy and data literacy are related in many different ways. Media literacy looks at media and how it influences our reality. Data literacy is the understanding of how platforms like social media apps and beyond interact with society. It is essential to know that much of the information we encounter on a daily basis can be misleading. A better understanding of media and data literacy can help someone navigate through various media. Many sources out there are reliable, but when it comes to knowing which sources are reliable, it can become tricky. That is why it is essential to get a better understanding of media and data literacy. It's important to recognize the intersectionality of media and data literacy with issues of race, gender, and social class. Understanding how these literacies intersect with broader social issues can provide a more holistic approach to media and data literacy. Knowing even just a little about both can greatly improve one's knowledge of the media we all consume today.

UNDERSTANDING DATA LITERACY SKILLS

There are many benefits to developing data literacy skills, including the ability to identify patterns and trends, make data-driven decisions, and communicate insights effectively. Data literacy is becoming increasingly important in many fields, including business, healthcare, and education, as more and more organizations are relying on data to make informed decisions. For example, in healthcare, data literacy can empower community health workers in underserved areas to better understand and address the specific health needs of their communities. In education, teachers in diverse classrooms can use data literacy to tailor their teaching methods to better serve students from various cultural and linguistic backgrounds. With the growing importance of data, there is also a growing demand for individuals with strong data literacy skills.

Inclusive case studies featuring a diverse range of individuals and communities can further enrich the understanding of data literacy. These case studies can serve as practical examples that resonate with a broader audience. With strong data literacy skills, there are many career opportunities an individual can have. One example of such a career is a marketing analyst. Marketing analysts identify customer behavior, measure

marketing campaign effectiveness, and optimize marketing strategies. Data literacy enables them to conduct accurate data analysis, segment audiences, track key metrics, and make data-driven recommendations for marketing decisions. Another job would be a data analyst. Data analysts play a crucial role in collecting, analyzing, and interpreting data to provide insights and support decision-making. With data literacy skills, they can effectively manipulate and analyze data, develop meaningful visualizations, and communicate data-driven findings to stakeholders.

Overall, data literacy is an essential skill for navigating the complex data landscape and making informed decisions. By developing data literacy skills, individuals can improve their understanding of data and use it to drive positive change in their organizations and communities.

In addition to the benefits of data literacy for individuals, there are also broader societal benefits. A more data-literate society can lead to better-informed decisions and policies, improved public health, and more effective and efficient use of resources. Data literacy can also address issues of inequality and social justice by providing insights into patterns and trends that may be affecting marginalized communities.

However, there are also challenges associated with data literacy, such as the potential for bias in data collection and analysis, the difficulty of interpreting complex data, and the risk of misinterpreting or misusing data. Individuals need to approach data with a critical mindset, be aware of the

limitations of the data, and seek out diverse perspectives when analyzing and interpreting data.

Data literacy is not a static skill set but a constantly evolving one. As new technologies and data sources emerge, individuals must be willing to adapt and continue learning to remain data literate. Today data literacy is becoming an increasingly important skill for success in many fields and industries. It will continue to be essential for individuals and organizations to stay ahead of the curve.

MEDIA LITERACY AND DATA LITERACY SKILLS

Developing media and data literacy skills is crucial for navigating the complex media landscape and making sense of the vast amounts of available data. When learning about media and data literacy, people need to understand how they work together better. I argue that teaching media and data literacy should be taught so the student can learn how to navigate through them without concrete courses devoted completely to that topic. Instead, already existing courses can assign work that can help improve one's knowledge of media and data literacy. The student should be taught the overall understanding of media and data literacy so they can understand how to make decisions and identify misinformation.

To effectively teach media literacy and data literacy in schools in Massachusetts, teachers should have small assignments that help build media literacy and data literacy. Students with no knowledge of media literacy and data literacy would gain more information on the topics due to the small workload that is provided. In the process of these assignments, students and teachers can be more understanding and share their knowledge to further gain information. Massachusetts

has also passed a bill that makes education for media literacy a high school graduation requirement. This also requires the Department of Elementary and Secondary Education to develop instructional guidelines in media literacy. The Department of Education has a working group to access and recommend revisions to policies and procedures on media literacy aligning with K-12 standards. This working group will consult with experts in media literacy including but not limited to academic experts and non-profit organizations. In the development of teaching and learning media literacy and data literacy, the Department of Education will assist in resources to aid and will provide and make sure media literacy and data literacy training opportunities are available.

Educators can incorporate these skills into various subjects, such as English, social studies, science, and math. Teachers can use real-world examples to demonstrate the importance of these skills and can provide opportunities for students to analyze media messages and data sets. Additionally, schools can offer classes or workshops specifically dedicated to teaching media and data literacy and can provide access to resources and tools that allow students to practice and develop these skills.

An example of how media literacy and data literacy could be taught is by engaging students in an interactive discussion about media and data topics. One example is encouraging students to share their perspectives and asking questions and then critically analyzing different media messages and data sets. Another example of how media literacy and data literacy could

be taught is by doing a media analysis assignment. In doing a media analysis, students will evaluate and critically analyze different types of media. Further, the assignment can require students to identify the intended audience, the message it is sending, and any biases that may be presented. It is important to adapt the teaching methods to a specific audience, group, education level, and learning style of the students.

Incorporating media literacy and data literacy into the curriculum of schools in Massachusetts is essential for preparing students to be critical thinkers and responsible consumers and producers of information in our increasingly media-saturated and data-driven world and preparing students for their future career opportunities. Employers increasingly seek individuals with these skills to work in various industries, including media, marketing, and technology. Some other opportunities that focus more on data literacy are data science, data engineering, business intelligence analysis, and data journalism. These are just a few jobs that focus on their employees having data literacy skills. The demand for these roles continues to grow as organizations recognize the value of data in decision-making and innovation.

According to a 2021 report by Burning Glass Technologies (Bursin, 2021), a labor market analytics firm, media literacy skills were listed as a desired competency in job postings across a variety of fields, including journalism, public relations, advertising, marketing, and social media management. The report found that jobs requiring media literacy skills were

growing at a rate of 6.5% annually. Equipping students from underserved communities with these skills through K-12 education can help level the playing field, potentially leading to a more diverse and equitable workforce. As these communities often face systemic barriers to employment in these growing fields, early media literacy education can be a step toward economic empowerment.

Additionally, having a solid foundation in media and data literacy can empower students to analyze and interpret information from multiple sources, leading to more informed decisions in their personal and professional lives. This is particularly impactful for marginalized communities, as being better-informed consumers and citizens can lead to more equitable access to opportunities and resources. Schools should also collaborate with experts in the field and consistently update their teaching methods to ensure students remain on the cutting edge of media and data literacy education. In turn, this comprehensive education will better position students to make valuable contributions to their chosen industries and communities, potentially leading to societal benefits such as a stronger, more diverse workforce and more equitable community development.

There can be many viewpoints on how we should teach media and data literacy, and many people might think there are better ways to teach literacy than what I have argued so far. Various opinions and criticisms exist regarding how these

literacies should be taught. Below I consider a few of them and how they can be addressed.

Some argue that media literacy and data literacy are not essential skills and should not be a priority in education. They believe other subjects like math and science should take precedence. However, with the rise of fake news and misinformation, individuals need to be able to distinguish between credible and unreliable sources. When getting information from an unreliable source, it can cause the work to lose credibility. It also may cause insurrections due to the fact that misinformation or fake news is dangerous and can cause people to act violently (Hoʻoulu Staff, 2017).

Additionally, data literacy is essential for making informed decisions in various fields, including business, health, and politics. Critics argue that media literacy and data literacy are too complex for the average person to understand. They believe these literacies require specialized training and should be left to experts. While media literacy and data literacy can be complex, it is possible to teach them in a way that is accessible and understandable for the general public. Teachers can use real-world examples and hands-on activities to make these skills more tangible and relevant to students. While media and data literacy may require different approaches, they are interconnected skills. Understanding how to analyze and interpret data is critical in evaluating media sources and vice versa. Teaching these literacies in conjunction with each other

is essential to provide students with a more holistic understanding of the information.

While there may be differing opinions and criticisms regarding how media and data literacy should be taught, it is essential to recognize the importance of these skills in today's information age. Teachers can use various strategies to make these skills accessible and relevant to students, including using real-world examples.

CONCLUSION

Media literacy and data literacy are essential and provide individuals with information and knowledge. Without media literacy, individuals would struggle to think critically about media and understand what's credible. Without data literacy, individuals would struggle to make data-driven decisions and communicate insights effectively. In today's world, these skills are extremely important and are vital in navigating the complexities of media and data effectively.

WRAP UP

Key Takeaways

- Media and data literacy are not just essential skills but also tools for social inclusion and empowerment, enabling individuals to make informed decisions and engage in societal discourse.
- The digital divide and systemic barriers like socio-economic status and educational opportunities can significantly impact access to media and data literacy, making education and outreach critical.
- Both media and data literacy are evolving fields that require ongoing education to keep pace with technological advancements;

innovative tools like virtual and augmented reality can enhance this educational process.

- Teaching media and data literacy is not just the responsibility of schools; it requires a multi-faceted approach involving educators, media professionals, policymakers, and parents to be truly effective.

Exercises

1. How do systemic barriers like socio-economic status, educational opportunities, and cultural factors impact access to media and data literacy in your community? Discuss specific examples.
2. Conduct a case study analysis of a media campaign or news story, identifying any biases, target audiences, and the techniques

used to convey the message. Discuss how media literacy skills could help someone critically evaluate this campaign or story.

3. Using publicly available data sets, perform a basic data analysis task, such as identifying trends or disparities in the data. You should also discuss any potential biases in the data and how data literacy skills can help them interpret the information.

REFERENCES

Bill S.213 188th (2013 – 2014). (n.d.). Massachusetts Legislature. Retrieved April 25, 2023, from https://malegislature.gov/Bills/188/S213

Bursin, J. (2021). *The Hybrid Job Economy*. https://www.burning-glass.com/wp-content/uploads/hybrid_jobs_2019_final.pdf

Confronting the Challenges of Participatory Culture: Media Education for the 21 Century. (2009). MacArthur Foundation. Retrieved April 4, 2023, from https://www.macfound.org/media/article_pdfs/jenkins_white_paper.pdf

Hobbs, R. (2010). *Digital and Media Literacy: A Plan of Action*. Aspen Institute. Retrieved April 4, 2023, from https://www.aspeninstitute.org/wp-content/uploads/2010/11/Digital_and_Media_Literacy.pdf

Ho'oulu Staff. (2017, January 7). The Real Consequences of Fake News – Ho'oulu. *The University of Hawaii Maui College*. https://maui.hawaii.edu/hooulu/2017/01/07/the-real-consequences-of-fake-news/

Jin, H. (2020, November 28). *How to Distinguish between Reliable and Unreliable Sources Online*. INKspire. Retrieved May 16, 2023, from https://inkspire.org/post/

how-to-distinguish-between-reliable-and-unreliable-sources-online/-MKNvPDSaxfUKjIneDIr

Media Literacy Defined. (n.d.). NAMLE. Retrieved April 4, 2023, from https://namle.net/publications/media-literacy-definitions/

Stobierski, T. (2021, February 23). *Data Literacy: An Introduction for Business*. HBS Online. Retrieved March 28, 2023, from https://online.hbs.edu/blog/post/data-literacy

Taylor, Kiara. "The Digital Divide: What It Is, and What's Being Done To Close It." Investopedia, April 15, 2022. https://www.investopedia.com/the-digital-divide-5116352.

"The 2020 Global State of Enterprise Analytics." Microstrategy, 2020. https://www.microstrategy.com/en/resources/research-and-reports/the-2020-global-state-of-enterprise-analytics.

PART V
THE AMERICAN MOTION PICTURE INDUSTRY AND BIG DATA

Figure 1. “A film reel being recorded by a film camera, digital art”. ‘Film Reel’ by Brendan Smith, created with DALL-E-2

Chapter Written by Brendan Smith

Learning Objectives

- Explain how big data has transformed the film industry, particularly in predicting box-office success and targeting marketing campaigns.
- Gain an understanding of the ethical and equity considerations associated with using big data in the film industry, including the potential for data bias and underrepresentation of certain groups.
- Identify the advantages and limitations of using big data in various stages of film production and distribution, from pre-production to post-release analysis.

INTRODUCTION

The film industry has undergone a major transformation in the past decade, thanks to the emergence of big data. With the ability to predict box-office success for upcoming releases, film production companies can make more informed decisions about everything from advertising to release dates. This is a significant improvement from earlier methods such as focus groups and analyzing box-office receipts, and it's all thanks to the incorporation of social media and other new data sources. By using big data, companies can now analyze marketing campaigns and audience responses to find potential hits or even decide to delay or cancel a film. But that's not all; big data can also be used to predict award winners and to identify potential films that audiences will want to see. In this chapter, we'll explore the ways in which big data has revolutionized the film industry, and we'll take a look at its potential for further advancements in predicting box-office success. This is a behind-the-scenes look at how big data is used and even transforming the film industry. As well as the challenges and limitations of these approaches.

PREDICTING BOX-OFFICE SUCCESS IN THE FILM INDUSTRY

In the world of the film industry, there are many films that are released each year to the public and are successful, but there are also films that do not see the same success as the others. In the United States alone, 449 films were released in 2022. This number is slowly rising back to the 2019 level of 792 films after falling during the COVID-19 pandemic. (*U.S. & Canada*, 2023). Film production companies want to be able to figure out whether or not they should fund a project based on its predicted box-office success. Currently the methods that are being used are not officially known, but there is evidence that suggests that data is being used for predictions by the companies in the American Motion Picture Industry. For a film production company to be able to predict the success of a film when it is in the stages of pre-production, there needs to be something out there for the film, such as a trailer for the people to be able to see a preview of the project, before companies are able to begin making predictions. 20th Century Studios has machine learning models to help before pre-production, but there are no known results. With that being

said, how and when can a production company within the American Motion Picture Industry predict the success of their films at the box-office? Read on to find out.

Before there was access to big data, companies used much simpler methods for prediction. Focus groups were employed to predict success for a film, as well as analyzing box office receipts to predict the potential success of a film based on similar films. There was one major issue with these methods, they all lacked easily accessible data, causing them to fail as reliable methods (Simon & Schroeder, 2019).

BIG DATA IN THE FILM INDUSTRY

Before continuing on, it is important to mention that the methods mentioned below are data sets and models created by third parties and not the companies within the American Motion Picture Industry. The third parties created models that they believe could be or could have been used by the companies. With the exception of the previously mentioned 20th Century Studios and Netflix, which will be talked about in detail later on, the companies in the industry tend to not release information about the models that they use. It can be assumed that this is the case due to the fact that the companies do not want competitors to be able to access or create similar models to them if they have seen relative success with their particular data sets and models. It is also important to note that, due to this being the case, we do not know from whom or where the data they use is collected from. This raises the question of potential data bias and how they are using such data.

The emergence of big data within the film industry has allowed for many avenues of box-office predictions to become available. But what is the big data within the film industry? The big data about the film industry used to be far more

limited. It only held box-office receipts, surveys, focus groups centered around the awareness and the attitude towards a particular film, and the outlets in which advertising for films were placed (Simon & Schroeder, 2019). However, the amount of available data has increased significantly. One of the main contributors to this is social media. The data now includes the online views gained by various trailers for a film, the social media posts about a particular film and the positive and negative engagement that they garner in regard to the yet-to-be-released film (Gold et al., 2013, Simon & Schroeder, 2019). Another area in which data was gathered was from the comments of the website Rotten Tomatoes. Rotten Tomatoes is a website where users are able to view critical and audience ratings of films and television shows, as well as comments made by the audience in regard to the films. These different areas and aspects of the data now allow for much more robust ways for companies to be able to predict the success of their yet-to-be-released films and to plan accordingly for the potential success or failure.

How can the data be used by companies within the American Motion Picture Industry? With the emergence of new data, these companies can analyze the responses of the audience toward their marketing campaigns and to see if there is interest garnering in regard to the movie before its planned release. With these comes the most important prediction for the companies, the box-office success of their film (Simon & Schroeder, 2019). Being able to predict how well their movie

will perform once it releases into theaters allows for the companies to have opportunities in front of them that may not have been present or usable without the knowledge of the data. These opportunities could include funding further advertising, delaying the release of the film, or even altogether canceling the film and its production.

HOW DATA HAS BEEN USED IN THE FILM INDUSTRY

Not all uses of the data go toward the success of the film at the box-office. The data can be used in a wide variety of ways. For example, in 2013, Farsite Group used data that they had gathered to predict the winners of six of the main awards for the 85th Oscars. They used Rotten Tomatoes ratings given by both critics and audience members, the box office success of each film, and if the films had won any awards at award shows that take place before the Oscars, such as the Directors Guild Awards, and the Golden Globes (Gold et al., 2013). In doing so, they were able to accurately predict 5 out of 6 of those awards. In 2014, Farsite Group once again predicted the Oscars (Pomerantz, 2014). There were no articles mentioning the outcomes of their predictions, but after cross referencing the Forbes article mentioning the predictions with the official Oscars website (*The 86th Academy Awards,* 2014), Farsite Group accurately predicted 6 out of 6 awards for that year.

Another example comes from a study done by Sitaram Asur and Bernardo A. Huberman (2013). Together, they analyzed 2.9 million tweets from 1.2 million different users about 24

different films. Considering the mentions of the film, the positivity of the tweets, etc. they were able to use a linear regression model that allowed for them to be able to show the relationship between the spread of people talking about the film and how successful the film was likely to be because of that. One example from their data set was the film *Avatar*. A week before its release it accumulated around 1212.8 tweets per hour. They were able to use this through their model to show that based on the number of tweets surrounding the film, it would be a successful film within its first week of release. Through their work, they were able to prove that data gathered from social media sites can effectively predict the future outcomes of a particular film's success at the box office. They were also able to prove that this method of analysis and prediction worked much better than the predictions of the Hollywood Stock Exchange. The Hollywood Stock Exchange is an online virtual stock market where users are able to buy and trade stock using virtual fake currency to make predictions for which movies will be a success at the box office. The method used by Asur and Huberman helped to prove that the more a film is positively talked about prior to its release, the better the film will perform at the box-office when it comes time for it to be released into theaters (Asur & Huberman, 2013).

An interesting tool that could eventually be used by many companies comes from 20th Century Studios. They have revealed that they began using machine learning models before

they even started pre-production. These machine learning models collect data and help find potential films that audiences will want to see. 20th Century Studios uses this to guide themselves when buying a script (Kapoor, 2021). They take labels created for their films and then they feed those through the machine learning models to help them discover potential scripts. This changes how the process works. Traditionally, a producer would have assistants who go through the scripts for them and author short reports on the scripts. It is then up to the producer to read the reports and decide which film they would like to make next. The machine learning models can now choose the scripts and then the assistants narrow down the chosen scripts rather than narrowing down every script. This could increase the output of more successful films overall. It can be seen as a potential huge money saver for companies looking to produce certain films. A lackluster script can be better avoided rather than be made and create a net loss after it has been produced into a movie, allowing for there to be less of a net loss when that film has a high budget and ends up not performing well at the box-office and after word spreads that there is not much substance to the film due to the script. The model considers what audiences might want to see next, but it is unable to account for any unexpected breakthroughs in the industry. This is due to the model relying on the earlier scripts that are considered successful by the studio.

"Predicting Movie Prices Through Dynamic Social Network Analysis" used both the Internet Movie Database

(IMDb) and the Rotten Tomatoes moving rating parameters, along with the "buzz" around a film and posts gathered from the IMDb forums (Simon & Schroeder, 2019). This allowed them to be able to make predictions for a film during the first four weeks after its release. But while they had some success with this method, the information does not really become useful for the studios as their film is already in theaters and they can figure how well their film will do based on their opening weekend.

Another study that garnered results came from, "Predicting consumer behavior with Web search" (Goel et al., 2010). They were able to use data gathered from Yahoo!'s search engine for box-office predictions (Simon & Schroeder, 2019). The data that they gained from the search engine came in the form of searches from individual users. They were able to compile the data based on whether or not there was a link to IMDb within the immediate search results and they then mapped out the movies based on which movie the IMDb link led to. With this, they were able to create predictions that worked well. They also found that the results worked particularly well due to users using the search engine to search for the film they were interested in and where they would eventually be able to see the film near them when it was released into theaters. But even still they pointed out an issue with this method, "the main advantage of using a search behavior may not be accurate but rather the ready availability of these data." (Simon & Schroeder, 2019, p.554). Having the data sets there to use

is indeed a fantastic advantage due to their availability, but having Simon and Schroeder say that they may not be the most accurate leads to the conclusion that those data sets would be most beneficial if they were a part of a model that takes data from multiple sources to be able to accurately create box-office predictions by analyzing multiple data sets from different sources rather than relying on the data from one particular source. This raises important equity considerations. For instance, the data could be skewed due to the digital divide, underrepresenting people from lower socio-economic backgrounds who may not have regular internet access. Additionally, the data may carry cultural and language biases, as it predominantly captures the behaviors and preferences of majority populations or those who are more active online.

INTO THE FUTURE

New ways of using data are regularly being invented and used. Surveys and categorization of age ranges that are used in data are becoming more finely sharpened. "That blunt instrument is fast giving way to computers that can render us in fine detail by picking up the trails of digital breadcrumbs we leave online and building them into predictive models of what we like and don't like." (*Big Data and Hollywood*, n.d.). The methods allow for a much more exact prediction model as to who will want to see which movie and how to get that movie to appeal to an even wider audience. "Studios can use these real-time opinion assessments to do all kinds of tweaking after a movie has been made; targeting specific demographics in marketing campaigns, tailoring trailers so that they appeal to the kinds of people who will be drawn to a particular movie, pushing distribution to geographic areas where the target audience lives." (*Big Data and Hollywood*, n.d.). implementing these into the film industry will allow a film company to take a hold of their box-office success to an extent. This will allow those companies to obtain a much wider audience for their films. And in doing so, they will be able to increase their box-office success in ways that were previously not available to them.

Another potentially game changing system comes from

Cinelytic. They have been working on an AI system that aids film production companies in new ways. "It licenses historical data about movie performances over the years, then cross-references it with information about films' themes and key talent, using machine learning to tease out hidden patterns in the data." (Vincent, 2019). This essentially creates an AI producer for companies. Analyzing all aspects of the data allows for it to look for those trends that work and when they may work and when to shift the focus to try to gain attention from a wider audience.

A problem also arises with the use of AI in film. For example, say that it was to be used to gather audience feedback before the film was released. It could then make recommendations that can be used to make the film more in line with what the audience is expecting it to be. For example, before the film *Snakes on a Plane* was released, it began to garner attention and so the studio decided to have reshoots for parts of the film so that they could incorporate feedback from the audience (Simon & Schroeder, 2019). And upon doing so, when it came time for the theatrical release of *Snakes on a Plane*, it fell short of what it was predicted to make at the box office. Taking this into consideration, while this is speculatory, if the AI is to take into account what this audience is interested in seeing based on their feedback, what is there to stop the method from being the same in terms of failure when relating this method to the one used for *Snakes on a Plane*?

THE ADVANTAGES OF BIG DATA IN THE FILM INDUSTRY

One of the advantages of big data is that it allows for a "...more accurate and detailed customer information at the individual level and uses the information for a very narrow and specific segmentation of customers..." (Rust & Huang, 2014, p.209). This was mentioned earlier with 20th Century Fox. It also allows for a better and more exact picture of the audience so that they become more understood by those who are creating films. This can allow for there to be a much more diverse and better representation of various groups of individuals who were once unrepresented or misrepresented within film.

> *"Various sources of data can be combined, including not just social media data but also geo-location (where which movies are popular), credit card data, and the like." (Simon & Schroeder, 2019, p.558).*

Having geo-location available is an essential part of the process for film production companies. Being able to know where certain genres are more popular than others allow for the companies to better plan out and spend money on the number

and locations of movie theaters where the film is released. Being able to limit how many movie theaters play a film in a certain location allows for money to be saved rather than be spent on too many theaters in areas where that film genre may not be as successful as it is in other regions. However, this approach raises ethical concerns, particularly for rural or economically disadvantaged areas that may not have the customer base to warrant the release of certain films. These areas could be left out, limiting their access to diverse cultural content. Moreover, the data could perpetuate existing biases, as big chains might focus only on genres that are already popular in specific geographic locations, thereby reinforcing existing cultural divides.

When a studio has their films shown at a chain like AMC, it tends to play in most if not all of the locations that that chain owns. But for smaller independent theaters, they don't always have the ability to showcase the films in their cinemas. Data could help with this to an extent. In a location where a major cinema chain isn't present, and a smaller cinema is, then if that area matches the demographic of those who have interest in a film, a studio can make a deal with that smaller independent cinema to have their film shown at that cinema to increase profit in the area. Allowing for a film to only be shown in theaters in locations in which those genres of films are popular amongst movie goers allows for there to be a better box office return for the companies, rather than if they were to

showcase their film in theaters where the genre of their film is less popular than other genres.

THE LIMITATIONS OF BIG DATA IN THE FILM INDUSTRY

One of the limitations is the lack of data on low-budget independent films (Simon & Schroeder, 2019). It becomes harder to successfully predict the box-office success of films that are smaller in scale. This is due to the lack of information in regard to the film. The less information that is available in a film, the harder it becomes for a prediction model to be able to use the data to accurately make a prediction.

Another limitation is that it does not consider who the potential audience for the film is and the audience for its actors. For example, the film *Ticket to Paradise* was not expected to do well in the United States (U.S.). Upon its release in the U.S. It had made $80 million overseas and was predicted to only accumulate $6.4 million in its opening week in the U. S. But in a surprise, *Ticket to Paradise* managed to secure $16 million in its U.S. debut (McClintock, 2022). It was reported that sixty-four percent of the audience for the film in its first weekend were older than 35 (*What the "Ticket to Paradise" Box Office Opening Says about the State of the Rom-Com*, 2022). The age of the audience is a key factor in this. In a movie

starring two older actors who are less known to the younger audiences, the main audience who is likely to go see the actors and actresses are familiar with those actors and actresses from when the audience themselves were younger and watch their films.

There are no officially known prediction models, except for one exception, which are able to make predictions for which actors should be hired for a particular project. The one exception to this is Netflix. In the past it has been revealed that for the Netflix series *House of Cards*, Netflix was able to use data gathered from the streaming platform to make a prediction that pairing David Fincher and Kevin Spacey would be a success. This was based on the popularity of films directed by David Fincher and films starring Kevin Spacey (Carr, 2013).

When it comes to predicting success at the box-office, there is one limitation that can't be accounted for. With predictions based upon social media users' activity in regard to the film before its release, there is no way to accurately predict the potential flop of the film at the box office until it has already happened. A film can be predicted to be a success with all the talk surrounding it prior to its release and have the estimated box-office receipt and how many theaters it will play in, but nothing can account for the word of mouth spread of negativity towards a film that is not rated well in the eyes of the audience. The audience who has seen it and disliked it overall can lead to the unpredictable possibility of the film becoming a failure at the box-office when it was previously seen as a

potential enormous success. Asur and Huberman (2013) have found that after the release of a film, sentiment on social media can affect the predictions of a film's box-office revenue.

Another limitation for big data within the industry comes down to when the usable models can create accurate predictions for the films. "Big Data Goes to Hollywood: The Emergence of Big Data as a Tool in the American Film Industry" brings this to light. Asur and Huberman's (2013) Twitter-based model was only able to make a prediction that could be seen as dependable the night before the film was released into theaters in the United States (Simon & Schroeder, 2019). While the model has shown that it can help to accurately predict which films will be a success, there is no real help when it can only supply the info the night before. At that point, it does not become beneficial for the studio to know at that point due to there being nothing for the studio to do about how the film is distributed.

A main point to remember from the Yahoo! and other search engines includes: when solely working off of the data sets from only one or a few sources at a time can lead to success with predictions, those predictions will contain a lot more failed predictions than there would be if you were to utilize many various data sets to create an overarching model in which box-offices predictions can become more and more successful. One major limitation of this analysis is that companies reveal little to no information on their prediction models, but we do

know that they create these overarching models to consider the possibilities from all sides (Kapoor, 2021).

OUTLOOK

In "Big Data Goes to Hollywood: The Emergence of Big Data as a Tool in the American Film Industry", Felix M. Simon and Ralph Schroder (2019, p. 560) bring up a valuable point: "...rationalization has of course affected the kinds of movies that are made and how they are made. And films with large budgets are affected more than lesser movies simply because the stakes are so high." Essentially, what is being argued is that while those working for studios claim that there is no effect upon the creative process of a film, there still is much effect that it has. One of the reasons a person writes a script or wants to direct a film is to use their creativity. But there comes a point at which that creativity begins to become hindered. And big data can be seen as a contributing factor to that hindrance. This can especially be seen with 20th Century Studios' models and Cinelytic's AI model. If a script is not seen as appropriate for what should be made, then no chance is taken on that script, or that script is then changed to meet the algorithmic recommendations rather than allowing for the full creativity of its creators. And, for the AI, if it looks for hidden patterns that will help make a film more successful, the studio would benefit and the creativity of the film would be hindered. If the film is to be changed to be more in line with what is supposedly the

best way to make it a complete success, then it can fail in the creative process as it tries to stick to a formula that is predicted to work. The use of big data in the film industry is beneficial when it comes to marketing in advertising as mentioned throughout this chapter, it is only when it enters the pre-production world that it becomes a moral question of creativity.

CONCLUSION

The film industry has evolved in the past decade with the emergence of big data, allowing film production companies to predict box-office success for their upcoming releases. However, it's crucial to question whether these data-driven methods are inclusive and equitable. For instance, the #OscarsSoWhite campaign highlighted the disparities in awards and recognition across different populations. The predictive models being used may replicate existing biases. Therefore, while this is a significant improvement from the earlier methods such as focus groups and analyzing box office receipts, problems remain The incorporation of social media and other new data sources has revolutionized the industry, allowing for more in-depth analysis and prediction of the success of upcoming films. Companies can use this data to analyze their marketing campaigns and audience response, leading to more informed decisions about advertising, delaying, or even canceling a film. However, as the industry leans more into data-driven decision-making, it's essential to ensure that these methods don't perpetuate existing inequalities or overlook diverse talents and stories. Additionally, big data can be used to its full extent through the numerous ways in which it can be used throughout the

entirety of the American Film Industry. The film industry has taken a big step forward, thanks to big data and its potential for further advancements in predicting box-office success, but it must also take steps to ensure that this progress is inclusive and equitable.

WRAP UP

Key Takeaways

- Big data has revolutionized the film industry by enabling more accurate predictions of box-office success, thereby informing decisions on everything from marketing to release strategies.
- While big data provides valuable insights, it also raises ethical concerns such as data bias, which can perpetuate existing inequalities and overlook diverse talents and stories in the film industry.
- Traditional methods like focus groups have been largely supplanted by big data analytics, but limitations still exist, such as the scarcity

of data on low-budget independent films and the late timing of reliable predictions.

- The application of big data is extending into various aspects of film production, even influencing script selection and casting decisions, which raises questions about the balance between data-driven decisions and creative integrity.

Exercises

1. How does the use of big data in the film industry compare to its use in other industries like healthcare, retail, or finance? Discuss both the advantages and ethical concerns that are unique to the film industry.
2. Consider a recent film that either succeeded or failed at the box office. How might big data

analytics have influenced the film's marketing strategy, release timing, or even content? Provide specific examples to support your analysis.

3. Debate the implications of using big data to influence creative decisions in filmmaking, such as script selection or casting. Do you think the use of data analytics enhances or hinders artistic creativity? Justify your position.

REFERENCES

2014 | Oscars.org | Academy of Motion Picture Arts and Sciences. (n.d.). Retrieved April 18, 2023. https://www.oscars.org/oscars/ceremonies/2014

Asur, S., & Huberman, B. A. (2013). Predicting the Future with Social Media. *Applied Energy*, *112*, 1536–1543. https://doi.org/10.1016/j.apenergy.2013.03.027

Big Data and Hollywood: A Love Story. (n.d.). Theatlantic.Com. Retrieved April 4, 2023. https://www.theatlantic.com/sponsored/ibm-transformation-of-business/big-data-and-hollywood-a-love-story/277/

Carr, D. (2013, February 25). Giving Viewers What They Want. *The New York Times*. https://www.nytimes.com/2013/02/25/business/media/for-house-of-cards-using-big-data-to-guarantee-its-popularity.html

Goel, S., Hofman, J. M., Lahaie, S., Pennock, D. M., & Watts, D. J. (2010). Predicting consumer behavior with Web search. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(41), 17486–17490. https://doi.org/10.1073/pnas.1005962107

Gold, M., McClarren, R., & Gaughan, C. (2013). The Lessons Oscar Taught Us: Data Science and Media &

Entertainment. *Big Data*, *1*(2), 105–109. https://doi.org/10.1089/big.2013.0009

Kapoor, A. (2021, November 3). *The Role of Data Science In Movie Making*. The Smart Cube. https://www.thesmartcube.com/resources/blog/datawatch-lights-camera-analytics-the-secret-role-data-plays-in-movie-making/

McClintock, P. (2022, October 22). Box Office: 'Black Adam' Heads for Solid $60M Opening While 'Ticket to Paradise' Travels in Style. *The Hollywood Reporter*. https://www.hollywoodreporter.com/movies/movie-news/black-adam-box-office-previews-ticket-to-paradise-1235246172/

Pomerantz, D. (n.d.). *Can Social Networking Accurately Predict Oscar Winners?* Forbes. Retrieved April 18, 2023, from https://www.forbes.com/sites/dorothypomerantz/2014/02/24/can-social-networking-accurately-predict-oscar-winners/

Rust, R. T., & Huang, M.-H. (2014). The Service Revolution and the Transformation of Marketing Science. *Marketing Science*, *33*(2), 206–221.

Simon, & Schroeder, R. (2019). Big Data Goes to Hollywood: The Emergence of Big Data as a Tool in the American Film Industry. In *Second International Handbook of Internet Research* (pp. 549–567). Springer Netherlands. https://doi.org/10.1007/978-94-024-1555- 1_63

U.S. & Canada: Movie releases per year 2022. (2023, February

13). Statista. https://www.statista.com/statistics/187122/movie-releases-in-north-america-since-2001/

Vincent, J. (2019, May 28). *Hollywood is quietly using AI to help decide which movies to make*. The Verge. https://www.theverge.com/2019/5/28/18637135/hollywood-ai-film-decision-script-analysis-data-machine-learning

What the "Ticket to Paradise" box office opening says about the state of the rom-com. (2022, October 23). Los Angeles Times. https://www.latimes.com/entertainment-arts/movies/story/2022-10-23/ticket-to-paradise-box-office-julia-roberts-george-clooney

PART VI
DATA IN SPORTS MARKETING

Chapter Written by Sophia Moore[1]

Learning Objectives

- Understand the various sources and types of data used in sports marketing, including community data, private organizational data, and public data repositories.
- Gain insights into the ethical considerations surrounding the collection and use of big data

1. This chapter was written with assistance from ChatGPT.

in sports marketing, particularly in relation to audience privacy and personalized marketing.

- Analyze and discuss the future implications of Artificial Intelligence (AI) in the field of sports marketing, including its potential for predictive analytics, real-time insights, and ethical concerns.

INTRODUCTION

Today, marketing has become part of nearly every business. Through marketing, businesses and companies are able to engage with their consumers, which allows them to increase sales and their brand's identity (Emeritus, 2022). A major part of marketing is reaching your target audience. In order for companies to understand who exactly their target audience is, big data begins to play a huge role. Data such as advertisement engagement, sales, audience demographics, and more help businesses understand how to market their brand better.

With all this data being collected on individuals to better understand what meets their needs and wants, there is an ethical line that must be drawn. Of course, data is essential and almost unavoidable for an effective marketing campaign, but this data should not breach a user's privacy rights. Later in this chapter, the topic of ethically right and wrong uses of data used for marketing will be analyzed.

In this chapter, the importance of big data used specifically for sports marketing will be discussed. Sports marketing is an industry that relies heavily on data analysis to help understand consumer behavior and drive better business decisions. Data overall has become a key component in sports marketing, as it helps teams, leagues, and sponsors identify key trends,

understand audience demographics, and make informed decisions about marketing strategy.

UNDERSTANDING BIG DATA

The technical definition of big data is "high-volume, high-velocity and/or high-variety information assets that demand cost-effective, innovative forms of information processing that enable enhanced insight, decision making, and process automation," ("The role of Big Data in sports marketing," 2023, para. 3). Big data benefits sports marketing by providing a wide set of data; with a wide range of data, more effective business strategies can be built. These business strategies not only can benefit the business but also the customers, employees, and players.

Effective sports marketing with the assistance of data can help a player gain more exposure and reach a wider audience. This can lead to increased recognition and a higher profile, which can translate into greater endorsement deals, better contracts, and greater opportunities for career advancement. Fans can also experience benefits such as better ticket pricing and content that caters to their specific needs.

Big data allows for a comprehensive understanding of the audience, even in industries with diverse and vast audiences, such as the sports industry. This understanding encompasses their traits, attitudes, and actions, enabling the prediction of

future actions. In essence, big data provides a precise and continually evolving map of the target audience, which can be used to tailor actions, timing, and targeting strategies to effectively reach them ("The role of Big Data in sports marketing," 2020).

SOURCES OF DATA IN SPORTS MARKETING

Sports stand to gain significant benefits from the increased availability of community data through blogging sites, social network trends, and content communities. This is evidenced by the fact that, in 2019, three of the top ten most searched stories on the Internet were related to sports (Moreno, 2019). In response, sports leagues are increasingly turning to digital programming to engage and connect with fans. Digital programming is essentially media found online such as websites, social media, and sports apps. The majority of sports fans interact on social media, which demonstrates the platform's effectiveness in facilitating engagement. Iconic soccer teams such as Real Madrid and Barcelona have amassed over 210 million Facebook and Twitter followers each, emphasizing the reach and impact of social media platforms in the world of sports. Brock and Khan (2017) have identified social media as a vast source of big data that can be utilized to derive insights and inform decision-making in sports.

The second development in sports data involves the utilization of private data collected within sports organizations through recording consumer transactions. This is achieved through the use of clickstream to categorize visits by location,

purchase, search, and the use of mobile applications. Clickstream data is the information collected about a user while they browse through a website or use a web browser (Gillis, 2022). The emergence and popularity of technology have contributed to the digitalization of sports, granting access to customers' and firms' digital interactions (Lazer and Radford, 2017). This interconnectedness of sports leagues and customers through digital platforms allows teams and sponsors to gather and share unique information. DeSchriver et al. (2021) provide an example of the potential benefits of this approach by collecting daily performance data on hotel occupancy rates and average daily rates to examine the impact of Southeastern Conference college football games on local hotel demand from 2003 to 2017. This study collected data from the hotel management analytics firm Smith Travel Research (STR), which represents proprietary data in the tourism and hospitality industry. The study highlights the use of volume and variety data to advance demand literature in sport and hospitality.

The third area of data availability in sports comes from a variety of sources such as government, research funding agencies, professional societies, universities, individual researchers, and other public data repositories. Uhlir and Schröeder (2007) have suggested that publicly funded data can be beneficial for reuse by a broad range of researchers, socioeconomic applications, and the general public. Large datasets related to sports have been published by the National

Collegiate Athletic Association, sports-reference.com, and official university websites. For instance, Jensen et al. (2020) collected 169,479 observations to explore donor behavior in the intercollegiate athletic industry. The study found that the probability of donor contraction increases with decreasing economic growth. Additionally, public data repositories offer historical and other data for professional and Olympic sports, as well as all co-educational postsecondary institutions that receive Title IV funding and have an intercollegiate athletics program. Overall, sport marketing researchers can benefit significantly from using public data produced by various entities to enhance the quality and productivity of research. (Mamo et al., 2021).

While there has been progress in collecting data from community, private, and public sources, there are still many opportunities for continued contributions. Sports blog websites and fan-hosted podcasts are emerging as potential sources of valuable data that can help sport marketers uncover fans' opinions and create new business opportunities. This data on fans' opinions can be gathered through comments and deciphering the positive from the negative feedback through tools such as sentiment analysis. The data collected can also be expanded from text data to image, audio, and video data from various sources (Mamo et al., 2021). This would allow researchers to gain a deeper understanding of the role that different forms of media play in the world of sports marketing.

DATA AND TARGET AUDIENCE

Sporting events bring in a massive amount of viewers. For example, the 2023 Super Bowl brought in an estimated total of 113 million viewers, while an estimated 1.5 billion people viewed the 2022 World Cup (Nielson, 2023). With such a huge audience, it may seem overwhelming to think about how franchises are supposed to cater to their needs. This is where data on audience demographics becomes essential.

To break down data demographics, let's use the world's most popular sport as an example: soccer. In a study found on Doxee ("The role of Big Data in sports marketing," 2020), the Italian soccer team, Juventus, found a change in their audience. Until a few years ago, Juventus had a fan base that consisted mainly of males, mostly from Italy or Europe. However, through data polled from social media following and engagement, it was found that more and more women have begun to follow soccer. More and more people who live outside of Europe and Italy also follow Juventus. Juventus' star player, Cristiano Ronaldo, who has 241 million followers on Instagram compared to Juventus' following of 38 million, has a demographic of 61.5% male and 38.5% female (Starngage, n.d.). This trend of more female engagement in sports has

become common throughout athletics. To find the data of this demographic, social media has become a big tool. Through social media, data such as followers, liked posts, and views can be broken down into demographics.

So what does this data mean for future sports marketing? Using data to understand a demographic allows sports marketers to make their franchise more personal. For example, from data showing a trend in women watching sports, franchises can respond by selling more merchandise made for women. This could mean merchandise made in women's sizes or more popular items for women such as leggings.

Another way data is collected for the target audience is through ticket sales. As Pete Giorgio (n.d.) has written: "With richer data, sports teams can know who was at the game, their in-stadium purchase history, and where they moved within the stadium. Having this specific information will enable more focused sponsor targeting and authentic engagement both inside and outside the stadium" Combing these metrics will allow a much deeper analysis of how fans are interacting with products and advertisements.

DATA USED FOR REVENUE

Not only do sporting events get a massive engagement through media, but in-person sales also generate a large amount of revenue. The average crowd size for an NFL game in 2022 was 69,442 people (Smith, 2023). To guarantee sports fans will be more likely to attend games, it is vital to set prices at an ideal price. Pricing is very important because it needs to bring in profit and be realistic for fans, the ideal price of a ticket can be found through data.

Researchers have created a tool to optimize sports ticket prices for both management and fans, in order to gain a better understanding of what people are willing to pay for sports tickets. This innovative approach is designed to improve the management of ticket prices and make them more attractive to fans, while also ensuring that they are fair and reasonable. This research was explored through Manufacturing & Service Operations Management by co-authors Robert Easley, John W. Berry Sr. Department Chair and Professor of Information Technology, Analytics and Operations at the University of Notre Dame, and Ovunc Yilmaz, assistant professor at the University of Colorado Boulder (Wampler, 2022).

The research team collaborated with an NCAA Division

I football program and analyzed its ticket sales data (Greene et al., 2017). Through a careful review of fans' purchasing behaviors and demographics, the researchers investigated two primary sales channels: season tickets and single-game tickets. By scrutinizing the audience segments within each channel, the research team was able to identify and differentiate between different groups of fans based on their purchasing patterns and preferences. This approach allowed for a more nuanced understanding of how fans engage with the ticketing process and how the program can better tailor its ticketing strategies to meet their needs.

Through this research, season ticket holders were broken down into three categories: big donors, the public, and employees. General ticket sales were also broken down into three categories: donors, alumni, and parents.

An unexpected discovery from the analysis of customer segment data was that as the number of available seats in a given section dwindled below a certain threshold, fans appeared to be less interested in purchasing seats in that section. This trend may suggest that fans do not perceive the remaining seats, which are typically located on the fringes of the section, to be a good value for their cost. Moreover, the research uncovered a divide among fans in terms of their price sensitivity: some fans prioritized watching the game from the best seats, regardless of the cost, while others were content with watching the game from the cheapest seats available. These insights shed light on the complex factors that influence

fans' ticket purchasing behaviors and can inform more targeted and effective ticket pricing strategies (Wampler, 2022).

DATA USED FOR CAMPAIGNS

Advertising campaigns are fundamental for marketing. They allow companies to capture the attention of their core audience. Campaigns can be done in multiple ways and reach their audience through multiple platforms. The key for a successful campaign is knowing which platform will reach their audience the best, this is where data is important. Data can identify which platform the target audience engages with most and use that information to create the most effective campaign.

There are a few brands that are immediately associated with sports when they are mentioned. These are brands such as Nike, Addias, and Under Armour that have been notorious for their collaborations in the sports industry. One of the most common sports marketing strategies is using popular players to promote products. Data is essential for this marketing approach because brands need an understanding of what individual will be well perceived by their target audience.

In 2019, Nike collaborated with former NFL player Colin Kaepernick to promote athletes following their dreams. This was a campaign that was very strategic and took an immense amount of data to predict audience reactions considering a

controversy surrounding the player at the time, which we will begin to explore further. Kaepernick began kneeling during the pre-game anthem in 2016 as a way to protest racial injustice (The Guardian, 2019). Despite negative responses from this campaign, including distasteful comments from former president, Donald Trump, the campaign proved to be successful with a resulting 5% increase in Nike stock.

In this particular campaign a variety of data has to be collected considering all perspectives. Data also has to be collected on how advertising can affect culture and individuals. This data is shown through audiences' positive and negative comments on social media posts and their engagement with the brand, such as the stock increase in Nike.

A study was done via questionnaires by Yun Kuan of Portland University in 2019 to get a better understanding of how audiences perceive controversial campaigns. In order to understand how controversial advertising impacts consumers' views on the social issue and the brand, participants were asked to react to two advertisements by different brands featuring different controversial issues. This was done to avoid bias on the specific issue or brand. "A major factor of this study includes participant's political views and past life experiences; therefore, an internet-mediated questionnaire was utilized to obtain participants from throughout the United States, thus ensuring a variety of backgrounds," (Yun Kuan, 2018, p. 20).

Through this research, the data revealed that controversial advertising is similar to a mirror, as it reflects individuals' pre-

existing values. The International Advertising Association (1977) explains that controversial advertising is attractive to individuals who already support the cause and thus, it mirrors their political and social perspectives. If controversial advertising is a mirror, then corporations may not be effective in using it to alter people's political and social opinions (Yun Kuan p. 32).

This shows the amount of thought Nike had to put into creating campaigns. Data must be collected to represent multiple different perspectives before producing any new marketing strategies. This is because sports bring in a very diverse audience that have many different views.

THE FUTURE OF SPORTS MARKETING

Artificial Intelligence (AI) is already changing the sports marketing landscape in numerous ways, and is poised to have an even greater impact in the future. The impacts include more personalized marketing, predictive analytics, real-time insights, and influential marketing.

AI technology enables marketers to collect and analyze vast amounts of data, which can be used to personalize marketing messages and promotions to individual fans. This data consists of the topics mentioned prior in this chapter, but, with AI, it will be gathered more efficiently. This will include helping identify the most effective social media influencers to partner with based on factors such as audience reach, engagement, and demographics. This can create a more engaging and relevant experience, leading to increased fan loyalty and engagement.

AI-powered predictive analytics can be used to forecast future trends in fan behavior, identify opportunities for growth, and optimize marketing campaigns for maximum impact. AI algorithms can also provide real-time insights into fan engagement, sentiment, and behavior, allowing marketers to make rapid adjustments to campaigns and messaging.

Since AI is able to collect a vast amount of data at a fast

rate, fans could experience a lot of personalized marketing in the future. An example of this is sports apps, such as the ESPN mobile app. When a consumer creates an account, AI can be used to collect data on this person. This could result in articles relating to the consumers favorite teams or players being promoted more in their feed. It could also result in advertisements being promoted for tickets to their favorite teams games or promotions for their favorite players merchandise.

For some consumers, the idea of more personalized marketing may seem appealing because their interests will be met. However, some may find an ethical concern with the amount of data that is being collected about themselves. There is a fine line between data being collected for the improvement of marketing and invading users privacy. The future of AI is still mysterious since it is still an evolving technology, but it is essential that its use is ethical and fair.

CONCLUSION

Sports marketing is of paramount importance due to its ability to generate significant revenue streams. Sports have a vast and passionate fan base, and effectively marketing sports events, teams, and merchandise can lead to increased ticket sales, merchandise purchases, and sponsorship deals. These revenue sources support the sports industry and contribute to local economies, creating jobs, and driving tourism. Furthermore, sports marketing plays a crucial role in building and maintaining the brand image of teams, athletes, and sporting events. Successful marketing campaigns can enhance brand visibility, recognition, and reputation, leading to increased fan loyalty and attracting new audiences.

Data has become a valuable asset in sports marketing, providing several benefits that contribute to its effectiveness and success. Data allows sports marketers to gain deep insights into their target audience. By analyzing demographic information, consumer behavior, and preferences, marketers can understand their fans' interests, motivations, and consumption patterns. This knowledge helps in creating personalized and targeted marketing campaigns that resonate with specific segments, leading to higher engagement and conversion rates.

Data also enables sports marketers to measure and track the effectiveness of their marketing efforts. With the availability of analytics tools and platforms, marketers can collect and analyze data on key performance indicators such as advertising campaign effectiveness, social media engagement, ticket sales, and merchandise purchases. This data-driven approach allows marketers to identify what strategies and channels are delivering the best results, enabling them to optimize their marketing campaigns and allocate resources more effectively.

The future of sports marketing will continue to improve and become more effective with the assistance of new AI technologies. Data will be able to be collected at faster rates resulting in more marketing strategies. Of course, as more data is collected more quickly, it will be difficult to ensure it is unbiased. There will also be an increase in personalization in marketing that will fit the interests of individual fans. The industry is constantly evolving along with technology and continues to change and shape the sports world.

WRAP UP

Key Takeaways

- Big Data plays a critical role in sports marketing by providing deep insights into audience demographics, preferences, and behaviors, enabling more targeted and effective marketing campaigns.
- Ethical considerations are paramount when collecting and using data for sports marketing to ensure that user privacy is respected and that the data is used responsibly.
- Advances in technology, particularly Artificial Intelligence (AI), are poised to revolutionize sports marketing by enabling real-time insights, predictive analytics, and highly personalized marketing strategies.

- Various sources of data, including community blogs, private organizational records, and public repositories, provide a rich set of information that can be leveraged to make informed decisions in sports marketing.

Exercises

1. How does the ethical use of data in sports marketing align or conflict with your personal views on privacy? Discuss the balance between effective marketing and consumer privacy.
2. Conduct a mini case study on a recent sports marketing campaign that utilized big data or AI. Analyze the campaign's effectiveness, ethical considerations, and the types of data used.

3. How do you envision the role of Artificial Intelligence in the future of sports marketing? Consider both the benefits and potential drawbacks, such as issues of data privacy or bias.

REFERENCES

Brock, Vitor, and Habib Ullah Khan. "Big Data Analytics: Does Organizational Factor Matters Impact Technology Acceptance?" *Journal of Big Data* 4, no. 1 (July 10, 2017): 21. https://doi.org/10.1186/s40537-017-0081-8.

DeSchriver, Timothy D., Timothy Webb, Scott Tainsky, and Adrian Simion. "Sporting Events and the Derived Demand for Hotels: Evidence from Southeastern Conference Football Games." *Journal of Sport Management* 35, no. 3 (May 2021): 228–38. https://doi.org/10.1123/JSM.2020-0268.

Emeritus. (2022, April 22). What is the importance of marketing for businesses? discover the undiscovered. Emeritus. https://emeritus.org/blog/what-is-the-importance-of-marketing-for-business/#:~:text=The%20importance%20of%20marketing%20for%20your%20business%20is%20that%20it,%2C%20reputation%2C%20competition%2C%20etc.

Giorgio, Pete. "How Data Can Help Drive Sports Sponsorship and Fan Engagement." Deloitte United States, n.d. https://www2.deloitte.com/us/en/pages/consumer-business/articles/fan-engagement-analytics-improve-fan-experiences.html.

Gillis, A. S. (2022, May 5). What is clickstream data (Clickstream Analytics)?. Customer Experience. https://www.techtarget.com/searchcustomerexperience/definition/clickstream-analysis-clickstream-analytics

Greene, Amanda, Kason O'Neil, and Gary Lhotksy. "Exploring a New Division 1 Football Program on a University Campus: An Application of Collaborative Action Research in Higher Education." *Journal of Work-Applied Management* 9, no. 1 (January 1, 2017): 5–17. https://doi.org/10.1108/JWAM-12-2016-0025.

The Guardian. (2019, September 16). "Nike's 'dream crazy' advert starring Colin Kaepernick wins Emmy." The Guardian. https://www.theguardian.com/sport/2019/sep/16/nikes-dream-crazy-advert-starring-colin-kaepernick-wins-emmy

Jensen, J.A., Wanless, L., Hedt, K. and Wayland, E. (2020), "Analyzing Big Data in intercollegiate athletics: an application to athletic donor behavior", Journal of Issues in Intercollegiate Athletics, Vol. 13 No. 1, pp. 292-311.

Kuan, Yun. "An Exploratory Study on Controversy Advertising's Effect on the Advertiser and the Controversial Issue." Portland State University, 2018. https://doi.org/10.15760/honors.561.

Lazer, David, and Jason Radford. "Data Ex Machina: Introduction to Big Data." *Annual Review of Sociology* 43, no. 1 (2017): 19–39. https://doi.org/10.1146/annurev-soc-060116-053457.

Mamo, Yoseph, Yiran Su, and Damon P.S. Andrew. "The Transformative Impact of Big Data Applications in Sport Marketing: Current and Future Directions." *International Journal of Sports Marketing and Sponsorship* 23, no. 3 (January 1, 2021): 594–611. https://doi.org/10.1108/IJSMS-03-2021-0073.

Moreno, J. (2019, December 25). These were the top google searches and trends of 2019. Forbes. https://www.forbes.com/sites/johanmoreno/2019/12/24/these-were-the-top-google-searches–and-trends-of-2019/?sh=26ad38e83089

Nielsen. "Super Bowl LVII Totals More than 113 Million Viewers, Ranks Second Most-Watched Game Ever." 2023. https://www.nielsen.com/news-center/2023/super-bowl-lvii-totals-more-than-113-million-viewers-ranks-second-most-watched-game-ever/.

PDXScholar: The Institutional Repository of Portland State University. Site. (n.d.). https://pdxscholar.library.pdx.ed

Smith, P. by M. D. (2023, January 16). NFL regular-season attendance hit six-year high in 2022. ProFootballTalk. https://profootballtalk.nbcsports.com/2023/01/16/nfl-regular-season-attendance-hit-six-year-high-in-2022/

StarNgage. "Cristiano.Ronaldoo's Instagram Audience Analytics and Demographics." Accessed May 22, 2023. https://starngage.com/app/pk/influencers/cristiano.ronaldoo.

Super Bowl LVII totals more than 113 million viewers, ranks

second most-watched game ever. Nielsen. (2023, February 14). https://www.nielsen.com/news-center/2023/super-bowl-lvii-totals-more-than-113-million-viewers-ranks-second-most-watched-game-ever/#:~:text=NEW%20YORK%20%E2%80%93%20February%2014%2C%202023,112.2%20million%20viewers%20on%20FOX.

The role of Big Data in sports marketing. Doxee. (2020, October 6). https://www.doxee.com/blog/digital-marketing/the-role-of-big-data-in-sports-marketing/

Uhlir, Paul F., and Peter Schröder. "Open Data for Global Science." *Open Science Journal* 6 (June 29, 2007). https://doi.org/10.2481/dsj.6.OD36.

Wampler, A. B. (2022, March 17). Researchers develop data-driven tool to optimize sports ticket pricing. Notre Dame News. https://news.nd.edu/news/whats-your-seat-worth/

PART VII
DATA IN PUBLIC RELATIONS, SOCIAL MEDIA, AND ADVERTISING

Chapter Written by Ana'aya McGowan Mozell

Learning Objectives

- Explain the role of data in modern public relations, specifically how it is used for campaign effectiveness, behavioral insights, and message adjustments.
- Gain an understanding of how big data can optimize various aspects of a business, such as resource management, operational

efficiency, and product development, in the context of public relations.

- Discuss the emerging role of Artificial Intelligence in the field of public relations, including its potential benefits and disruptions.

Data is used or generated by nearly everything we do in society. It's no surprise data is a big part of the modern public relations (PR) that we know today. PR businesses use the knowledge of data as their number one priority as a digital marketing strategy since public relations is known as a business of influence and understanding of who, what, and how their audience is connected with them and how they resonate with the brand or business. This strategy gives the brand, organization, or company a well-rounded understanding of what and who their target audience is, their profit numbers, and high and low insights on what could be the next waves of marketing.

But, first, what is public relations?

Public relations is the practice of managing and building a positive public image for a company or organization. The PR experts learn how to create media for their brand for press releases of social media messages that help shape the public's

opinion on the brand or company. This helps to increase brand awareness and with the data, – big or small, – it's collected by organizing and analyzing the information. At first glance, public relations and data might not go hand-in-hand together since PR is more focused on social skills than numbers.

But data is a key essential for successful campaigns and better-targeted media outreaches. The data are very ideal key essentials used for proving the effectiveness of PR campaigns, data analysis, behavioral insights, adjusting messaging, and the value of service.

KEY ESSENTIALS

The first data essential is proof of PR campaign effectiveness. Without data, it's not easy to measure the effectiveness and efficiency of a public relation campaign and it's harder to prove that your campaign is influenced by the public decisions to favor your brand or company. Luckily, PR experts are able to record and document the campaigns by spoken words and are able to collect and analyze the data as a present proof of the campaign.

The second data essential is behavioral insights. PR experts gather data by watching and analyzing how a group of people, or the public reacts to what the brand or company is putting out (Lotame, 2022). In a natural way, humans respond more to the environment around them and everything with it. The behavioral data can help you achieve effective ways of making public-approach strategies and making a better image for the company. Understanding how the public reacts and acts towards your company is one of the key functions of public relations.

The third data essential is adjusting messaging. We all know words have a powerful effect on people to influence how someone can communicate or react towards a subject. Inappropriate messaging can damage a brand, an individual,

or a company's image in an instance. However, PR experts can craft careful messaging to target a certain group of people and data from trends that can help realign the message the company is sending.

The fourth data essential is the value of service. The quality of the data that is collected to influence the public can help improve your value as a PR expert. The data is a valuable tool in public relations. It shows the growth and success as a brand or company.

BIG DATA

Now, what is the main focus on data? Why is it important for Public Relations?

The focus of data is the relationships revealed by big data. The term "big data" is used to describe data that is hard to manage or too large in masses that can be both unstructured and structured for use (Big Data: What it is and why it matters, n.d.). To save time, PR experts figure out what's important to analyze through the given data. This helps to improve quicker decision making and strategic planning for businesses. The importance of big data is how you use it. By taking the source of data in Twitter, Facebook, or Instagram for example, companies can find their answers in five ways.

The first way is through streamlining resource management (CFI Team, 2023). A great way to optimize a business is streamlining. Companies work better with optimizing effective operations to minimize their cost and profits. It helps businesses get to their highest potential, saving time and money, and minimizing high risks.

The second way is improving operational efficiency. Operational efficiency is important because businesses find ways to reduce costs, waste, improve their productivity, and improve their quality of products and services. It involves