Artificial Intelligence, Fairness and Productivity

by

Di Yuan

BMgmt, Shanghai University of Finance and Economics, Shanghai, China, 2009

MAcc, University of Melbourne, Melbourne, Australia, 2011

Submitted to the Graduate Faculty of

the Joseph M. Katz Graduate School of Business in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2023

UNIVERSITY OF PITTSBURGH

JOSEPH M. KATZ GRADUATE SCHOOL OF BUSINESS

This dissertation was presented

by

Di Yuan

It was defended on

April 19th 2023

and approved by

Manmohan Aseri (Chair), Katz Graduate School of Business, University of Pittsburgh

Narayan Ramasubbu, Katz Graduate School of Business, University of Pittsburgh

Tridas Mukhopadhyay, David A. Tepper School of Business, Carnegie Mellon University

Dennis Galletta, Katz Graduate School of Business, University of Pittsburgh

Zia Hydari, Katz Graduate School of Business, University of Pittsburgh

Abstract

Artificial Intelligence, Fairness and Productivity

Di Yuan, PhD

University of Pittsburgh, 2023

The widespread integration of advanced AI systems into business activities has substantially transformed how markets operate. AI in the workplace holds immense potential to enhance productivity and revolutionize how knowledge is disseminated among employees. While the benefits of AI are clear, there are still concerns about its use, particularly in terms of fairness and productivity, creating challenges for policymakers and businesses alike as they seek to ensure that the benefits of AI are shared fairly across society. One issue is that AI adoption can lead to unintended consequences that conflict with fairness. For example, in online advertising, AI has enabled advertisers to target users with greater precision than ever before, leading to concerns about discrimination and the potential for bias. Another issue is that AI may not always lead to expected productivity gains. While AI has the potential to drive productivity and economic growth, it is important to recognize that it may also affect the motivation of the workforce.

To address these concerns, my research examines the economic incentives associated with AI adoption and explores potential remedies to mitigate the associated side effects. Through game theoretical models, my research concludes that the disparity in online advertising display may not necessarily stem from purposeful discrimination on the part of advertisers or algorithmic bias, but rather, may arise from the characteristics of ad-auctions. Furthermore, the study highlights that introducing AI in the workplace may result in undesired consequences if firms do not consider the impact of AI on employees' motivations. Using research findings, we formulate recommendations for policies that can prevent negative outcomes and optimize the benefits of AI implementation. These policy suggestions may include promoting fairness in advertising and redesigning reward schemes.

Overall, this paper aims to provide insights into how organizations can adopt AI efficiently while ensuring fairness and productivity. By understanding the potential pitfalls of

AI and the economic incentives of stakeholders affected by AI, policymakers and business owners can develop strategies that maximize the benefits of AI while minimizing its risks.

Keywords: Artificial Intelligence, Algorithm, Fair Advertising, Algorithmic Fairness, Productivity, Pay-for-performance, Human Capital Management .

Table of Contents

List of Tables

List of Figures

Preface

The completion of this work would not have been possible without the support and guidance I have received throughout my academic journey. First, I want to express my deepest appreciation to my advisor, Dr. Manmohan Aseri, for his invaluable guidance and patience. His commitment to academic excellence has pushed me to challenge my own limits. I also want to thank Dr. Narayan Ramasubbu and Dr. Tridas Mukhopadhyay for their insightful advice and constructive criticism in shaping this dissertation. I am indebted to my family for their love, understanding, and encouragement. Their belief in me, especially during moments of self-doubt, has been a constant source of strength. Lastly, I would like to extend my thanks to all those who have inspired and supported me throughout this process, including the professors who have taught me, the supporting staff at Pitt, my dear friends who have been there during the highs and lows of this academic pursuit.

1.0 Introduction

With the exponential growth of data and computation power, artificial intelligence with capabilities resembling humans has become a reality. While state-of-art AI systems cannot fully think, reason, and learn like humans, their scalability and computation speed make them highly appealing. In the past decade, we have witnessed a proliferation of AI-based business models and technological solutions, with new business models created and traditional industries reshaped. However, as AI becomes ubiquitous, concerns have emerged. This raises questions such as: What are AI's business values? What organizational factors should companies pay attention to during AI adoption? What are the implications of ubiquitous AI to the economy and society at large? Are there any ethical and fairness concerns with algorithms, and what can we do about them?

In this dissertation, I describe two substantial side effects of AI, examine why it happens, and evaluate possible solutions. In Chapter 1, I review the trend in Artificial Intelligence applications and the status of management and economics research on related topics. Chapter 2 delves into the fairness issue in online advertising, and Chapter 3 discusses the potential pitfalls of adopting AI for performance enhancement. Finally, I conclude with a summary of the two essays and the implications of the findings.

1.1 AI in Organizations

More and more organizations recognize AI as a means to gain competitive advantages (Ransbotham et al. 2020). The broader definition of AI encompasses any computer system capable of completing tasks that typically require human intelligence, including rule-based expert systems (Collins et al. 2021). In this paper, the term 'AI' refers to systems that require data, algorithms, and computation power to perform intelligent tasks and can be easily scaled up. The impact of AI on how businesses and society operate has been substantial. This section will summarize two of the most prevalent AI-enabled phenomena.

First, in many industries, AI-powered business models enable newcomers to challenge the incumbent. Specifically, online markets became a prevalent model, reshaping many industries. For instance, online advertising has overtaken traditional channels as the prevailing marketing solution; ride-sharing platforms have made taxis obsolete in many countries; and streaming services have redrawn the entertainment industry landscape. The common feature of these new markets is the crucial role that algorithms play in matching players on the markets. For example, in online advertising, when a user comes online, the opportunity to show an ad to that user, known as an impression, is sold via ad auctions. In these auctions, several advertisers compete through bids to show their ads. Dominating platforms, namely Google, Meta, and Amazon, use algorithms to match advertisers to ad viewers. Every second, such real-time-bidding (RTB) systems fulfill billions of ad auctions. Other examples include recommendation algorithms used by online social media to distribute content and enhance user engagement, and dynamic pricing systems developed by platform businesses such as Uber, Lyft, and Airbnb to adjust pricing based on demand and supply. Collectively, these new business models have revolutionized how ordinary people obtain information online, make purchasing decisions, and determine the cost of products and services.

Secondly, organizations in business settings that were not traditionally tech-centric are also embracing AI to automate repetitive tasks, assist humans in making decisions, and improve employee performance. AI systems have penetrated almost every industry, from the judicial and finance system to the manufacturing, healthcare, and entertainment industries. For instance, in manufacturing, AI systems can design the optimal maintenance schedule to minimize machine downtime and improve the quality control process with computer vision technology to spot defects in real-time (Arinez et al. 2020). Fraud detection algorithms help banks and insurance companies identify risky transactions, thereby preventing losses (Ryman-Tubb et al. 2018) In healthcare, physicians utilize AI to assess patient risks and allocate medical resources. Scientists even use AI programs to discover new medicines (Yu et al. 2018). In human resource management, much of the hiring process has been automated with resume screening programs. For ordinary employees, their daily activities are also impacted as AI is used for performance monitoring and employee training, becoming a standard feature in the work environment.

1.2 Review of Business Research on AI

Management and economic research have undergone a significant evolution because of AI. On the one hand, algorithms have enriched the arsenal of methodologies, enabling the analysis of the exploding volume of unstructured data and the development of sophisticated solutions to address business questions. On the other hand, as AI adoption spreads from specific business functions to almost every industry and every task, research topics have expanded from assessing the effectiveness of individual AI applications, such as recommendation systems, to exploring human-AI collaboration and the future of work and organizations. Moreover, there is an increasing body of research on AI's economic and societal impacts. This section will briefly review the literature on AI's business value and economic impact, drawing on works from business school scholars and economists.

1.2.1 Harnessing AI's Business Value

The early presence of AI and algorithms in businesses concentrates on consumer-facing applications, sparking a large research interest, especially among marketing and IS scholars, in examining individual applications (Jacobides et al. 2021). For example, the stream of studies on recommender systems[1] suggests that they can increase sales (Oestreicher-Singer and Sundararajan 2012, Kumar and Tan 2015) and expand diversity in purchased products (Lee et al. 2020, Li et al. 2022). Personalized marketing is another effective use of AI across various marketing channels (Kumar et al. 2019, Reisenbichler et al. 2022, Tong et al. 2020). As Chatbots, either by text or voice, gained popularity as an alternative to human customer representatives, another stream of studies found that factors like the nature of the product and service (Castelo et al. 2019, Longoni and Cian 2022), and how Chatbots interact with customers (Luo et al. 2019, Schanke et al. 2021) affect the performance and consumer satisfaction.

With AI applications introduced as decision aids for internal operations, AI-augmented decision-making and Human-AI collaboration have become another research frontier. Stud-

[1]Recommendation systems could date back to the 1990s. Its evolution mirrors the development of AI from simple regression algorithms to more sophisticated systems based on deep learning.

ies incorporate qualitative observations, lab experiments, and field data to answer urgent questions, such as user perceptions of AI, what features of AI systems affect users' willingness to follow AI advice, and how humans and AI collaborate to correct each other's errors. Key concepts determining whether users follow AI's advice include users' appreciation (Logg et al. 2019) and aversion toward algorithms (Dietvorst et al. 2015). Users' own ability (Logg et al. 2019), the nature of decisions (Castelo et al. 2019), and design features of AI systems, such as explaining how AI makes predictions (Zhang et al. 2020), revealing the accuracy and confidence level (Bansal et al. 2019) and giving users control (Dietvorst et al. 2018, Kawaguchi 2021), can affect the likelihood of user adherence to algorithms. With complex decision-making processes, where domain expertise cannot be fully captured by AI yet, such as drug discovery, Allen and Choudhury (2022), Lou and Wu (2021), van den Broek et al. (2021) find that collaboration between human experts and AI, instead of humans being substituted, leads to better performance. In a knowledge-intensive business environment, early-stage evaluation of AI, two-way delegation between humans and AI, and balancing user conformity (Fügener et al. 2021, 2022, Lebovitz et al. 2021) are additional factors that require consideration.

AI's ubiquitous existence across business functions also creates vibrant research agenda in other business disciplines. Examples of operational-level business activities facilitated by AI include procurement, pricing, and supply chain management (Calvano et al. 2020, Cui et al. 2022, Mahroof 2019). Managing new dynamics in organizational structure, control, and learning (Kellogg et al. 2020, Sturm et al. 2021), innovating with AI (Burström et al. 2021, Sjödin et al. 2021) and developing an AI-oriented firm strategy (Jingyu Li et al. 2021, Raisch and Krakowski 2021) are important organizational and strategic decisions in the era of AI.

1.2.2 AI and Labor

AI's potential negative impact on the labor market is a topic that economists have heatedly discussed(Makridakis 2017, Webb 2020). Since the beginning of industrialization, automation has led to job loss and instability in the labor market (Autor 2015). The recent

wave of AI-induced automation has triggered fears of a similar effect on future employment. New methods have been proposed to map AI capabilities to occupation tasks, model the AI's impact on job displacement, and project the future of work (Acemoglu and Restrepo 2019, Agrawal et al. 2019). However, the rapid development in AI and the lack of high-quality data make estimating the impact on labor a challenging task (Frank et al. 2019).

In addition to the concerns about AI causing unemployment and short-term labor market disruption, the productivity paradox remains an unresolved question (Furman and Seamans 2019). On the one hand, state-of-the-art AI systems can perform tasks and generate predictions more accurately than humans, and economists hold a positive perspective on AI's potential (Frank et al. 2019). On the other hand, it is puzzling why the corresponding increases in productivity have not yet been observed–annual productivity growth staggered at 1.3 percent from 2005 to 2016, less than half of the rate for the decade prior (Brynjolfsson et al. 2018). Several explanations have been proposed to reconcile the discrepancy between the seemingly powerful AI capacity and the lack of productivity growth. Brynjolfsson et al. (2018) suggests that mismeasurement of productivity and implementation lags could explain why productivity growth from AI has not been captured by aggregate-level economic statistics. At the micro-firm level, the lack of AI-related organizational resources and negative employee perceptions of AI could also be why technology capacity has not been translated into productivity and performance growth (Jarrahi 2018, Tong et al. 2021). In the second essay, "Backfiring AI? Examining AI Deployment in Pay-For-Performance Regimes", I offer a new perspective to explain the AI-productivity puzzle by examining how AI deployment interacts with a firm's compensation scheme.

1.2.3 AI and Fairness

While AI has brought many positive changes, unexpected societal side effects are also becoming apparent. These concerns encompass algorithm fairness, ethical conundrums regarding data privacy, and political implications (Gallego and Kurer 2022, Tucker 2018). Whether AI systems are acting fairly is particularly concerning. AI plays an increasingly vital role in functions where domain experts traditionally held the decision-making authority.

However, the scalability and computational speed of AI-assisted decision-making systems can result in significant adverse outcomes if the system is biased against certain marginalized groups. In the judicial system, for instance, algorithms aid judges in making decisions about parole and bail. In 2016, *ProPublica* reported that many courts' risk assessment systems were racially discriminatory (Angwin et al. 2016). Healthcare is another industry in which AI has been extensively adopted. A 2019 *Science* article on racial bias in commercial algorithms has sparked a new round of debate on algorithm fairness (Obermeyer et al. 2019). As evidence of biased AI practices in other business domains, such as consumer lending, hiring, and online advertisement, emerged, there have been strong advocates for systematic investigations of algorithm fairness (Barocas and Selbst 2016, Bartlett et al. 2022).

The growing body of literature on this topic has made promising development in recent years. We find that the reasons why algorithms make biased predictions are multifaceted. The current consensus is that unrepresentative training data, programmers, and the selection of learning objectives could all contribute to bias (Chouldechova and Roth 2018, Cowgill and Tucker 2020). In Chapter 2– the essay entitled "Is Fair Advertising Good for Platforms?"– I examine the fairness issue in the online advertising industry and evaluate the welfare implications of possible solutions.

2.0 Fair Advertising

There is sufficient empirical evidence that some groups, e.g., females, are less likely to see advertisements related to economic opportunities, such as employment ads or education degree program ads. More importantly, such biases in advertisements may not be due to deliberate discrimination by advertisers. Instead, they may occur due to the nature of ad-auctions. For example, females are very lucrative customers for retailers like Macy's and Target; thus, these retailers place a very high bid in ad-auctions for female impressions and, therefore, win most of these impressions. As a result, an economic-opportunity advertiser, such as an employer, ends up showing its ad to the remaining (male) users. In this paper, we analyze some popular methods of ensuring fairness in the outcome of ad-auctions, on advertising platforms like Facebook, Google, etc. Specifically, we try to understand how these methods of fair advertising affect the incentives and welfare of various stakeholders.

A popular fairness notion in the literature, referred to as *equal-exposure* in our paper, requires the advertising platforms to artificially increase the bid of an economic-opportunity advertiser for female impressions in ad-auctions (or give away some free female impressions). The increased bid makes economic-opportunity advertisers more competitive against retailers on female impressions and ensures that both males and females are *equally exposed* to economic-opportunity ads. However, requiring a profit-maximizing platform to artificially increase the bid of an advertiser might lead to a loss of revenue for the platform. Contrary to this conventional wisdom, our results suggest that enforcing equal-exposure fairness in advertising might increase the profit of advertising platforms. This is because equal-exposure fairness intensifies the competition between an economic-opportunity advertiser and a retail advertiser (e.g., Macy's). This intensified competition leads to higher ad spending by both types of advertisers, which increases the profit of the advertising platform. This result highlights that it is in the interest of the advertising platforms to adopt equal-exposure fairness.

2.1 Introduction

With sophisticated machine learning systems penetrating every corner of our daily life, we witness a massive momentum of automating business activities with algorithms. One industry that has been dramatically transformed by technology is advertising, with digital advertising becoming a dominant marketing channel. In the United States, the spending on digital channels accounts for more than half of the $200 billion spent on advertising in 2019 (Internet Advertising Bureau and PwC 2020). Within the digital advertising ecosystem, Statista (2021) estimates that four out of five dollars are spent on algorithms, also known as programmatic advertising (i.e., real-time bidding or RTB). When a user comes online, the opportunity to show an ad to that user, also known as an impression, is sold via these ad auctions in real time. Several advertisers compete through bids in ad-auctions to show their ads to users. Big publishers like Facebook, Google, etc., can internally run ad-auctions to sell impressions of their users. On the other hand, small publishers can sell their inventory of impressions by participating in external ad-auctions on ad-exchanges, where an aggregate inventory of many small publishers is sold.

The speed and efficiency of automated ad delivery dramatically expand the processing capacity and audience base, drawing more businesses to participate. However, this exponential growth might have unintended societal side effects for some groups. It is widely reported in the popular press that women are less likely to see ads related to economic opportunities, such as employment ads. In Figure 1, we see that civil rights groups and other media outlets have expressed concerns over discrimination in advertising on several platforms like Facebook and Google. More importantly, this discrimination exists even in some heavily regulated ad categories such as housing, credit[1] and employment[2]. Researchers from several disciplines have found consistent empirical evidence for the disparity in ads in protected demographics, such as gender, race, age, and location (Lambrecht and Tucker 2019).

Advertising plays an important role in informing people about economic opportunities such as employment, education, loan, etc., (Bagwell 2007). Thus, discrimination in adver-

[1]See "When Algorithms Don't Account for Civil Rights", *The Atlantic*, Mar. 7, 2017

[2]See "Facebook Accused of Allowing Bias Against Women in Job Ads", *New York Times*, Sep. 18, 2018

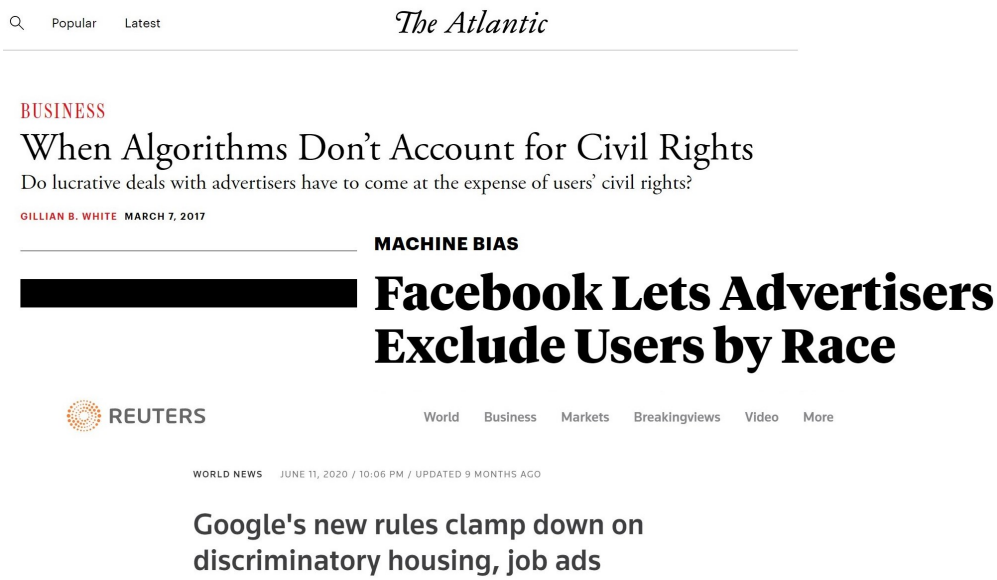


Figure 1: Media Coverage of Fairness Issues in Online Advertising

tising can have far-reaching consequences for users. For example, if women are shown fewer employment ads, they will be less informed about employment opportunities, affecting their labor participation and representation. Equal access to information is critical in ensuring the equality of opportunity (Roemer and Trannoy 2016). As more advertising continues to move online, it becomes imperative to ensure fairness in advertising. Platforms, such as Facebook, are vital information sources for people. A large part of this information is sponsored by advertisers. In this paper, we take Facebook as a running example to illustrate the reasons for the unfair outcome in advertising and the economic impact of fair advertising policies. However, our results are generalizable to settings where ad allocation happens via ad-auctions, such as ad-exchange, Google search ads, etc. From the growing body of reports and studies on digital ads, we summarize several sources of disparity in ad delivery:

- **Intentional Advertiser Bias:** Advertisers could explicitly or implicitly bias against certain protected groups. Although some platforms prohibit explicit targeting with sensitive information, malicious advertisers could still use other tools to differentiate, such as Special Ad Audience in Facebook Ads manager (Speicher et al. 2018).

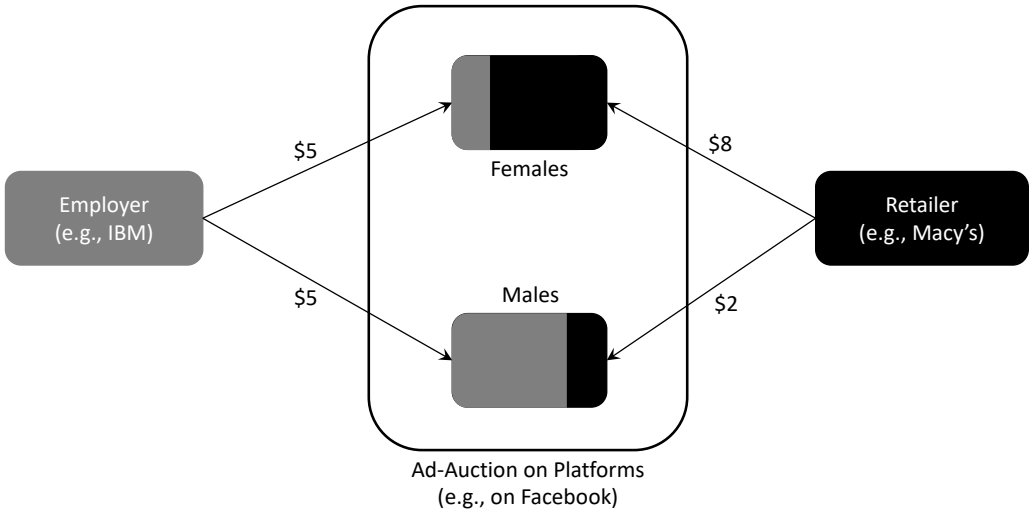
- **Unintentional Algorithmic Bias:** The algorithm that predicts how 'relevant' the ads are to any given user and decides the Real-Time Bidding (RTB) results could be biased due to either algorithm design or historical training data.
- **Economic Incentives:** Even if all parties and algorithms act fairly, the difference in economic incentives among advertisers could lead to biased ad delivery (Lambrecht and Tucker 2019). Figure 2 pictorially illustrates this unfair advertising outcome due to the difference in the preferences of advertisers. Females are considered lucrative customers for retailers. Thus, they bid disproportionately high for females compared to males in ad-auctions. In the toy example of Figure 2, the retailer bids \$8 for females and only \$2 for males. On the other hand, a neutral employment advertiser bids \$5 for both males and females. As a result, the employment advertiser loses most of the female ad-auctions and wins most of the male ad-auctions (depicted as the larger proportion of gray color in the box representing males). Therefore, the employment ads are shown mainly to males. Overall, the strong preference of retailers to target females leads to fewer employment ads shown to females.

The solution to the first source of bias roots in extensive regulations on employment, credit, and housing advertisements. In our paper, we refer to these types of ads as economic-opportunity ads because they provide information on some economic opportunities. Several measures have been taken to address bias in these economic-opportunity ads. For example, the U.S. Department of Housing and Urban Development, in the Fair Housing Act, regulates that the housing providers should not conduct selective geographic advertisements[3]. Authorities supervising other economic opportunity ads have similar provisions, such as the Federal Trade Commission (FTC) on credit opportunity ads[4] and the U.S. Equal Employment Opportunity Commission (EEOC) on employment ads[5]. There are clauses prohibiting selective advertising practices, which became the legal foundation of the lawsuits against ad platforms. Confronted with the social and legal pressure on their online advertising prac-

[3]See *U.S. Department of Housing and Urban Development* for Fair Housing Act (Sec 109.25), https://www.hud.gov/program_offices/fair_housing_equal_opp/fair_housing_act_overview

[4]See *Federal Trade Commission* for Equal Credit Opportunity Rights, https://www.consumer.ftc.gov/articles/0347-your-equal-credit-opportunity-rights

[5]See *U.S. Equal Employment Opportunity Commission* for Prohibited Employment Practices on Job Advertisements, https://www.eeoc.gov/prohibited-employment-policiespractices



A neutral employment advertiser loses in most of the female ad-auctions and wins most of the male ad-auctions. Therefore, the employment ads are shown largely to males (depicted as the larger proportion of gray color in the box representing males).

Figure 2: Retailers bid disproportionately high for females in ad-auctions

tices, Facebook and Google released product updates to prohibit discriminatory targeting using sensitive demographic information for certain types of advertisements (Facebook 2019, Google 2020).

However, Kingsley et al. (2020) find that the disparity continues even after the ad platform imposed restrictions. This is because using sensitive demographic information is one of many sources of discriminatory outcomes in online advertising. The legal perspective on fairness in advertising is slightly different from fairness in other areas (e.g., fairness in hiring). For example, all employers need to satisfy the fairness laws in hiring. However, not all advertisers must adhere to fair advertising regulations. For example, a retailer selling a feminine product is not required to show its ad to males. The advertising regulations apply only to advertisements related to employment, housing, loan, etc.

Researchers have made significant progress in developing methods to mitigate the bias from the first two sources mentioned above, i.e., the intentional advertiser bias and the unintentional algorithmic bias. In this paper, we focus on the third source of bias mentioned

above, i.e., the bias in advertising due to the economic incentives of advertisers. The existence of such biases is well known in the literature (Lambrecht and Tucker 2019), and researchers have also proposed feasible solutions to mitigate bias due to incentives. For instance, Celis et al. (2019) design and empirically test an auction mechanism that can achieve fair allocation and, at the same time, satisfy incentive compatibility. Their method introduces a 'shift' to advertisers' bids, which can be interpreted as an artificial boost. Ilvento et al. (2020) demonstrate that it is feasible to realize fair distribution of ads in a multi-category user environment. These methods of fair advertising, although mathematically robust, overlook stakeholders' incentives and strategic behavior. We contribute to the literature by developing an analytical model that decomposes the dominating business model in the online advertising industry and compares the impact of fair advertising solutions on the competition mechanism and stakeholders' welfare, thereby offering policy insights on designing incentive-compatible strategies to achieve fair advertising.

Most fairness notions can be classified into two categories: (i) *Individual fairness* notions, which require all individuals to be subject to the same standards, and (ii) *Group fairness* notions, which require equalizing a group level statistic across groups. After reviewing the current practices of online ad platforms and theoretical proposals on fair advertising, we analyze the following three methods of fair advertising:

1. **No Restriction (NR):** This is a benchmark policy where the platform doesn't enforce any fairness. We refer to this policy as the *No-Restriction* (NR) policy. The no-restriction policy gives advertisers complete control over targeting options. Before the increasing social pressure to regulate discriminatory ad placement, most ad platforms used (and some still use) this free-market business model with minimum intervention.
2. **Equal Treatment (ET):** This is an individual-level fairness method, and it is currently practiced by platforms like Facebook and Google. In this method, economic-opportunity advertisers (e.g., employment advertisers) are not allowed to target users on the basis of protected demographic attributes (e.g., gender, race). We refer to this fairness policy as the *Equal-Treatment* policy because the platform requires the economic-opportunity advertisers to treat users in different demographic groups equally. In Figure 2, the employer satisfies this fairness notion of equal-treatment by placing equal bids for both males and

females. However, as seen in this illustrative case, the outcome of the ad-auction can still be discriminatory. Since equal-treatment only ensures parity of ad-auction bids between user groups, the outcome can still be discriminatory. Nevertheless, this notion of equal-treatment fairness prevents any deliberate bias by economic-opportunity advertisers.

3. **Equal Exposure (EE):** This is a group-level fairness policy that aims to ensure that all the demographics are *equally exposed* to economic opportunity ads. For example, an equal proportion of males and females should be shown an employer's ad. Presently, this notion of fairness exists only in theory. However, it is popular among researchers and is often promoted for adoption in practice (Celis et al. 2019, Chawla and Jagadeesan 2020, Nasr and Tschantz 2020). This fairness policy can be implemented in the following two ways:
   - **Centralized Implementation:** In the centralized implementation of equal-exposure, the platform actively intervenes and ensures that all the demographic groups are equally exposed to the economic opportunity ads. The platform can ensure this by artificially increasing the bid of the economic-opportunity advertiser to make it more competitive against the retailers.
   - **Decentralized Implementation:** In the decentralized implementation, the platform delegates the responsibility of ensuring equal-exposure to economic opportunity advertisers. In this case, the equal-opportunity advertiser will have to bid higher on the protected group to ensure that all demographics are equally exposed to its ads.

   In this paper, we analyze both centralized and decentralized implementations of equal-exposure fairness policy.
4. **Equal Exposure with Equal Treatment (EET):** We propose a fourth policy that imposes both individual and group fairness. On top of the equal-treatment requirement for economic opportunity advertisers, the platform would intervene and ensure equal exposure to economic opportunity ads across demographic groups. Hence, we refer to this policy as *Equal Exposure with Equal Treatment*. This policy aims to achieve group-level fairness in a manner aligned with the legal environment.

Although there are various theoretically proven auction algorithms to achieve equal exposure, the underlying mechanism can be abstracted to subsidizing some advertisers by

adjusting bids or ad allocation results. One of the notable obstacles in promoting these fair-advertising algorithms into real business practice is the concern that subsidizing advertisers is not aligned with platforms' financial interests[6]. However, it has come to our attention that most theoretical algorithm models take advertisers' valuation of online users as exogenous and do not consider that advertisers could react strategically to the changes in the auction mechanism. Our model takes a holistic perspective of all stakeholders and provides a realistic insight into what would happen if the platform imposes *equal-exposure* constraint. Contrary to the prevailing perception that fair advertising would make advertising platforms worse off, we find that the equal-exposure policy would benefit platforms.

The primary contribution of this paper is two-fold: first, showing that the current fair-advertising policy (i.e., equal treatment policy) is not only unable to achieve a fair outcome, but it is also a bad policy in terms of the platform's profit. In fact, we show that this policy can be worse than having no fairness policy (i.e., no-restriction policy) for the platform's profit. After showing that the current fair-advertising policy doesn't work, our second main contribution is to identify the equal-exposure fairness policy from the literature and show that this policy not only achieves a fair outcome but also leads to higher profit for the platform.

2.2 Related Work

Our paper builds on and contributes to four streams of research on fairness issues in digital advertising and the design and governance of online markets. First, our findings echo the growing empirical evidence on skewed ad delivery. Due to the black-box nature of RTB algorithms, the lack of voluntary reporting on ad delivery by platforms, and the lack of well-defined fairness measurements, it is not straightforward to reach a verdict on whether discrimination exists. Many scholars use experimental methods to collect first-hand data, such as creating fictitious users or running ad campaigns, and find that online ad delivery is skewed for many protected groups. Sweeney (2013) experiments with Google Search Ads and

[6]In some of these proposals, the authors estimate that the platform would not be better off.

finds ad delivery rates vary by race. Datta et al. (2015) set up fictitious internet users to show that female users are less likely to see ads for high-paying jobs. By creating test ad campaigns, Lambrecht and Tucker (2019) show that STEM career ads disproportionately reach more male audiences even though the advertisers do not express gender preference. Recently, Facebook made the historical data public for specific ad categories via Ads Library APIs[7]. Using this official data source, Kingsley et al. (2020) find direct evidence that disparity persists among gender divisions on credit and employment ads.

Second, our theoretical model provides a framework to disentangle the mechanisms behind discrimination within online markets, a challenging research question when platforms are reluctant to open up. We identify three main reasons that have been hypothesized in the literature - *human, technology, and market*. The *human* part comes from the deliberate bias of market participants. In the case of online ads, malicious advertisers could exploit targeting tools to exclude protected users. Even though the current practices prohibit direct targeting by sensitive user features, studies have demonstrated that it is feasible to discriminate by either including features that are correlated with gender, race, or age or by using a look-alike audience tool provided by the platform (Speicher et al. 2018). Regulators and researchers have limited access to data on ad campaigns to delineate the prevalence of deliberate discrimination in ad targeting; however, constant news reporting and the settlements by Facebook over lawsuits indicate that such unlawful practices still exist. The causal evidence on the second route - *technology* - is thin, if not inconclusive. Although there is some consensus that bias could be coded into the system through training historical data (Chouldechova and Roth 2018, Cowgill and Tucker 2020), research in this direction is at its nascent stage for the digital ad industry. Ali et al. (2019) provide some evidence, though indirect, on the bias of ad delivery systems. Their experimental ad campaigns show that Facebook evaluates similar ads with different creatives to be more relevant to some demographic groups than others. The third angle to explain the disparate rate of ad exposure is the *market*, also identified as spillover effects of competition (Cowgill and Tucker 2020). In our work, we present an economic analysis that isolates the effect of market competition among advertisers on the fairness of the outcome.

[7]See https://www.facebook.com/ads/library

The third area of literature we contribute to relates to prescriptive solutions to fairness concerns with algorithms and online markets. Scholars in business and economics have started to pivot from discrimination detection to implementable measures of bias reduction. There are new algorithm designs to address unfair machine learning prediction (Kleinberg et al. 2018) and discussions about unexpected impacts of fair algorithms (Jung et al. 2020, Fu et al. 2021, Shimao et al. 2021). Within the scope of digital ad platforms, the proposals cover both market and algorithm redesign. For example, building on the seminal correspondence design on Airbnb by Edelman et al. (2017), Cui et al. (2019) pioneer an empirical attempt to reduce the bias of Airbnb hosts with better platform design. The recommendations from computer science researchers are either from the platform's perspective to incorporate different fairness notions into auction design (Celis et al. 2019, Dwork and Ilvento 2018, Chawla and Jagadeesan 2020), or they are advertiser-centered (Gelauff et al. 2020, Nasr and Tschantz 2020). Abstracted from the mathematical specifications of the proposed auction algorithm, our model compares high-level fairness notions that the platform should adopt and provides insights on incentive compatibility and social welfare.

Last, theoretical research on digital advertising platforms falls within the broader discussion of online information markets. Essentially, the indirect sale of user information makes digital advertising platforms more attractive than traditional channels (Bergemann and Bonatti 2019). Although it is beyond the scope of this paper to survey the vast literature on this topic, we summarize a few design considerations for designing digital ad markets. Previous works either focus on solving the optimal bidding strategy for one of the ecosystem participants, such as the ad exchange (Aseri et al. 2018) and advertisers (Balseiro et al. 2015, Iyer et al. 2014), or the trade-offs in designing the market, such as the tension between information disclosure policy and market thickness (Levin and Milgrom 2010, Marotta et al. 2023). Fairness concerns on market outcomes and related welfare evaluation have rarely been considered. Our study extends the understanding of the dynamics of the digital ad market under different fairness constraints.

2.3 Model Setup

To analyze the effect of fair-advertising methods, we model the competition among advertisers via bidding on ad-auctions. We focus on the strategic interactions among the following stakeholders:

1. **Platform (denoted by P):** It is the platform (e.g., Facebook) that allocates the ads via ad-auctions. In the real business setting, the platform can optimize its profit via multiple routes. For example, the platform chooses the auction design to induce truthful bids from advertisers. Theoretical models and empirical evidence show that the widely-adopted Generalized Second Price (GSP) auction mechanism can reach stable and efficient results. Due to its possession of sophisticated technology and valuable data on users' behavior and ad performance, existing literature emphasizes the platform's vital role in regulating the market to achieve a fairness requirement. In this regard, the decision and possible trade-off faced by the platform is the impact of fair advertising policies. Therefore, we assume that the platform has already adopted profit optimization strategies (such as incentive compatible auction mechanisms), and the platform's main decision is to choose the overarching fairness policy from three options identified in Section 2.1 . The focus of our analyses is on the trade-offs, if any, between the platform's financial incentive and the fairness goal.
2. **Advertisers:** Advertisers compete in ad-auctions for user attention, and we categorize them into two types:
   - **Economic Opportunity Advertiser (denoted by E)**: The first type relates to those who provide economic opportunities, such as education, financial credit, and employment. We denote this type of advertisers by E.
   - **Retail Advertisers (denoted by R)**: The other type of advertisers that compete with E comprises those who extract economic values from users directly, such as retailers (denoted by R).

   We model both advertisers as strategic players who make their decisions to maximize their profits.

3. **Users**: Users are the ad recipients. We model that the users are of two types – (i) protected users (denoted by p) and (ii) regular users (denoted by r). Protected users represent the group of users that have been historically discriminated against, e.g., women. The rest of the users are regular users. We use N_p and N_r to represent the number of impressions available for ad-auction in each user group.

2.3.1 Ad-auctions

Billions of impressions are sold every day via real-time bidding. To participate in ad auctions on platforms like Facebook or Google, an advertiser sets the targeting user attributes and a budget for a certain period (e.g., a day or a week), then lets the platforms bid and spend the budget on their behalf. Although there are various pricing schemes, the platform's revenue from advertisers could be simplified as the total bidding budgets from advertisers. This is because the most popular bidding option normally leads to the budget spent in full.[8]

We model the competition on ad auctions as a game between advertisers. This game, during a given period, plays out as follows: (1) advertisers' main decision is to set a total auction expenditure on each user group. Let e_{ij} represents the ad-auction expenditure, or total bidding budget, allocated by advertiser i on user group j, where $i \in \{E, R\},\ j \in \{p, r\}$. (2) The advertisers' profit from engaging in online advertising comes from the potential revenue when a user is exposed to the advertisement, minus the bidding expenses. Let $\alpha_{i,j}$ denote the benefit for advertiser i from a user of type j, when the user clicks or converts, where $i \in \{E, R\},\ j \in \{p, r\}$. (3) For each ad auction, we assume that the likelihood of a user clicking on one ad, or the quality of user (q), to be a constant. Later we will relax this assumption and allow users to be heterogeneous on q (please see Section 2.5.4). Without loss of generality, we normalize $q = 1$. Thus, $\alpha_{i,j}$ is the expected revenue an advertiser would receive when a user is exposed to its ad. (4) We assume that the share of total ad impressions f_{ij} of the advertiser i from group j is $\frac{e_{ij}}{e_{ij}+e_{\bar{i}j}}$, where $e_{\bar{i}j}$ is the auction total expenditure from i's competitor. This functional form of the aggregated market share is consistent with seminal papers in advertising competition (see Erickson 1985) and has also been used in recent

[8] https://www.facebook.com/business/m/one-sheeters/facebook-bid-strategy-guide

studies (Aseri et al. 2021, Dwork and Ilvento 2018, Ilvento et al. 2020). (5) The total bidding expenditure set by advertisers eventually becomes platform's revenue.

The main feature that separates protected users from regular ones is the differences in valuation from advertisers. Consistent with empirical findings, we model protected users to be more valuable for advertiser R than for E at both unit and total values, i.e., we assume $\alpha_{Rp} > \alpha_{Rr}$. On the other hand, we assume that both protected and regular users are equally valuable for the economic-opportunity advertiser (E). Specifically, we assume that $\alpha_{E,p} = \alpha_{E,r} = \alpha_E$.

To evaluate the exposure parity for ads from E, we introduce a metric for fairness level θ as the ratio of E's ad-market share between protected and regular users ($\theta = \frac{f_{Ep}}{f_{Er}}$). Figure 3 summarizes the sequence of events in the game, and Table 1 lists the notation used in our model. We now proceed to analyze fairness policies mentioned in Section 2.1: (i) No-Restriction (NR), (ii) Equal-Treatment (ET), (iii) Equal-Exposure (EE), and (iv) Equal Exposure and Treatment (EET).

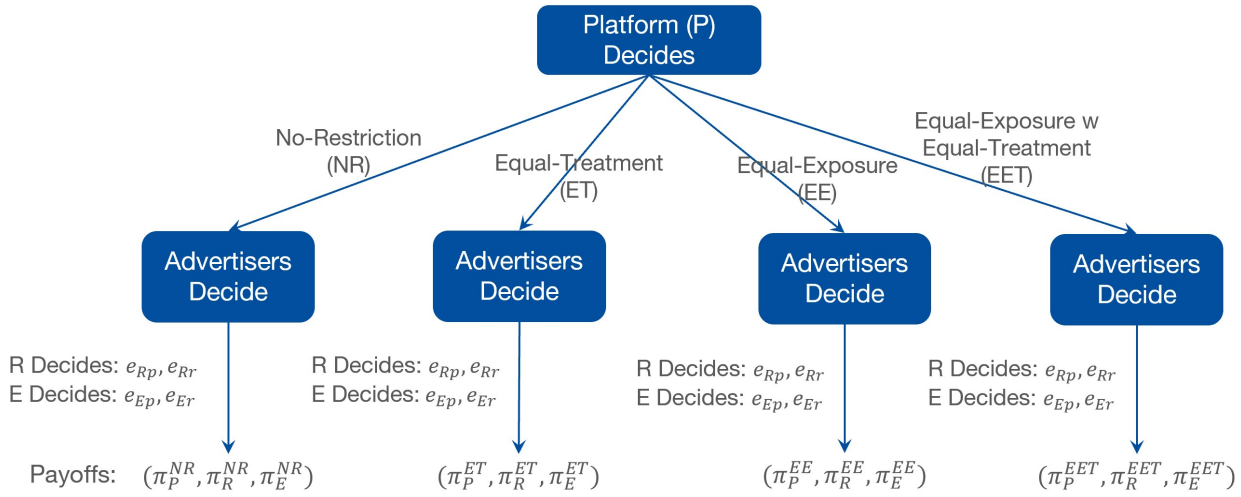


Figure 3: Sequence of Events

2.3.2 No Restriction Policy

The first policy we examine is the no-restriction case, in which the platform or the advertisers don't operate under any fairness constraint. This is the most 'natural' competition

Table 1: Fair Ads: Notations

Notation	Description
E	Index for Economic-Opportunity Advertisers (e.g., employment ads).
R	Index for Retail Advertiser (e.g., ads for clothing sales).
P	The platform that runs the online ad market.
p	Index for protected users.
r	Index for regular users.
α_{E}	Benefit to the economic-opportunity advertiser, when a user clicks on its ad.
α_{Rp}	Benefit to the retail advertiser, when a protected user clicks on its ad.
α_{Rr}	Benefit to the retail advertiser, when a regular user clicks on its ad.
N_p	Number of protected users.
N_r	Number of regular users.
e_{ij}	Ad Expenses from advertiser i on user group j, for $i \in \{E, R\}$ and $j \in \{p, r\}$.
$f_{ij}(e_{ij}, e_{\bar{i}j})$	Advertiser i's share in user group j, as a function of ad-budgets of both advertisers.
θ	Measure of Fairness: Ratio of E's ads shown to protected and regular users.
π_i	Profits of i, $i \in \{E, R, P\}$.
NR	No Restriction policy
ET	Equal Treatment policy
EE	Equal Exposure policy

condition for the online advertising industry. Before recent updates by some platforms under legal pressure from regulatory authorities and social advocacy organizations[9], most online ad markets operated under the No-Restriction model. Advertisers compete over the entire user base, and their relative ad budget levels determine the share of users shown with their ads. The cost of advertisers is equal to the sum of ad expenditures on both users group. The

[9]See lawsuit settlement between Facebook and ACLU: https://www.aclu.org/other/summary-settlements-between-civil-rights-advocates-and-facebook

profit of platform P is the sum of ad expenses spent by both advertisers. Thus, the profits of E, R, and P can be written as follows:

$$\begin{aligned}\pi_{E} &= \alpha_{E}\left(N_{p}f_{Ep}+N_{r}f_{Er}\right)-\left(e_{Ep}+e_{Er}\right)\\ &= \alpha_{E}\left(N_{p}\frac{e_{Ep}}{e_{Ep}+e_{Rp}}+N_{r}\frac{e_{Er}}{e_{Er}+e_{Rr}}\right)-\left(e_{Ep}+e_{Er}\right);\end{aligned} \tag{1}$$

$$\begin{aligned}\pi_{R} &= \left(\alpha_{Rp}N_{p}f_{Rp}+\alpha_{Rr}N_{r}f_{Rr}\right)-\left(e_{Rp}+e_{Rr}\right)\\ &= \left(\alpha_{Rp}N_{p}\frac{e_{Rp}}{e_{Ep}+e_{Rp}}+\alpha_{Rr}N_{r}\frac{e_{Rr}}{e_{Er}+e_{Rr}}\right)-\left(e_{Rp}+e_{Rr}\right);\end{aligned} \tag{2}$$

$$\pi_{P} = e_{Ep}+e_{Er}+e_{Rp}+e_{Rr}. \tag{3}$$

Since the platform exerts no constraint on advertisers' decisions, advertisers could target different demographic groups with distinct ad expenditure levels. Thus, the ad budget set by advertisers for one user group can be different from that for the other group. Thus, we can derive equilibrium values of es under the no-restriction policy from unconstrained optimization. Specifically, E and R solve the following optimization problems:

$$\max_{e_{Ep},\,e_{Er}}\pi_{E} = \max_{e_{Ep},\,e_{Er}}\alpha_{E}\left(N_{p}\frac{e_{Ep}}{e_{Ep}+e_{Rp}}+N_{r}\frac{e_{Er}}{e_{Er}+e_{Rr}}\right)-\left(e_{Ep}+e_{Er}\right); \tag{4}$$

$$\max_{e_{Rp},\,e_{Rr}}\pi_{R} = \max_{e_{Rp},\,e_{Rr}}\left(\alpha_{Rp}N_{p}\frac{e_{Rp}}{e_{Ep}+e_{Rp}}+\alpha_{Rr}N_{r}\frac{e_{Rr}}{e_{Er}+e_{Rr}}\right)-\left(e_{Rp}+e_{Rr}\right). \tag{5}$$

Solving the above optimization problems, we get the following result[10]:

Lemma 2.1. *Under the no-restriction policy, the equilibrium values of ad expense devoted by economic-opportunity advertiser (E) and the retailer (R) are*

$$e_{Ep}^{NR} = \frac{\alpha_{E}^{2}\alpha_{Rp}N_{p}}{\left(\alpha_{E}+\alpha_{Rp}\right)^{2}},\ e_{Er}^{NR} = \frac{\alpha_{E}^{2}\alpha_{Rr}N_{r}}{\left(\alpha_{E}+\alpha_{Rr}\right)^{2}},$$
$$e_{Rp}^{NR} = \frac{\alpha_{E}\alpha_{Rp}^{2}N_{p}}{\left(\alpha_{E}+\alpha_{Rp}\right)^{2}},\ e_{Rr}^{NR} = \frac{\alpha_{E}\alpha_{Rr}^{2}N_{r}}{\left(\alpha_{E}+\alpha_{Rr}\right)^{2}}.$$

The equilibrium ad market share for advertiser E are

$$f_{Ep}^{NR} = \frac{\alpha_{E}}{\alpha_{E}+\alpha_{Rp}},\ f_{Er}^{NR} = \frac{\alpha_{E}}{\alpha_{E}+\alpha_{Rr}}.$$

[10]To facilitate the interpretation of results, we add a superscription (XX) to the notation to represent the equilibrium value under policy XX. For example, e_{Ep}^{ET} is the equilibrium expense level on ad auction of advertiser E on protected group under Equal Treatment policy and θ^{NR} is the fairness level achieved under No Restriction policy.

By comparing the equilibrium values, we get the following result.

Proposition 2.1. *Under no-restriction policy, the equilibrium values of ad expenses have the following relationships:* $e_{Ep}^{NR} < e_{Rp}^{NR}$ *and* $e_{Er}^{NR} > e_{Rr}^{NR}$ *when* $\alpha_{Rr} < \alpha_E < \alpha_{Rp}$*;*

In addition, regular users would see economic opportunity ads more often than the protected users: $f_{Ep} < f_{Er}$*.*

The above proposition shows that R is willing to spend more money on the protected users than E is. Intuitively, this is because when $\alpha_{Rr} < \alpha_E < \alpha_{Rp}$, the value of showing an ad to a protected user is higher for R. Similarly, for the regular group, the ad expense spent by R is lower than that of E, i.e., $\alpha_{Rr} < \alpha_E$. Thus, the regular group is shown more ads of E. Under the no-restriction policy, the protected group is shown more ads from R. For the second half of Proposition 2.1, we find that the equilibrium ads share under the baseline policy mirrors the reality that protected users are less likely to see economic opportunity ads. Our result is consistent with the empirical findings - the differences in advertisers' economic incentives among user groups drive the disproportionate ads displayed to protected users.

2.3.3 Equal Treatment Policy

Facing increasing social and legal accusations that online ad markets enable discrimination based on sensitive user information, a few leading online ad platforms, including Facebook and Google, have tried to reduce explicit discrimination by limiting advertiser E's access to differentiate advertising based on demographic characteristics. For example, both platforms would not allow advertisers to launch an employment ad campaign that only targets one gender or racial group. This policy significantly reduces the possibility of deliberate discrimination using protected information by advertiser E. However, its impact on the fairness level of ad exposure and other players' welfare is unclear. Under the equal-treatment policy, the advertiser E needs to ensure that the per-capita budgets for protected and regular users are equal. That is,

$$\frac{e_{Ep}}{N_p} = \frac{e_{Er}}{N_r}. \tag{6}$$

Hence, the profit maximization problem of E can be written as:

$$\max_{e_{Ep},\, e_{Er}} \pi_E = \max_{e_{Ep},\, e_{Er}} \alpha_E \left(N_p \frac{e_{Ep}}{e_{Ep}+e_{Rp}} + N_r \frac{e_{Er}}{e_{Er}+e_{Rr}} \right) - (e_{Ep}+e_{Er}),$$
$$s.t., \frac{e_{Ep}}{N_p} = \frac{e_{Er}}{N_r}.$$

The profit maximization problem of R remains the same as that under no constraint:

$$\max_{e_{Rp},\, e_{Rr}} \pi_R = \max_{e_{Rp},\, e_{Rr}} \left(\alpha_{Rp} N_p \frac{e_{Rp}}{e_{Ep}+e_{Rp}} + \alpha_{Rr} N_r \frac{e_{Rr}}{e_{Er}+e_{Rr}} \right) - (e_{Rp}+e_{Rr}). \tag{7}$$

Define $B_1 = \sqrt{\alpha_{Rp}}\alpha_{Rr}N_p + \sqrt{\alpha_{Rr}}\alpha_{Rp}N_r$ and $B_2 = (\alpha_E + \alpha_{Rp})\alpha_{Rr}N_p + \alpha_{Rp}(\alpha_E + \alpha_{Rr})N_r$. Solving the above optimization problems, we get the following result:

Lemma 2.2. *Under the equal-treatment policy, the equilibrium values of ad expenses devoted by the economic-opportunity advertiser (E) and the retailer (R) are*

$$e_{Ep}^{ET} = N_p \left(\frac{B_1}{B_2}\alpha_E \right)^2,\ e_{Er}^{ET} = N_r \left(\frac{B_1}{B_2}\alpha_E \right)^2,$$
$$e_{Rp}^{ET} = \frac{B_1 \left[\sqrt{\alpha_{Rp}}\alpha_{Rr}(N_p+N_r) + \alpha_E \left(\sqrt{\alpha_{Rp}} - \sqrt{\alpha_{Rr}} \right) N_r \right]}{B_2^2} \alpha_E \alpha_{Rp} N_p,$$
$$e_{Rr}^{ET} = \frac{B_1 \left[\sqrt{\alpha_{Rr}}\alpha_{Rp}(N_p+N_r) + \alpha_E \left(\sqrt{\alpha_{Rr}} - \sqrt{\alpha_{Rp}} \right) N_p \right]}{B_2^2} \alpha_E \alpha_{Rr} N_r.$$

The equilibrium ad market share for advertiser E are

$$f_{Ep}^{ET} = \frac{\alpha_E B_1}{\sqrt{\alpha_{Rp}} B_2},\ \ f_{Er}^{ET} = \frac{\alpha_E B_1}{\sqrt{\alpha_{Rr}} B_2}.$$

Define $\hat{\alpha}_E^{ET-l} = \frac{\alpha_{Rp}\alpha_{Rr}(N_p+N_r)}{2\sqrt{\alpha_{Rp}\alpha_{Rr}}N_p - \alpha_{Rr}N_p + \alpha_{Rp}N_r}$ and $\hat{\alpha}_E^{ET-h} = \frac{\alpha_{Rp}\alpha_{Rr}(N_p+N_r)}{2\sqrt{\alpha_{Rp}\alpha_{Rr}}N_r + \alpha_{Rr}N_p - \alpha_{Rp}N_r}$. By comparing the equilibrium values of ad-budgets, we get the following result.

Proposition 2.2. *(i) Under equal-treatment policy, the equilibrium values of ad-budgets have the relationships of $e_{Ep}^{ET} < e_{Rp}^{ET}$ and $e_{Er}^{ET} > e_{Rr}^{ET}$ when $\hat{\alpha}_E^{ET-l} < \alpha_E < \hat{\alpha}_E^{ET-h}$.*

(ii) The regular users would see economic opportunity ads more often than the protected users: $f_{Ep} < f_{Er}$ when $\alpha_{Rr} < \alpha_{Rp}$.

The above proposition shows that the comparison between two advertisers' ad expenses is qualitatively similar under the equal-treatment policy to that under the no-restriction policy (in Proposition 2.1). That is, R allocates more advertising budget than E to the protected group and less ad expense than E to the regular group. The intuition of this result also remains the same as that in Proposition 2.1. That is, retailer R allocates more advertising budget to the protected group because R derives more utility from this group. We should note that the fairness constraint of ET applies only to E; thus, it makes E less competitive.

2.3.4 Equal Exposure

The analysis of the equal-treatment policy in the previous section shows that the group-level outcome can differ for different demographic groups. That is, the protected group is less exposed to the ad of E. To overcome this, we now analyze the equal-exposure fairness policy, which ensures that all demographic groups are equally exposed to the ads of the economic opportunity advertiser. As discussed in Section 2.1, the platform can implement this fairness policy in centralized and decentralized ways. In the centralized implementation, the platform actively intervenes and gives some protected users' ads for free to E, to ensure that both protected and regular groups are equally exposed to the ads of E. However, the intervention by the platform to ensure fairness might be too cumbersome for the platform in practice. Thus, we also consider a decentralized implementation, in which E needs to place different bids for different demographic groups to ensure the equal-exposure (and the platform only monitors).

2.3.4.1 Centralized Equal-Exposure (CEE):

In the centralized implementation of equal-exposure, the platform makes the reallocation criteria public to both advertisers, and states that when the protected group is less likely to see an ad of E, the platform would artificially boost advertiser E's budget on the protected group to make its ads share among the protected users equal to that among the regular ones. Next, both advertisers allocate the ad budgets between user groups freely. Mathematically, it means that when $f_{Ep} \leqslant f_{Er}$ (i.e., $\frac{e_{Ep}}{e_{Ep}+e_{Rp}} \leqslant \frac{e_{Er}}{e_{Er}+e_{Rr}}$), the platform boosts advertiser E's budget

by Δe to ensure that the protected user ads-share of E after adjustment (denoted as f^a_{Ep}) is equal to E's ad-share among the regular users (i.e., $f^a_{Ep} = f_{Er}$). Hence, the updated ads share is $f^a_{Ep} = \frac{e_{Ep}+\Delta e}{e_{Ep}+\Delta e+e_{Rp}}$. For the other possible direction of reallocation that is when $f_{Ep} > f_{Er}$, although unlikely, the platform would also make the adjustment in favor of advertiser E. The difference between $f_{Ep} \leqslant f_{Er}$ and $f_{Ep} > f_{Er}$ cases is that when $f_{Ep} \leqslant f_{Er}$, Δe would represent the free ads budget for advertiser E on "protected" users to match the exposure. On the other hand, when $f_{Ep} > f_{Er}$, Δe would represent the ads budget boost for advertiser E on "regular" users. The updated profit functions are listed in Table 2.

Table 2: Adjusted Advertisers' Profits

Case	π_R	π_E
$f_{Ep} \leqslant f_{Er}$	$\alpha_{Rp}N_p f^a_{Rp} + \alpha_{Rr}N_r f_{Rr} - (e_{Rp} + e_{Rr})$	$\alpha_E \left(N_p f^a_{Ep} + N_r f_{Er}\right) - (e_{Ep} + e_{Er})$
$f_{Ep} > f_{Er}$	$\alpha_{Rp}N_p f_{Rp} + \alpha_{Rr}N_r f^a_{Rr} - (e_{Rp} + e_{Rr})$	$\alpha_E \left(N_p f_{Ep} + N_r f^a_{Er}\right) - (e_{Ep} + e_{Er})$

We now present the profit maximization and the solution for both scenarios. When $f_{Ep} \leqslant f_{Er}$, the profit maximization problem can be specified as follows, with Δe satisfying the equal-exposure condition $\frac{e_{Ep}+\Delta e}{e_{Ep}+\Delta e+e_{Rp}} = \frac{e_{Er}}{e_{Er}+e_{Rr}}$:

$$\max_{e_{Ep},e_{Er}} \pi_E = \max_{e_{Ep},e_{Er}} \alpha_E \left(N_p \frac{e_{Ep}+\Delta e}{e_{Ep}+\Delta e+e_{Rp}} + N_r \frac{e_{Er}}{e_{Er}+e_{Rr}}\right) - (e_{Ep}+e_{Er});$$
$$\max_{e_{Rp},e_{Rr}} \pi_R = \max_{e_{Rp},e_{Rr}} \left(\alpha_{Rp}N_p \frac{e_{Rp}}{e_{Ep}+\Delta e+e_{Rp}} + \alpha_{Rr}N_r \frac{e_{Rr}}{e_{Er}+e_{Rr}}\right) - (e_{Rp}+e_{Rr}).$$

For the other case where the unadjusted ad share satisfies $f_{Ep} > f_{Er}$, we have:

$$\max_{e_{Ep},e_{Er}} \pi_E = \max_{e_{Ep},e_{Er}} \alpha_E \left(N_p \frac{e_{Ep}}{e_{Ep}+e_{Rp}} + N_p \frac{e_{Er}+\Delta e}{e_{Er}+\Delta e+e_{Rr}}\right) - (e_{Ep}+e_{Er});$$
$$\max_{e_{Rp},e_{Rr}} \pi_R = \max_{e_{Rp},e_{Rr}} \left(\alpha_{Rp}N_p \frac{e_{Rp}}{e_{Ep}+e_{Rp}} + \alpha_{Rr}N_r \frac{e_{Rr}}{e_{Er}+\Delta e+e_{Rr}}\right) - (e_{Rp}+e_{Rr}).$$

Solving the above optimization problems, we get the following results. Define $\hat{e}_E = \frac{\alpha_E^2(N_p+N_r)^2(\alpha_{Rp}N_p+\alpha_{Rr}N_r)}{[(\alpha_E+\alpha_{Rp})N_p+(\alpha_E+\alpha_{Rr})N_r]^2}$ and $\hat{e}_R = \frac{\alpha_E(N_p+N_r)(\alpha_{Rp}N_p+\alpha_{Rr}N_r)^2}{[(\alpha_E+\alpha_{Rp})N_p+(\alpha_E+\alpha_{Rr})N_r]^2}$.

Lemma 2.3. *There are two equilibria possible:*

$$(i)\ \begin{Bmatrix} e_{Ep}^{CEE}=0, & e_{Er}^{CEE}=\hat{e}_{E}, \\ e_{Rp}^{CEE}=0, & e_{Rr}^{CEE}=\hat{e}_{R} \end{Bmatrix}, \quad (ii)\ \begin{Bmatrix} e_{Ep}^{CEE}=\hat{e}_{E}, & e_{Er}^{CEE}=0, \\ e_{Rp}^{CEE}=\hat{e}_{R}, & e_{Rr}^{CEE}=0 \end{Bmatrix}. \tag{8}$$

We should note that advertiser E and R's total ad-expenditures are the same under both equilibria. That is, $e_{Ep}^{CEE} + e_{Er}^{CEE} = \hat{e}_{E}$ in equilibrium (i) is equal to $e_{Ep}^{CEE} + e_{Er}^{CEE} = \hat{e}_{E}$ in equilibrium (ii). Similarly, $e_{Rp}^{CEE} + e_{Rr}^{CEE} = \hat{e}_{R}$ in equilibrium (i) is equal to $e_{Rp}^{CEE} + e_{Rr}^{CEE} = \hat{e}_{R}$ in equilibrium (ii).

In equilibrium (i) of Lemma 2.3 above, both advertisers allocate zero budget to the protected group, and the ad-share of an advertiser (i.e., f_{Ep} & f_{Rp}) is determined by the competition for the regular users. In other words, the outcome of competition for regular users will not only determine the ad-share of advertisers for the regular users but it will also determine how much ad-share an advertiser will get from the protected users. For example, if E and R allocate \$100 and \$300 ad-budget respectively for the regular group (i.e., $e_{Er}^{CEE} = \$100$ and $e_{Rr}^{CEE} = \$300$), then E gets 25% of regular users and R gets remaining 75%. Then, E will be given 25% of the protected users to ensure equal exposure and R will get 75% of the protected users. Similarly, in equilibrium (ii) of Lemma 2.3, the exact opposite effect happens. That is, the ad-shares are determined by the competition for the protected users.

In equilibrium (i) of Lemma 2.3, on the surface it appears that advertiser E has an incentive to deviate from this equilibrium and increase its ad budget on the protected users (i.e., e_{Ep}) by a very small amount ϵ so that E could get the entire protected users. However, if E makes such a move, the setup moves to equilibrium (ii) where $f_{Ep} > f_{Er}$ and R would react accordingly, leading to equilibrium (ii) in Lemma 2.3, which has the same welfare implication for advertisers, (i.e., the advertisers' net utility remains the same in both equilibria (the detailed proof is in the Appendix). Therefore, E won't increase e_{Ep} by ϵ or deviate from the decisions presented in Lemma 2.3.

The transfer of free ads to E in equilibrium (i) resembles closely with the real-world, unlike in equilibrium (ii). Thus, we focus on equilibrium (i) for the rest of this paper. Define $\bar{\alpha}_{R} = \frac{\alpha_{Rp}N_p + \alpha_{Rr}N_r}{N_p + N_r}$ as the weighted average ad return for R. We compare the optimal values of ad-budgets in equilibrium (i) from two advertisers and get the following result:

Proposition 2.3. *Under equal-exposure policy, the equilibrium values of ad-budgets have the following relationships:* $e_{Er}^{CEE} < e_{Rr}^{CEE}$ *if* $\alpha_E < \bar{\alpha}_R$.

The intuition of the above result is as follows: The term $\bar{\alpha}_R$ represents the weighted average valuation of users by R. On the other hand, α_E is the valuation of users by E. Thus, when the weighted average valuation of R is higher than that of E, R values the users more and, therefore, spends more.

2.3.4.2 Decentralized Equal Exposure (DEE):

As mentioned in Section 2.1, there are different ways to achieve equal-exposure. The one discussed in the centralized equal-exposure emphasizes the platform's critical role in the allocation process. One critique of this approach is whether it is practical for the platform to ensure exposure parity across numerous protected features. This section discusses a decentralized approach, which requires economic opportunity advertisers to take responsibility for fair ads. Under this policy, the platform would force E to ensure equal-exposure by choosing appropriate levels of ad budget by monitoring the ad performance. Mathematically, we add a constraint on ad share to the generic profit maximization for E.

$$\max_{e_{Ep},e_{Er}} \pi_E = \max_{e_{Ep},e_{Er}} \alpha_E \left(N_p \frac{e_{Ep}}{e_{Ep}+e_{Rp}} + N_r \frac{e_{Er}}{e_{Er}+e_{Rr}} \right) - (e_{Ep}+e_{Er})$$
$$\text{s.t. } f_{Ep} = f_{Er}$$

The optimization problem for R remains as:

$$\max_{e_{Rp},\, e_{Rr}} \pi_R = \max_{e_{Rp},\, e_{Rr}} \left(\alpha_{Rp} N_p \frac{e_{Rp}}{e_{Ep}+e_{Rp}} + \alpha_{Rr} N_r \frac{e_{Rr}}{e_{Er}+e_{Rr}} \right) - (e_{Rp}+e_{Rr}).$$

Solving the constrained profit optimization, we have the following results:

Lemma 2.4. *The equilibrium values of ad expenditures from the economic-opportunity advertiser (E) and the retailer (R) are*

$$
\begin{aligned}
e_{Ep}^{DEE} =& \frac{\alpha_E^2 \alpha_{Rp} N_p \left(N_p + N_r\right)^2}{2\left[\left(\alpha_E + \alpha_{Rp}\right) N_p + \left(\alpha_E + \alpha_{Rr}\right) N_r\right]^2}, \\
e_{Er}^{DEE} =& \frac{\alpha_E^2 \alpha_{Rr} N_r \left(N_p + N_r\right)^2}{2\left[\left(\alpha_E + \alpha_{Rp}\right) N_p + \left(\alpha_E + \alpha_{Rr}\right) N_r\right]^2}, \\
e_{Rp}^{DEE} =& \frac{\alpha_E \alpha_{Rp} N_p \left(N_p + N_r\right)\left(\alpha_{Rp} N_p + \alpha_{Rr} N_r\right)}{2\left[\left(\alpha_E + \alpha_{Rp}\right) N_p + \left(\alpha_E + \alpha_{Rr}\right) N_r\right]^2}, \\
e_{Rr}^{DEE} =& \frac{\alpha_E \alpha_{Rr} N_r \left(N_p + N_r\right)\left(\alpha_{Rp} N_p + \alpha_{Rr} N_r\right)}{2\left[\left(\alpha_E + \alpha_{Rp}\right) N_p + \left(\alpha_E + \alpha_{Rr}\right) N_r\right]^2}.
\end{aligned}
$$

By comparing the equilibrium values, we have the following results:

Proposition 2.4. *Under the decentralized equal-exposure policy, the equilibrium values of ad-budgets have the following relationships*

$$
\begin{aligned}
& e_{Rr}^{DEE} < e_{Rp}^{DEE}, \; e_{Er}^{DEE} < e_{Ep}^{DEE} \\
& e_{Ep}^{DEE} < e_{Rp}^{DEE}, \; e_{Er}^{DEE} < e_{Rr}^{DEE} \; \textit{if } \alpha_E < \bar{\alpha}_R.
\end{aligned}
$$

We observe that R's budget is higher than E's on both user groups when $\alpha_E < \bar{\alpha}_R$ and the intuition behind this result is the same as that for centralized equal-exposure: when the weighted-average return for R is higher than that of E ($\alpha_E \leqslant \bar{\alpha}_R$), R exerts more effort on both user groups. Next, we present the analysis for the last policy - Equal-exposure with Equal-treatment.

2.3.5 Equal Exposure with Equal Treatment

In the implementation of equal-exposure with equal-treatment, we assume that E is also operating under equal-effort constraint, i.e., E cannot bid differently for different user groups. This is aligned with the current practice and therefore rules out any deliberate discrimination by E. The solution is very similar to the centralized equal-exposure (CEE) policy, with the main difference being that advertiser E faces the equal-treatment constraint. Therefore, we have the profit maximization and the solution as:

When $f_{Ep} \leqslant f_{Er}$, the profit maximization problem can be specified as follows, with Δe satisfying the equal-exposure condition $\frac{e_{Ep}+\Delta e}{e_{Ep}+\Delta e+e_{Rp}} = \frac{e_{Er}}{e_{Er}+e_{Rr}}$ and E's decision satisfying the equal-treatment condition $\frac{e_{Ep}}{N_p} = \frac{e_{Er}}{N_r}$:

$$\max_{e_{Ep},e_{Er}} \pi_E = \max_{e_{Ep},e_{Er}} \alpha_E \left(N_p \frac{e_{Ep}+\Delta e}{e_{Ep}+\Delta e+e_{Rp}} + N_r \frac{e_{Er}}{e_{Er}+e_{Rr}} \right) - (e_{Ep}+e_{Er});$$
$$\max_{e_{Rp},e_{Rr}} \pi_R = \max_{e_{Rp},e_{Rr}} \left(\alpha_{Rp} N_p \frac{e_{Rp}}{e_{Ep}+\Delta e+e_{Rp}} + \alpha_{Rr} N_r \frac{e_{Rr}}{e_{Er}+e_{Rr}} \right) - (e_{Rp}+e_{Rr}).$$

For the other case where the unadjusted ad share satisfies $f_{Ep} > f_{Er}$, we have:

$$\max_{e_{Ep},e_{Er}} \pi_E = \max_{e_{Ep},e_{Er}} \alpha_E \left(N_p \frac{e_{Ep}}{e_{Ep}+e_{Rp}} + N_p \frac{e_{Er}+\Delta e}{e_{Er}+\Delta e+e_{Rr}} \right) - (e_{Ep}+e_{Er});$$
$$\max_{e_{Rp},e_{Rr}} \pi_R = \max_{e_{Rp},e_{Rr}} \left(\alpha_{Rp} N_p \frac{e_{Rp}}{e_{Ep}+e_{Rp}} + \alpha_{Rr} N_r \frac{e_{Rr}}{e_{Er}+\Delta e+e_{Rr}} \right) - (e_{Rp}+e_{Rr}).$$

Solving the above two optimization problems also leads to the same solution:

Lemma 2.5. *The equilibrium values of ad budget exerted by the economic-opportunity advertiser (E) and the retailer (R) are*

$$e_{Ep}^{EET} = \frac{\alpha_E^2 N_p \left(N_p+N_r\right)^2 \left(\alpha_{Rp} N_p + \alpha_{Rr} N_r\right)}{\left[\left(\alpha_E+\alpha_{Rp}\right) N_p + \left(\alpha_E+\alpha_{Rr}\right) N_r\right]^2}, e_{Er}^{EET} = \frac{\alpha_E^2 N_r \left(N_p+N_r\right)^2 \left(\alpha_{Rp} N_p + \alpha_{Rr} N_r\right)}{\left[\left(\alpha_E+\alpha_{Rp}\right) N_p + \left(\alpha_E+\alpha_{Rr}\right) N_r\right]^2};$$
$$e_{Rp}^{EET} = \frac{\alpha_E N_p \left(\alpha_{Rp} N_p + \alpha_{Rr} N_r\right)^2}{\left[\left(\alpha_E+\alpha_{Rp}\right) N_p + \left(\alpha_E+\alpha_{Rr}\right) N_r\right]^2}, e_{Rr}^{EET} = \frac{\alpha_E N_r \left(\alpha_{Rp} N_p + \alpha_{Rr} N_r\right)^2}{\left[\left(\alpha_E+\alpha_{Rp}\right) N_p + \left(\alpha_E+\alpha_{Rr}\right) N_r\right]^2}.$$

We compare the equilibrium values of advertisers' decisions and get the following result:

Proposition 2.5. *Under the equal-exposure with equal-treatment policy, the equilibrium values of ad-budgets have the following relationships:* $e_{Ep}^{EET} < e_{Rp}^{EET},\ e_{Er}^{EET} < e_{Rr}^{EET}$ *if* $\alpha_E < \bar{\alpha}_R$.

$$\frac{e_{Ep}^{EET}}{e_{Er}^{EET}} = \frac{e_{Ep}^{EET}}{e_{Er}^{EET}} = \frac{N_p}{N_r}$$
$$e_{Ep}^{DEE} < e_{Rp}^{DEE},\ e_{Er}^{DEE} < e_{Rr}^{DEE} \text{ if } \alpha_E < \bar{\alpha}_R.$$

Proposition 2.5 reveals an interesting property of EET policy. Both advertisers allocate their bidding expenses proportionately to the population size of user groups. Therefore, the ads shares before the platform's intervention already satisfy the equal-exposure condition ($f_{Ep} = f_{Er}$). Thus, the mere existence of the platform's threat to reallocate impressions ensures equal-exposure, and no actual intervention is required from the platform. The above proposition also shows that R is willing to spend more ad-budgets on both user groups when the weighted-average return for R is higher than that of E ($\alpha_E \leqslant \bar{\alpha}_R$).

Comparing the profits of all stakeholders (P, E, and R) of three implementations equal-exposure (CEE, DEE and EET), we can make an interesting observation: Although different in their implementation details, equal-exposure measures have identical impacts on fairness and stakeholders' welfare – that is, the total ad expenses from each advertiser, the ad allocation, and the profits are all at the same level. Formally, we conclude the following proposition:

Proposition 2.6. *The profits of P, E, and R under centralized and decentralized equal-exposure measures are identical*

$$f_{Ep}^{CEE} = f_{Ep}^{DEE} = f_{Ep}^{EET},\ f_{Er}^{CEE} = f_{Er}^{DEE} = f_{Er}^{EET};$$
$$\pi_P^{CEE} = \pi_P^{DEE} = \pi_P^{EET},\ \pi_E^{CEE} = \pi_E^{DEE} = \pi_E^{EET},\ \pi_R^{CEE} = \pi_R^{DEE} = \pi_R^{EET}.$$

The main difference is the impact on competition dynamics – although the total ad budgets are the same, the allocation between user groups varies. The centralized approach incentivizes advertisers to compete on the regular group as advertiser E is motivated to increase the exposure gap between protected and regular users while R tries to narrow it down. Under the decentralized method, protected users, rather than the regular ones, become the focus of the competition. This is because the fairness constraint forces E to compete for the protected users and R will react accordingly to defend the ads share. As for equal-exposure with equal-treatment, it 'forces' both advertisers to treat every user equally.

Thus far, we have solved the problems of both E and R under different fairness policies. In the next section, we compare these fairness policies to determine which policy should be adopted by the platform and how the welfare of other stakeholders is affected by different policies.

2.4 Welfare Effect of Fairness Policies

In this section, we analyze the effect of each fairness policy on all the stakeholders involved. As we saw in the previous section, the centralized and decentralized equal-exposure policies, and equal-exposure with equal-treatment lead to mathematically the same profits for all stakeholders. We refer to these three policies as equal-exposure (EE) only in this section. We begin by analyzing the impact of fairness policies on the advertising platform.

2.4.1 Effect on the Platform's Profit

The platform is at the central stage of the debate on rectifying unfair advertising. On the one hand, the general expectation is to have platforms take more aggressive measures to alleviate the unbalanced ad delivery between user groups, in addition to the current practice of prohibiting deliberate discrimination by advertisers. On the other hand, the main arguments against fairness intervention, such as equal-exposure, are that such interventions are not aligned with the financial incentives of advertising platforms because these policies are perceived to give freebies to the economic-opportunity advertisers. Comparing the platform's profit under equal-exposure and no-restriction policies, we get the following result.

Theorem 2.1. *The platform's profit under equal-exposure policy is higher than that under no-restriction policy, i.e.,* $\pi_P^{EE} > \pi_P^{NR}$.

Contrary to conventional wisdom, the above result shows that the platform's profit is higher under the equal-exposure policy than under no-restriction (NR). The intuition behind this result is that the equal-exposure policy makes both advertisers E and R compete more fiercely. This is because, under the no-restriction policy, R spends a lot on the protected group and very low on the regular group (because R values the protected group much more than regular users). Further, because R spends less on the regular group, the neutral advertiser E finds it easier to win regular impressions, and, therefore, spends more on the regular group compared to the protected group. Overall, R largely advertises only to the protected group, and E largely advertises only to the regular group. This creates a natural differentiation between advertisers because they are focused on different user groups, which

helps them avoid competition.

The equal-exposure policy changes everything and takes away the differentiation advantage created by the no-restriction policy. When the platform implements the centralized equal-exposure policy, it promises to close any gap in exposure to E's ads by giving some free impression to E (or by artificially inflating the budget of E). This creates two opposite incentives for E and R: E is incentivized to increase the exposure gap further (to receive more free ads), and R is incentivized to decrease the exposure gap (to decrease the number of free ads). To this end, E increases its ad-budget on the regular group (to obtain even more regular impressions) and decreases its ad-budget on the protected group (to obtain even fewer protected impressions) to widen the exposure gap between the protected and the regular group. Advertiser R, on the other hand, would also increase its budget for regular users and decrease its ad-budget on the protected users to narrow down the exposure gap, which in turn reduces the number of free protected impressions reallocated to its competitor E. As a result, both advertisers compete intensely on the regular user group, unlike in the no-restriction policy, where both focus on different user groups. This lack of differentiation under equal-exposure policy leads to intense competition in the regular group, benefiting the platform.

Under the decentralized implementation of equal-exposure, the responsibility of closing the exposure gap between the protected and regular groups is on advertiser E. E can close this gap by increasing its ad-budget on the protected group and decreasing it on the regular group (or by moving ad-dollars from regular to the protected group). This increases competition on the protected group and forces R to also move its ad-budget from the regular group to the protected group. In this way, both the advertisers are again focused only on the protected group and lose the benefit of differentiation (unlike in the no-restriction policy). This lack of differentiation leads to intense competition on the protected group and benefits the platform.

When equal-exposure is implemented with the additional equal-treatment constraint on advertiser E, two forces are in play. First, two advertisers experience opposite incentives due to the platform's promise to eliminate the exposure gap. Second, the equal-treatment requirement restricts advertiser E's ability to expand the gap. As a result, both advertisers allocate the ad-budgets proportionally to the population size of user groups, and the com-

petition increase in both the protected and the regular group. This increased competition makes advertisers spend more ad-dollars on the platform, which makes the platform better off.

Another noticeable facet of our result is that Theorem 2.1 holds regardless of the parameter values. That is, no matter how much difference there is among advertisers' valuations (α), the equal-exposure policy leads to a higher profit level for the platform.

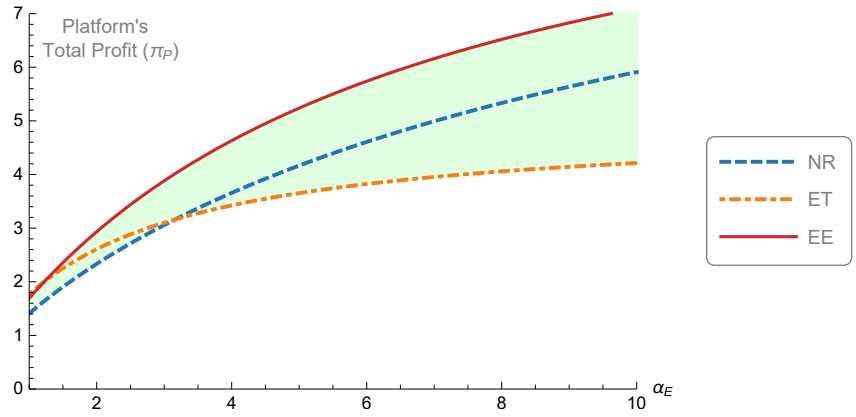
A fair advertising policy, such as equal-exposure, is proposed to remedy disparities in information dissemination in digital ad markets. Our analysis indicates that the equal-exposure policy would make advertisers compete more fiercely. Contrary to the widely feared concerns that equal-exposure fair advertising is not aligned with a profit-maximizing platform's incentives, we contend that it is in the interest of platforms to enforce equal-exposure fair advertising.

Next, we move onto the comparison of two fair advertising policies. Defining $\hat{\alpha}_{E}^{\pi_P} = \frac{\sqrt{\alpha_{Rp}\alpha_{Rr}}(\alpha_{Rp}N_p+\alpha_{Rr}N_r)}{\left(\alpha_{Rp}+\alpha_{Rr}+\sqrt{\alpha_{Rp}\alpha_{Rr}}\right)(N_p+N_r)}$, we compare the profit of the platform under equal-exposure and equal-treatment and get the following result.

Theorem 2.2. *The platform's profit is higher under equal-exposure (EE) policy compared to that under equal-treatment (ET) policy, i.e.,* $\pi_{P}^{EE} > \pi_{P}^{ET}$ *if* $\alpha_{E} > \hat{\alpha}_{E}^{\pi_P}$.

Figure 4 pictorially depicts the dominance of the equal-exposure policy over the other two policies. We also note that as E's valuation of advertising (α_{E}) increases, the platform could benefit more by implementing equal-exposure fairness. This is because when α_{E} is high, E naturally has incentives to spend more for impressions of protected users, which forces R to allocate more ad auction spending in reaction.

The intuition behind why the profit of the platform under equal-exposure is higher than that under equal-treatment at high values of α_{E} (i.e., $\alpha_{E} > \hat{\alpha}_{E}^{\pi_P}$) is as follows: Under equal-exposure, the rate of spending of advertiser E with respect to α_{E} is higher compared to that under equal-treatment. This is because under equal-exposure, advertiser E also receives some free ads to close the exposure gap. Since E spends more at a high value of α_{E}, R also does so. Consequently, as α_{E} increases, E and R both spend more, and equal-exposure becomes more profitable for the platform. Thus, at high values of α_{E}, equal-exposure leads to higher



Platform's total profit among three policies as advertiser E becomes more competitive (α_E increases), with the same standardized parameters ($N_p = N_r = \alpha_{Rr} = 1$).

Figure 4: Comparison of Platform's Profits under three policies

profit for the platform compared to equal-treatment. On the other hand, when α_E is very low, the incentive of free ads is not enough to incentivize E to spend more.

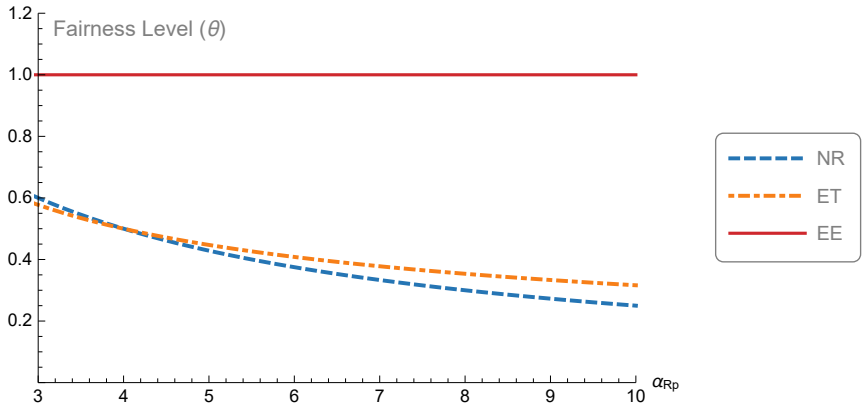
Another noticeable observation we made from Figure 4 is that for the platform, the equal-treatment policy can be worse than the baseline policy (NR). This result is mainly because, under equal-treatment, the increasing rate of advertiser E's ad spending with respect to α_E is lower than that under no-restriction. Due to R's differential valuation of two types of users, E's willingness to compete would increase at different rates for the two user groups when α_E increases. In the natural competition environment of NR, the competition for protected users intensifies naturally as α_E increases. Under the equal-treatment policy, on the other hand, advertiser E decides on the ad budget by balancing between user groups; therefore, advertiser E's overall willingness to compete is lower than that under NR when α_E is high.

2.4.2 Impact on Platform Users

An essential aspect in evaluating any market design policy is understanding to what extent the policy helps the players that it is designed to help and whether the mechanism

through which the intervention takes place has any unexpected side effects. In the case of fair advertising via equal-exposure, the protected user group is the target beneficiary of the fairness policy. We now analyze the effect of fairness policies on both user groups.

Fairness level: We evaluate the welfare of platform users from two perspectives - the relative fairness level and consumer surplus in the form of ad market share. For the relative level of fairness, we use the measurement $\theta = \frac{f_{Ep}}{f_{Er}}$ (as defined in Section 2.3). Comparing the fairness level under different policies, we find that the fairness level under the equal-exposure policy strictly dominates the fairness level achieved under other policies. Mathematically, we have $\theta^{EE} > \theta^{NR},\ \theta^{ET}$.



This plot shows that: (i) Equal-exposure policy, with the perfect fairness level of 1, outperforms other policies. (ii) The more valuable the protected users are to advertiser R compared to regular ones (that is, the larger α_{Rp} is relative to α_{Rr}), the larger the ad exposure deficit.

Figure 5: Compare the fairness level among three policies

Figure 5 plots fairness level θ against α_{Rp}. This figure illustrates the differences in fairness level θ among the three policies and leads us to the following observations. First, the equal-exposure policy, with a perfect fairness level of 1, outperforms other policies. Second, the plot confirms the intuition that the more valuable the protected users are to advertiser R compared to regular ones (that is, the larger α_{Rp} is relative to α_{Rr}), the larger the ad exposure deficit we would observe for protected users (or a lower θ value).

Comparing the values of fairness levels under different policies, we get the following

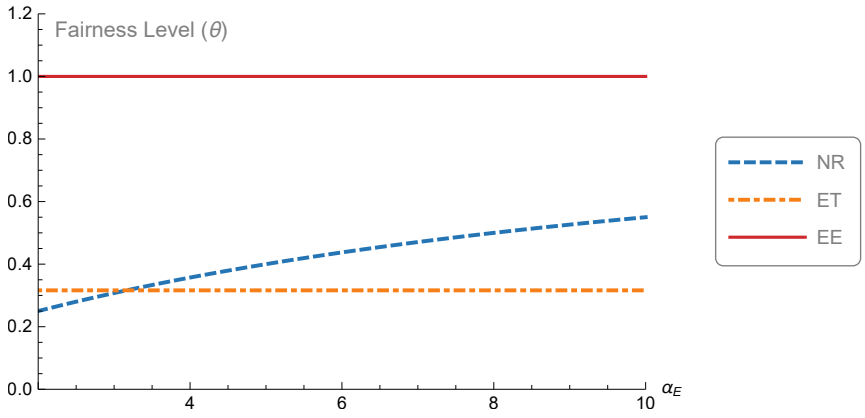


Figure 6: Comparison of fairness level among three policies

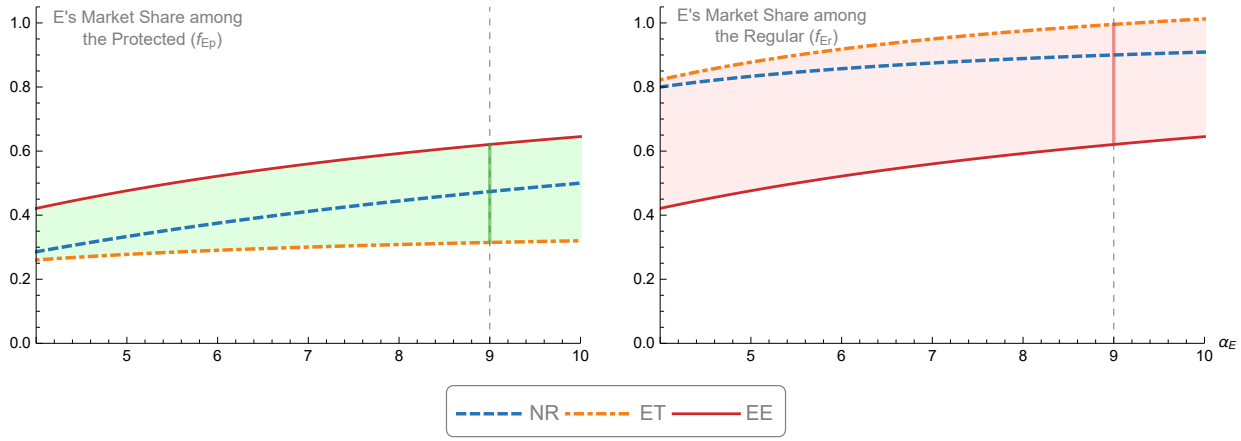
interesting result. Define $\hat{\alpha}_E^{\theta} = \sqrt{\alpha_{Rp}\alpha_{Rr}}$.

Proposition 2.7. *Equal-treatment (ET) can leads to a lower fairness level than the no-restriction (NR) policy, when* $\alpha_E > \hat{\alpha}_E^{\theta}$. *That is,* $\theta^{NR} > \theta^{ET}$ *when* $\alpha_E > \hat{\alpha}_E^{\theta}$.

The above result delivers an intriguing message that equal-treatment can lead to a lower fairness level than the baseline no-restriction policy. Figure 6 plots the fairness level (θ) against α_E and pictorially depicts this result. This is a surprising result because the equal-treatment policy consciously tries to achieve fairness, while the no-restriction policy is completely driven by market forces without regard for fairness. The intuition of this result is that when α_E is high, E is naturally capable of competing well with the retailer for the protected users without any support or intervention from the platform. In this situation, not having any constraint (NR policy) helps E compete further with the retailer. On the other hand, having a constraint of equal-treatment restricts E and hinders it from competing fiercely. Thus, no-restriction leads to higher fairness levels compared to equal-treatment (please see Figure 6).

Consumer Surplus: Apart from the fairness level θ, we are also interested in how the equal-exposure policy affects consumer welfare. We assume that a consumer obtains a utility of 1 by seeing an ad of E. On the other hand, we normalize the utility of seeing R's ad to 0. Thus,

consumer welfare equals the number of users that become aware of economic opportunities through E's ads in each segment. As the goal of the equal-exposure policy is to help advertiser E on the protected group, we expect that it will lead to a larger proportion of protected users and a lower proportion of regular users exposed to ads from advertiser E. Mathematically, we find that $f_{Ep}^{EE} > f_{Ep}^{NR}$, f_{Ep}^{ET} and $f_{Er}^{EE} < f_{Er}^{NR}$, f_{Er}^{ET} within the common parameter regions. We illustrate this result with Figure 7. When the platform switches from no-restriction or equal-treatment policy to equal-exposure, it leads to opposite effect on protected and regular groups. Advertiser E improves ad market share on protected users as the policy is designed for this purpose. However, the market share for regular users decreases for E.



These two plots show how E's share changes in each user segment as it becomes more resourceful (α_E increases), with the other parameter at $N_p = N_r = \alpha_{Rr} = 1$. The green shading area represents the market share gained on protected users under EE, while the light red shades are the loss on regular users.

Figure 7: Advertiser E's Market Share By User Groups

We are also interested in the overall welfare impact of fair-advertising policies. By comparing the total user welfare among the three policies, we have the following results. Define

$$\hat{\alpha}_E^{f_E} = \sqrt{\alpha_{Rp}\alpha_{Rr}} + \frac{\left(\sqrt{\alpha_{Rp}} + \sqrt{\alpha_{Rr}}\right)\left(\sqrt{\alpha_{Rr}}N_p + \sqrt{\alpha_{Rp}}N_r\right)}{N_p + N_r}.$$

Proposition 2.8. *The total number of users shown E's ads under the equal-exposure policy is lower than that under the other two policies if* $\alpha_E < \hat{\alpha}_E^{f_E}$.

In the regular competition environment of the online ad market, fair advertising would lead to a lower total number of people exposed to ads of E. These propositions reveal a possible trade-off between improving fairness and consumer welfare, which has been debated in other fairness settings. For protected users, fair advertising levels the playground and leads to a higher consumer surplus. However, the regular users bear the welfare loss. Even though E pours more resources on regular users, R also invests heavily in this user group. The competition in the regular segment intensifies drastically, and additional ad-budgets do not bring in market share gain for E; instead, a much lower proportion of regular users would see the ads of E.

Before concluding that the equal-exposure policy makes users worse off, we should extend the discussion to the measurement of consumer surplus. So far, we assume the value of information on economic opportunities is linear in the number of ad viewers. However, if the value of information follows the standard economics setup of diminishing marginal utility, then the increased ad exposure of the protected group, from a very low exposure before, can add a lot of consumer surplus. On the other hand, for the regular group, the loss of consumer surplus due to a decrease in exposure from a very high level of exposure can be small. The net effect of the equal-exposure policy can be positive if the gain for the protected group is higher than the loss for the regular group.

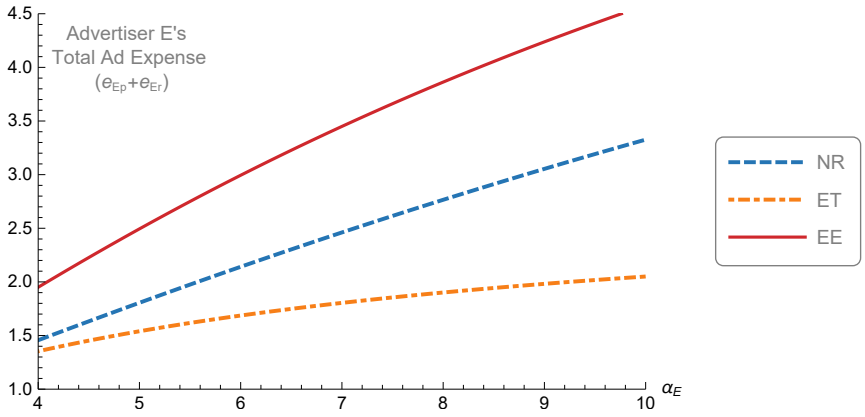
2.4.3 Impact on Advertisers' Decisions

We now analyze the impact of fair-advertising policies on both advertisers E and R. Due to the complexity of the model, it is challenging to derive results in closed form. Thus, we numerically analyze the impact of fairness policies on advertisers. First, we analyze the impact of fairness policies on the decisions of both advertisers.

2.4.3.1 E's Decisions

First, we take a closer examination at how advertiser E chooses its ad expenses level under equal-treatment. Figure 8 shows that for any given parameter value, the total ad expenses from E are at the highest level under the equal-exposure policy and the lowest

under equal-treatment policy. Also, E's ad expenditure increases with α_{E}.[11]



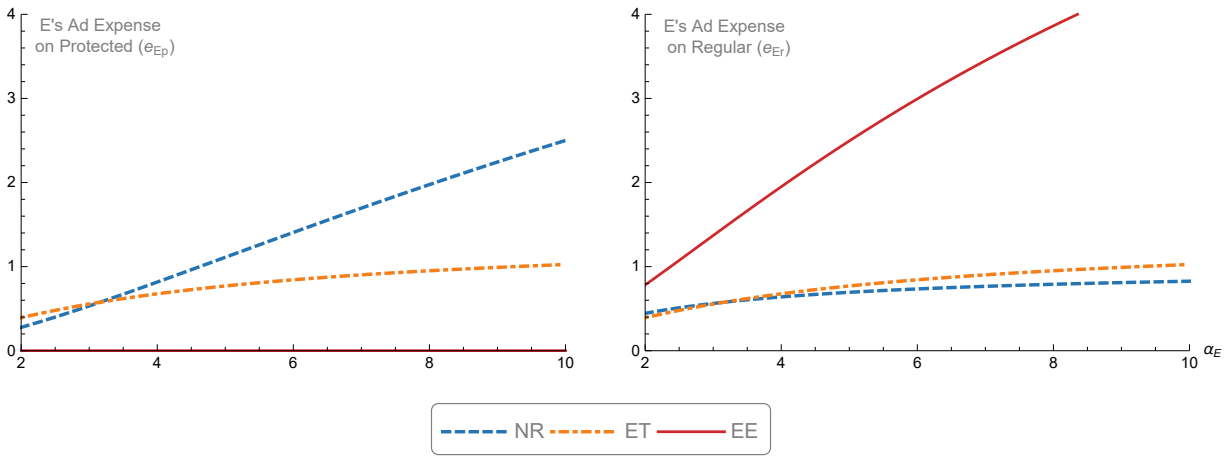
This plot shows advertiser E's total ad expenses among three policies as α_{E} increases, with standardized population and α_{Rr} ($N_p = N_r = \alpha_{Rr} = 1$)

Figure 8: Advertiser E's Total Ad Expenditures

The change in total ad expenses would assist the intuitive understanding of change in profit, while ad-budgets breakdown by user groups could offer insights on competition dynamics. Figure 9 illustrates the intuition that intensified competition in the regular group is the driving force behind the increase in E's total ad expenses. We observe that equal-treatment pushes advertiser E to boost its overall ad spending, and the higher level of the total budget is driven by the increased investment in regular users. From the numerical comparison of E's total ad expenditure and that broken down by user groups under different fair advertising policies, we make the following observation:

Observation 2.1. *Advertiser E's ad expenditure is the highest under the equal-exposure policy and this is driven by the intensified competition in the regular group.*

[11] All the numerical analysis in this section was done assuming equal-exposure as centralized equal-exposure.



This plot breaks down advertiser E's total ad expenses by user group and compares no-restriction and equal-exposure under the same standardized parameters ($N_p = N_r = \alpha_{Rr} = 1$).

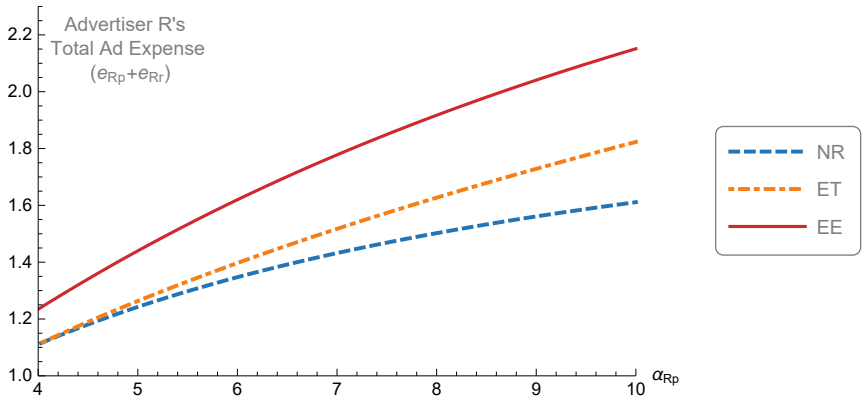
Figure 9: Advertiser E's Ad Expenses By User Groups

2.4.3.2 R's Decisions

Even though advertiser R is not at the central stage when evaluating fair advertising policy, understanding how R adapts its decisions would provide a clear picture of the shifts in competition dynamics. The process of solving for R's optimal decisions under equal-treatment offers a clue on how fair advertising would shift R's focus of competition. Platform's direct intervention is equivalent to taking away R's market share. Therefore, once R realizes that no matter how much advantage it has in the protected user segment it would be wiped away, she would choose to spend the 'wasted' money on the other user group. Hence, we expect R to steer its ad budget from the protected group to the regular one.

Observation 2.2. *Driven by the intensified competition in the regular group, advertiser R's ad expenditure is the highest under the equal-exposure policy.*

Figure 10 shows a noticeable increase in R's total ad expenses. When we dissect R's ad-budget by user segments in Figure 11, we observe that R would reallocate a significant amount of investment from protected users to the regular ones under the equal-exposure



This plot shows advertiser R's ad expenses among three policies as α_{Rp} increases, with the same standardized parameters ($N_p = N_r = \alpha_{Rr} = 1$).

Figure 10: Advertiser R's Total Ad Expenses

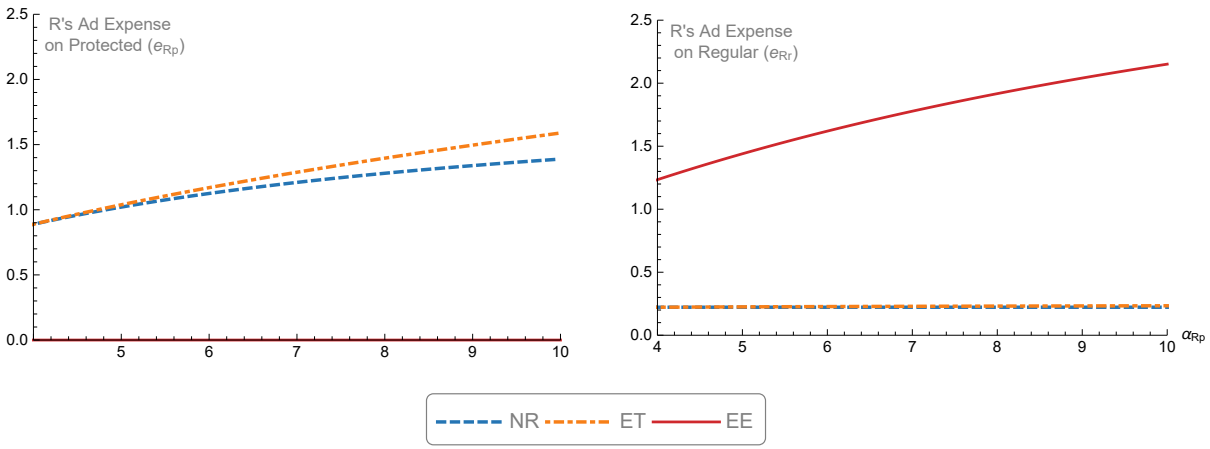
policy. Combining Figure 10 with the illustration of E's ad expenses breakdown in Figure 9, we have an overview of how the competition intensifies in the regular group. As R takes on a more aggressive position on regular users, E's advantage would extenuate even though it continues to invest heavily in the regular segment as compared to under the no-restriction and equal-treatment regimes.

We now proceed to analyze the impact of fairness policies on the profit of both advertisers.

2.4.4 Impact on Advertisers' Profit

Impact on E's profit: In this subsection, we focus on the welfare analysis of advertiser E. As we saw in the previous subsection, the equal-exposure fairness policy pushes advertiser E to invest more in protected user impressions. Also, the total number of users shown with E's ad is lower under equal-exposure policy. Thus, we expect that the total profit of E under equal-exposure will be lower than that under the other two policies.

Figure 12 confirms our intuition that E is worse off under equal-exposure policy. This



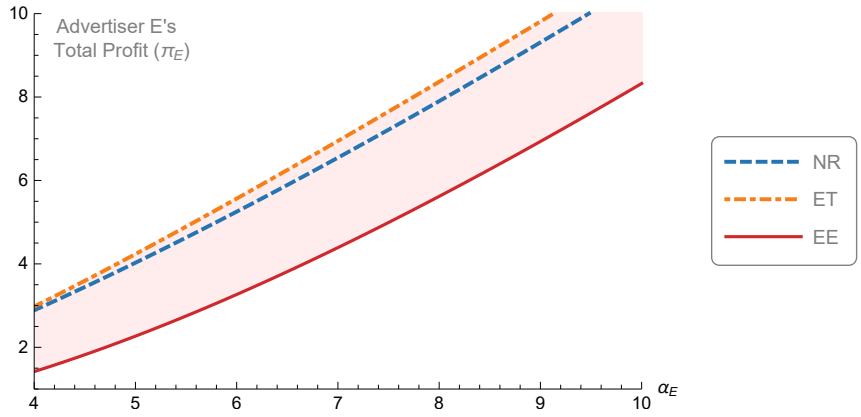
This plot breaks down advertiser R's ad expenses by user groups, with the same standardized parameters ($N_p = N_r = \alpha_{Rr} = 1$).

Figure 11: Advertiser R's Ad Expenses by User Groups

is counter-intuitive because, on the surface, the fairness policy seems to give away some impression for free to E. This presents another intriguing trade-off the policy designer must consider: between fairness and the stakeholder's profit. The decrease in advertiser E's profit is mainly due to the hyper-competitive environment for regular users' attention. The 'free' ad exposure among protected users cannot compensate for the loss in the regular segment. Although fair advertising makes the playground even for users in terms of access to information, this policy has an unfavorable impact on E's profit.

Impact on R's profit: The analysis of the policy's impact would be incomplete without assessing R's profit. The other side of the story on E's ad viewership (Figure 7) is that R loses its ground to E among protected users and gains market share on regular users. Since the protected users create more value for R on average, we expect the total revenue to decrease under equal-exposure. In terms of cost, fair advertising would force R to incur a higher level of ad expense. Therefore, we expect the overall impact on R's profit to be negative.

Figure 13 depicts the impact of equal-exposure on R's profit. It confirms our expectation



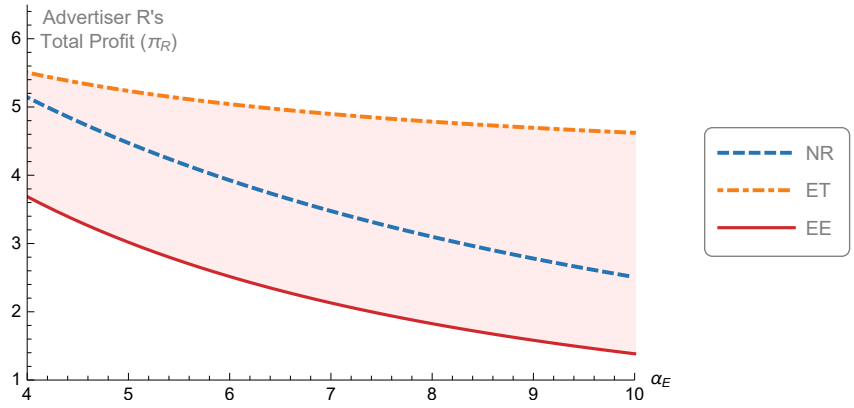
This plot shows advertiser E's total profit among three policies as it becomes more resourceful (α_E increases), with the same standardized parameters ($N_p = N_r = \alpha_{Rr} = 1$).

Figure 12: Advertiser E's Total Profit

that R would incur a reduction in profit under the equal-exposure policy. From the numerical comparisons of two advertisers' profits under different fair advertising policies, we make the following observation:

Observation 2.3. *Advertiser E and R experience a drop in their overall profit under the equal-exposure policy because of the intensified competition in the regular group.*

These plots also show that the profit gaps between equal-exposure and two existing policies expand as the competition in the protected segment intensifies (or as α_E increases). As an average ad audience becomes more valuable for E, its lead in the regular users also enlarges under equal-treatment and no-restriction policies. Under the equal-exposure policy, the platform would need to reallocate a larger portion of protected user traffic away from R. Eventually, this process leads to a more significant reduction in the retailer's profit.



This plot shows advertiser R's total profit among three policies as its competitor becomes more resourceful (α_E increases), with the same standardized parameters ($N_p = N_r = \alpha_{Rr} = 1$).

Figure 13: Advertiser R's Total Profit

2.4.5 Cost of fair advertising

From the policymaker's perspective, the equal-exposure policy implies a trade-off between fair advertising of economic opportunity information and the welfare of players in online ads. The analysis thus far shows that the benefits of the equal-exposure policy include a fair allocation of economic opportunity ads and an increase in the platform's profit. We also show that both advertisers bear the cost of fair advertising in Section 2.4.4. We numerically break down the cost of fair advertising to understand it. In Figure 14, we compare the profit losses experienced by two advertisers when the platform moved from the baseline policy (NR) to the equal-exposure policy (EE). It shows that advertiser R bears a higher level of profit loss under fair advertising.

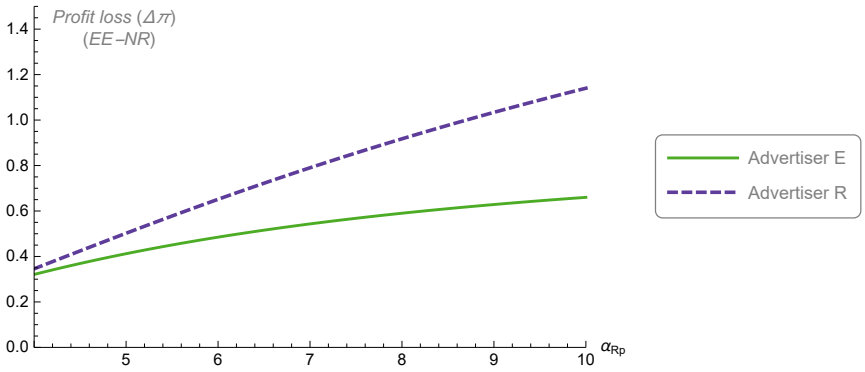


Figure 14: Compare the profit loss of two advertisers (EE vs. NR)

2.5 Robustness

In this section, we check the robustness of our main results by adding new features. Specifically, we check the robustness of our results by adding the following extensions to our base model:

- **Multiple Advertisers:** The base model describes the case where two advertisers compete on the ad platform. We now allow more than two advertisers to participate in the ad auctions.
- **Outside Option:** Our base model assumes the focal platform as a monopoly player. Thus, the advertisers cannot leave the platform and go to an alternative platform. We now introduce an outside option for advertisers.
- **Endogenize the Population of Platform Users:** Our base model assumes the size of the ad audience to be exogenously given. We now relax this assumption and model the case where the platform's decision can affect the number of users attracted to the platform.
- **User Heterogeneity:** In our base model, a user's probability of clicking on an ad, i.e., q, was the same for all the users and, therefore, did not play any role in what ad the

user will see. That is, the ad-allocation was independent of q. We now model users to be heterogeneous in quality q and allow the ad-allocation function to depend on a user's quality q.

In the following, we present the analysis of each of these extensions for the no-restriction and centralized equal-exposure policies.

2.5.1 Multiple Advertisers

We assume that N_E economic opportunity advertisers and N_R retailers compete for user attention. Advertisers of the same type are identical in their valuation of ad impressions. That is, the valuation of all N_E economic-opportunity advertisers for protected and regular users is α_E. Similarly, the valuation of all N_R retail advertisers for the protected group is α_{Rp}, and the valuation for the regular group is α_{Rr}. We introduce additional notations:

- $e_{ij}^{(n)}$: ad expenses set by the n^{th} advertiser of type i on user group j, for $n = 1, 2, \cdots, N_i$, $i \in \{E, R\}$ and $j \in \{p, r\}$.
- $f_{ij}^{(n)}$: the ad market share of the n^{th} advertiser of type i on user group j. Following the setup of the base model, the proportion of ad-impressions obtained by the n^{th} economic opportunity provider is

$$f_{Ej}^{(n)} = \frac{e_{Ej}^{(n)}}{\sum_{i=1}^{N_E} e_{Ej}^{(i)} + \sum_{i=1}^{N_R} e_{Rj}^{(i)}}$$

Since advertisers of a type are identical in terms of their valuation of users, we solve for a symmetric equilibrium in which advertisers with the same valuation allocate the same ad budget. That is, all economic-opportunity advertisers bid e_{Ep} on protected users and e_{Er} on the regular users. Similarly, all retail advertisers bid e_{Rp} and e_{Rr} on protected and regular users respectively. After solving the model, we obtain the following optimal bid values:

Lemma 2.6. *Denote* $\bar{\alpha}_E = \frac{\alpha_E}{N_E}$, $\bar{\alpha}_{Rp} = \frac{\alpha_{Rp}}{N_R}$ *and* $\bar{\alpha}_{Rr} = \frac{\alpha_{Rr}}{N_R}$, *individual advertisers' optimal*

decisions under the no-restriction are:

$$
\begin{aligned}
e_{Ep}^{NR} &= \bar{\alpha}_{E}\bar{\alpha}_{Rp}N_p\frac{(N_E+N_R-1)\left(\alpha_{E}-\alpha_{Rp}+\bar{\alpha}_{Rp}\right)}{N_E\left(\bar{\alpha}_{E}+\bar{\alpha}_{Rp}\right)^2},\\
e_{Er}^{NR} &= \bar{\alpha}_{E}\bar{\alpha}_{Rr}N_r\frac{(N_E+N_R-1)\left(\alpha_{E}-\alpha_{Rr}+\bar{\alpha}_{Rr}\right)}{N_E\left(\bar{\alpha}_{E}+\bar{\alpha}_{Rr}\right)^2},\\
e_{Rp}^{NR} &= \bar{\alpha}_{E}\bar{\alpha}_{Rp}N_p\frac{(N_E+N_R-1)\left(\alpha_{Rp}-\alpha_{E}+\bar{\alpha}_{E}\right)}{N_R\left(\bar{\alpha}_{E}+\bar{\alpha}_{Rp}\right)^2},\\
e_{Rr}^{NR} &= \bar{\alpha}_{E}\bar{\alpha}_{Rr}N_r\frac{(N_E+N_R-1)\left(\alpha_{Rr}-\alpha_{E}+\bar{\alpha}_{E}\right)}{N_R\left(\bar{\alpha}_{E}+\bar{\alpha}_{Rr}\right)^2}.
\end{aligned}
$$

Denote the maximum value that individual advertiser could achieve as $\Phi_E=\frac{\alpha_E(N_p+N_r)}{N_E}$ *and* $\Phi_R=\frac{\alpha_{Rp}N_p+\alpha_{Rr}N_r}{N_R}$*, the equilibrium values of ads expense under the equal-exposure policy can be formulated as:*

$$
\begin{aligned}
e_{Ep}^{EE}=0,\ e_{Er}^{EE} &= \frac{\Phi_E\Phi_R}{\left(\Phi_E+\Phi_R\right)^2}\frac{(N_E+N_R-1)}{N_E}\left[\left(\alpha_{E}-\alpha_{Rp}+\bar{\alpha}_{Rp}\right)N_p+\left(\alpha_{E}-\alpha_{Rr}+\bar{\alpha}_{Rr}\right)N_r\right];\\
e_{Rp}^{EE}=0,\ e_{Rr}^{EE} &= \frac{\Phi_E\Phi_R}{\left(\Phi_E+\Phi_R\right)^2}\frac{(N_E+N_R-1)}{N_R}\left[\left(\alpha_{Rp}-\alpha_{E}+\bar{\alpha}_{E}\right)N_p+\left(\alpha_{Rr}-\alpha_{E}+\bar{\alpha}_{E}\right)N_r\right].
\end{aligned}
$$

Lemma 2.7. *The platform's profit under the no-restriction and equal-exposure policies are:*

$$
\begin{aligned}
\pi_{P}^{NR} &= (N_E+N_R-1)\left[N_p\frac{\bar{\alpha}_{E}\bar{\alpha}_{Rp}}{\bar{\alpha}_{E}+\bar{\alpha}_{Rp}}+N_r\frac{\bar{\alpha}_{E}\bar{\alpha}_{Rr}}{\bar{\alpha}_{E}+\bar{\alpha}_{Rr}}\right],\\
\pi_{P}^{EE} &= (N_E+N_R-1)\frac{\Phi_E\Phi_R}{\Phi_E+\Phi_R}.
\end{aligned}
$$

Comparing the platform's profit under two policies, we conclude that our main result continues to hold. That is, the equal-exposure policy leads to a higher profit for the platform and achieves fair ad exposure between user groups. Especially, our conclusion does not depend on the number of advertisers in each type (N_E & N_R).

2.5.2 Outside Option

In the base model, we analyzed an environment with one platform and two advertisers, and the platform has no direct competition. We showed that the platform does not necessarily face the trade-off between fairness and profit because advertisers bear most of the costs. One concern, however, is that there are multiple online ad platforms in the real world, and any change in the platforms' policies could drive the advertisers away. We now extend our model by introducing an outside option for advertisers and use u_i, $i \in \{E,\ R\}$, to denote the profit advertiser i could earn if it spends the budget on another platform. Advertiser i will choose to stay with the focal platform if the realized profit π_i is higher than the outside option u_i. Otherwise, the advertiser will leave. Interestingly, we note that if one advertiser leaves the platform, the remaining one will get the entire market at zero cost. Denote ϕ_i as the maximum profit that advertiser i could achieve when i is the only player left on the platform. Thus, we have $\phi_{E} = \alpha_{E}\left(N_p + N_r\right),\ \phi_{R} = \alpha_{Rp}N_p + \alpha_{Rr}N_r$. When both advertisers choose to stay with the focal platform, their profits under equal-exposure can be rearranged and written as follows in terms of ϕ_i: $\pi_{E}^{EE} = \phi_{E}\left(\frac{\phi_E}{\phi_E+\phi_R}\right)^2,\ \pi_{R}^{EE} = \phi_{R}\left(\frac{\phi_R}{\phi_E+\phi_R}\right)^2$. Therefore, we have the following payoff matrix:

Table 3: Advertisers' Payoff with an Outside Option

	R stays	R leaves
E stays	$\left(\phi_{E}\left(\frac{\phi_E}{\phi_E+\phi_R}\right)^2,\ \phi_{R}\left(\frac{\phi_R}{\phi_E+\phi_R}\right)^2\right)$	$(\phi_{E},\ u_{R})$
E leaves	$(u_{E},\ \phi_{R})$	$(u_{E},\ u_{R})$

We assume $u_{E} \leqslant \pi_{E}^{NR}$ and $u_{R} \leqslant \pi_{R}^{NR}$. That is, both advertisers stay with the focal platform under the no-restriction (NR) policy. An interpretation of this assumption is that the outside option is an alternative for advertisers but not a perfect substitute. This is mainly because different ad platforms offer different user segments. For example, LinkedIn offers more exposure to professionals, while the younger generation prefers TikTok and Snapchat.

From the payoff matrix in Table 3, one can easily observe that there are two pure strategy Nash equilibria: (i) $(u_{E},\ \phi_{R})$ and (ii) $(\phi_{E},\ u_{R})$. That is, only one advertiser chooses to stay

on the platform, and the other one leaves. Because both advertisers have a strong incentive to stay with the platform and hope the other party to leave, a mixed strategy equilibrium should be considered (and depicts the reality more accurately). We use p_i to denote the probability for advertiser i to stay on the platform. The expected profits for advertisers and the platform are:

$$\pi_{E} = p_{E}\left[\frac{\alpha_{E}^{3}\left(N_p+N_r\right)^{3}}{\left[\left(\alpha_{E}+\alpha_{Rp}\right)N_p+\left(\alpha_{E}+\alpha_{Rr}\right)N_r\right]^{2}}p_{R}+\alpha_{E}\left(N_p+N_r\right)\left(1-p_{R}\right)\right]+\left(1-p_{E}\right)u_{E}, \quad (9)$$

$$\pi_{R} = p_{R}\left[\frac{\left(\alpha_{Rp}N_p+\alpha_{Rr}N_r\right)^{3}}{\left[\left(\alpha_{E}+\alpha_{Rp}\right)N_p+\left(\alpha_{E}+\alpha_{Rr}\right)N_r\right]^{2}}p_{E}+\alpha_{E}\left(N_p+N_r\right)\left(1-p_{E}\right)\right]+\left(1-p_{R}\right)u_{E}. \quad (10)$$

By solving for advertisers' best response function, we obtain the following results:

Lemma 2.8. *The mixed strategy probabilities of advertisers to stay on the platform are:*

$$p_{E} = \frac{1-u_{R}/\phi_{R}}{1-(\frac{\phi_{R}}{\phi_{E}+\phi_{R}})^{2}},$$
$$p_{R} = \frac{1-u_{E}/\phi_{E}}{1-(\frac{\phi_{E}}{\phi_{E}+\phi_{R}})^{2}}.$$

The profits for two advertisers and the platform under the mixed strategy equilibrium when there is an outside option are:

$$\pi_{E}^{EE} = u_{E},$$
$$\pi_{R}^{EE} = u_{R},$$
$$\pi_{P}^{EE} = \frac{1-u_{R}/\phi_{R}}{1-(\frac{\phi_{R}}{\phi_{E}+\phi_{R}})^{2}} \times \frac{1-u_{E}/\phi_{E}}{1-(\frac{\phi_{E}}{\phi_{E}+\phi_{R}})^{2}} \times \frac{\phi_{E}\phi_{R}}{\phi_{E}+\phi_{R}}.$$

The interpretation of the mixed strategy equilibrium is that both advertisers allocate resources between the focal platform and the outside option. The stable equilibrium is when the advertisers are indifferent between the expected profit from the mixed strategy and the outside option. In addition to advertisers' decisions and profits, we are most interested in whether the platform can still benefit from the equal-exposure policy in the presence of competing ad platforms. Our main finding, as stated in Theorem 2.1 continues to hold under this extension. Define

$\hat{u}_{E} = \phi_{E}\left[1-\frac{1}{1-u_{R}/\phi_{R}}\frac{(\phi_{E}+2\phi_{R})(2\phi_{E}+\phi_{R})}{(\phi_{E}+\phi_{R})^{3}}\pi_{P}^{NR}\right]$ and $\hat{u}_{R} = \phi_{R}\left[1-\frac{1}{1-u_{E}/\phi_{E}}\frac{(\phi_{E}+2\phi_{R})(2\phi_{E}+\phi_{R})}{(\phi_{E}+\phi_{R})^{3}}\pi_{P}^{NR}\right]$.

Proposition 2.9. *With the presence of an outside option for advertisers, the platform's profit under equal-exposure policy is still higher than that under no-restriction policy, i.e.,* $\pi_P^{EE} > \pi_P^{NR}$ *when* $u_E < \hat{u}_E$ *or* $u_R < \hat{u}_R$.

The above result reveals a very interesting dynamic about the behavior of competing advertisers in the presence of an outside option: Even if the advertisers can earn a higher utility by switching to the outside option, they might choose to stay with the focal platform and continue earning a lower utility. This counter-intuitive behavior is due to the fact that if one advertiser moves to the outside option the other advertiser gets the entire user audience on the focal platform at zero cost. Thus, both the advertisers choose to wait for the other one to leave and, in equilibrium, both of them stay on the focal platform unless the outside option is too attractive (i.e., either $u_E \geqslant \hat{u}_E$ or $u_R \geqslant \hat{u}_R$).

2.5.3 Endogenize user population

In the base model, we assume that the total number of ad impressions available for auction (N_p & N_r) is a constant. We now endogenize this traffic and consider that the total traffic depends on the consumer surplus from being exposed to informative ads about economic opportunities. This is because platform users gain positive utility from ads on economic opportunities.

To describe how the user base increases with consumer surplus and such a relationship affects the dynamics of fair advertising policy, we assume that a user gains utility of γ by seeing an economic opportunity ad. Without loss of generality, we normalize users' utility from seeing retail ads to zero. The followings are new parameters:

- $\overline{N_j}$: the maximum possible ad impressions available for auction in user group j, for $j \in \{p, r\}$;
- γ_j: the utility of seeing an ad E for a user in group j, for $j \in \{p, r\}$.

The ad impressions up for bid (N_j) can be modeled as $N_j = \gamma_j f_{Ej} \overline{N_j}$, with $\gamma_j f_{Ej}$ representing user j's average utility after seeing an ad. That is, the total number of ad audiences is positively associated with the total user utility. The profit functions for the two advertisers become:

$$\pi_{E} = \gamma_p f_{Ep} \overline{N_p} \alpha_{E} f_{Ep} + \gamma_r f_{Er} \overline{N_r} \alpha_{E} f_{Er} - (e_{Ep} + e_{Er}), \tag{11}$$

$$\pi_{R} = \gamma_p f_{Ep} \overline{N_p} \alpha_{Rp} (1 - f_{Ep}) + \gamma_r f_{Er} \overline{N_r} \alpha_{Rr} (1 - f_{Er}) - (e_{Rp} + e_{Rr}). \tag{12}$$

Solve for profit maximization under both no-restriction and equal-exposure policies, we obtain the following optimal values:

Lemma 2.9. *Advertisers' decision on ad expenditures under the no-restriction policy are:*

$$\begin{aligned}
e_{Ep}^{NR} &= \frac{\alpha_{E} \alpha_{Rp} \left(2\alpha_{E} + \alpha_{Rp}\right)^2}{4 \left(\alpha_{E} + \alpha_{Rp}\right)^3} \gamma_p \overline{N_p}, \\
e_{Er}^{NR} &= \frac{\alpha_{E} \alpha_{Rr} \left(2\alpha_{E} + \alpha_{Rr}\right)^2}{4 \left(\alpha_{E} + \alpha_{Rr}\right)^3} \gamma_r \overline{N_r}, \\
e_{Rp}^{NR} &= \frac{\alpha_{E} \alpha_{Rp}^2 \left(2\alpha_{E} + \alpha_{Rp}\right)}{4 \left(\alpha_{E} + \alpha_{Rp}\right)^3} \gamma_p \overline{N_p}, \\
e_{Rr}^{NR} &= \frac{\alpha_{E} \alpha_{Rr}^2 \left(2\alpha_{E} + \alpha_{Rr}\right)}{4 \left(\alpha_{E} + \alpha_{Rr}\right)^3} \gamma_r \overline{N_r}.
\end{aligned}$$

The decisions of advertisers under the equal-exposure policy are:

$$e_{Ep}^{EE} = 0,\ e_{Er}^{EE} = \frac{\alpha_{E} \left(\gamma_p \overline{N_p} + \gamma_r \overline{N_r}\right) \left(\alpha_{Rp} \gamma_p \overline{N_p} + \alpha_{Rr} \gamma_r \overline{N_r}\right) \left[\left(2\alpha_{E} + \alpha_{Rp}\right) \gamma_p \overline{N_p} + \left(2\alpha_{E} + \alpha_{Rr}\right) \gamma_r \overline{N_r}\right]^2}{4 \left[\left(\alpha_{E} + \alpha_{Rp}\right) \gamma_p \overline{N_p} + \left(\alpha_{E} + \alpha_{Rr}\right) \gamma_r \overline{N_r}\right]^3},$$

$$e_{Rp}^{EE} = 0,\ e_{Rr}^{EE} = \frac{\alpha_{E} \left(\gamma_p \overline{N_p} + \gamma_r \overline{N_r}\right) \left(\alpha_{Rp} \gamma_p \overline{N_p} + \alpha_{Rr} \gamma_r \overline{N_r}\right)^2 \left[\left(2\alpha_{E} + \alpha_{Rp}\right) \gamma_p \overline{N_p} + \left(2\alpha_{E} + \alpha_{Rr}\right) \gamma_r \overline{N_r}\right]}{4 \left[\left(\alpha_{E} + \alpha_{Rp}\right) \gamma_p \overline{N_p} + \left(\alpha_{E} + \alpha_{Rr}\right) \gamma_r \overline{N_r}\right]^3}.$$

Lemma 2.10. *The platform's profit under the no-restriction and equal-exposure policies are:*

$$\begin{aligned}
\pi_{P}^{NR} &= \frac{\alpha_{E}}{2} \left(\frac{\alpha_{Rp} \left(2\alpha_{E} + \alpha_{Rp}\right)}{\left(\alpha_{E} + \alpha_{Rp}\right)^2} \gamma_p \overline{N_p} + \frac{\alpha_{Rr} \left(2\alpha_{E} + \alpha_{Rr}\right)}{\left(\alpha_{E} + \alpha_{Rr}\right)^2} \gamma_r \overline{N_r} \right), \\
\pi_{P}^{EE} &= \frac{\alpha_{E} \left(\gamma_p \overline{N_p} + \gamma_r \overline{N_r}\right) \left(\alpha_{Rp} \gamma_p \overline{N_p} + \alpha_{Rr} \gamma_r \overline{N_r}\right) \left[\left(2\alpha_{E} + \alpha_{Rp}\right) \gamma_p \overline{N_p} + \left(2\alpha_{E} + \alpha_{Rr}\right) \gamma_r \overline{N_r}\right]}{2 \left[\left(\alpha_{E} + \alpha_{Rp}\right) \gamma_p \overline{N_p} + \left(\alpha_{E} + \alpha_{Rr}\right) \gamma_r \overline{N_r}\right]^2}.
\end{aligned}$$

The comparison of the platform's profit under the two policies confirms our main finding that the platform is better off with the equal-exposure policy.

2.5.4 User Heterogeneity

In the base model, we assume all the users from the same group are homogeneous in their 'quality' - the likelihood of clicking on the ads. Also, in the base model, the platform only considers advertisers' bids to decide the auction result, and the quality did not play any role. We now relax this assumption and allow users to be heterogeneous in quality. We assume that the quality of a user is uniformly distributed along two dimensions - $q_{\cdot}$ and $q_{\cdot}$. These qualities can be interpreted as the user's click probabilities for ads by E and R, respectively. We assume these qualities to be uniformly distributed as $q_{\cdot} \sim U[0,1],\ q_{\cdot} \sim U[0,1]$. For each ad auction, the platform can reliably estimate the quality and considers both the bid and user quality when deciding the winner. Specifically, now the probability of E's ad being shown to a user j with qualities $q_{\cdot}$ and $q_{\cdot}$ is: $\frac{q_{\cdot} e_{\cdot j}}{q_{\cdot} e_{\cdot j} + q_{\cdot} e_{\cdot j}}$. Advertisers' profit still follows the specification in the first part of Equation 1 & 2:

$$\pi_{\cdot} = \alpha_{\cdot} \left(N_p f_{\cdot p} + N_r f_{\cdot r}\right) - (e_{\cdot p} + e_{\cdot r}),$$
$$\pi_{\cdot} = (\alpha_{\cdot p} N_p f_{\cdot p} + \alpha_{\cdot r} N_r f_{\cdot r}) - (e_{\cdot p} + e_{\cdot r}).$$

However, ad market share f_{ij} requires integrating the probability of winning an ad auction over the entire distribution of user quality $q_{\cdot} \sim U[0,1],\ q_{\cdot} \sim U[0,1]$. That is,

$$f_{ij} = \int_0^1 \int_0^1 \frac{q_{ij} e_{ij}}{q_{ij} e_{ij} + q_{\bar{i}\bar{j}} e_{\bar{i}\bar{j}}} dq_{ij} dq_{\bar{i}\bar{j}}.$$

We then obtain the advertisers' profit functions as follows:

$$\begin{aligned}\pi_{\cdot} =& \frac{\alpha_{\cdot} N_p}{2}\left[1 + \frac{e_{\cdot p}}{e_{\cdot p}} \ln\left(\frac{e_{\cdot p}}{e_{\cdot p} + e_{\cdot p}}\right) + \frac{e_{\cdot p}}{e_{\cdot p}} \ln\left(\frac{e_{\cdot p} + e_{\cdot p}}{e_{\cdot p}}\right)\right] \\ &+ \frac{\alpha_{\cdot} N_r}{2}\left[1 + \frac{e_{\cdot r}}{e_{\cdot r}} \ln\left(\frac{e_{\cdot r}}{e_{\cdot r} + e_{\cdot r}}\right) + \frac{e_{\cdot r}}{e_{\cdot r}} \ln\left(\frac{e_{\cdot r} + e_{\cdot r}}{e_{\cdot r}}\right)\right] - (e_{\cdot p} + e_{\cdot r}),\end{aligned} \tag{13}$$

$$\begin{aligned}\pi_{\cdot} =& \frac{\alpha_{\cdot p} N_p}{2}\left[1 + \frac{e_{\cdot p}}{e_{\cdot p}} \ln\left(\frac{e_{\cdot p}}{e_{\cdot p} + e_{\cdot p}}\right) + \frac{e_{\cdot p}}{e_{\cdot p}} \ln\left(\frac{e_{\cdot p} + e_{\cdot p}}{e_{\cdot p}}\right)\right] \\ &+ \frac{\alpha_{\cdot r} N_r}{2}\left[1 + \frac{e_{\cdot r}}{e_{\cdot r}} \ln\left(\frac{e_{\cdot r}}{e_{\cdot r} + e_{\cdot r}}\right) + \frac{e_{\cdot r}}{e_{\cdot r}} \ln\left(\frac{e_{\cdot r} + e_{\cdot r}}{e_{\cdot r}}\right)\right] - (e_{\cdot p} + e_{\cdot r}).\end{aligned} \tag{14}$$

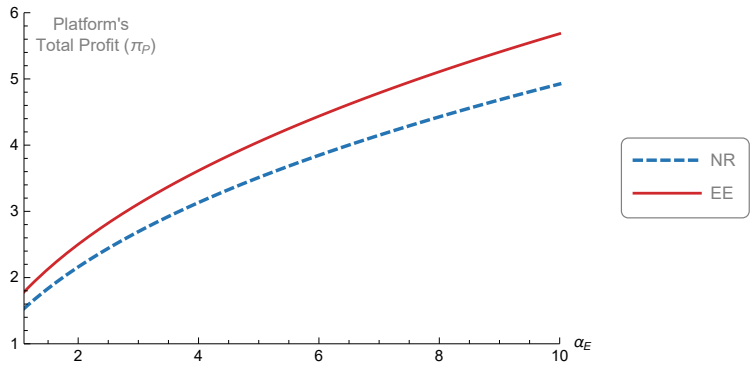
Lemma 2.11. *Advertisers' decision on ad expenditures under the no-restriction and equal-exposure policy follows the same relationship as in the base model:*

$$\frac{e_{Ep}^{NR}}{e_{Rp}^{NR}} = \frac{\alpha_E}{\alpha_{Rp}}, \frac{e_{Er}^{NR}}{e_{Rr}^{NR}} = \frac{\alpha_E}{\alpha_{Rr}},$$
$$\frac{e_{Er}^{EE}}{e_{Rr}^{EE}} = \frac{\alpha_E (N_p + N_r)}{\alpha_{Rp} N_p + \alpha_{Rr} N_r}.$$

The platform's profit under two policies are:

$$\pi_P^{NR} = \frac{(\alpha_E + \alpha_{Rp})N_p}{2}\left[\frac{\alpha_E}{\alpha_{Rp}}\ln\left(\frac{\alpha_E + \alpha_{Rp}}{\alpha_E}\right) + \frac{\alpha_{Rp}}{\alpha_E}\ln\left(\frac{\alpha_E + \alpha_{Rp}}{\alpha_{Rp}}\right) - 1\right]$$
$$+ \frac{(\alpha_E + \alpha_{Rr})N_r}{2}\left[\frac{\alpha_E}{\alpha_{Rr}}\ln\left(\frac{\alpha_E + \alpha_{Rr}}{\alpha_E}\right) + \frac{\alpha_{Rr}}{\alpha_E}\ln\left(\frac{\alpha_E + \alpha_{Rr}}{\alpha_{Rr}}\right) - 1\right]$$
$$\pi_P^{EE} = \frac{\phi_E + \phi_R}{2}\left[\frac{\phi_E}{\phi_R} ln\left(1 + \frac{\phi_R}{\phi_E}\right) + \frac{\phi_R}{\phi_E} ln\left(1 + \frac{\phi_E}{\phi_R}\right) - 1\right].$$

Numerically, we observe that our main finding (Theorem 2.1) continues to hold. As shown in Figure 15, the platform could still benefit from the equal-exposure policy.



This plot shows the platform's total profit as advertiser E becomes more competitive (α_E increases), with the same standardized parameters ($N_p = N_r = \alpha_{Rr} = 1$).

Figure 15: Comparison of Platform's Profits (NR vs. EE)

2.5.4.1 Different User Quality Distribution

The analysis thus far in this Section assumes that the quality distribution of two user groups is identical, i.e., the quality of both user groups is distributed uniformly between 0 and 1. We now relax this assumption and allow the user quality of both groups to be distributed differently. Specifically, we assume that $q_{Ej} \sim U[0, Q_{Ej}]$ and $q_{Rj} \sim U[0, Q_{Rj}]$, $j \in \{p, r\}$. Thus, we have

$$f_{ij} = \frac{1}{Q_{ij}Q_{\bar{i}j}} \int_0^{Q_{\bar{i}j}} \int_0^{Q_{ij}} \frac{q_{ij}e_{ij}}{q_{ij}e_{ij} + q_{\bar{i}j}e_{\bar{i}j}} dq_{ij}dq_{\bar{i}j},\ i \in \{E, R\},\ j \in \{p, r\}.$$

We continue to using the notation $\phi_E = \alpha_E (N_p + N_r)$, $\phi_R = \alpha_{Rp}N_p + \alpha_{Rr}N_r$ and denote $K = \frac{Q_{Er}}{Q_{Rr}}$. After solving the model, we have the following results:

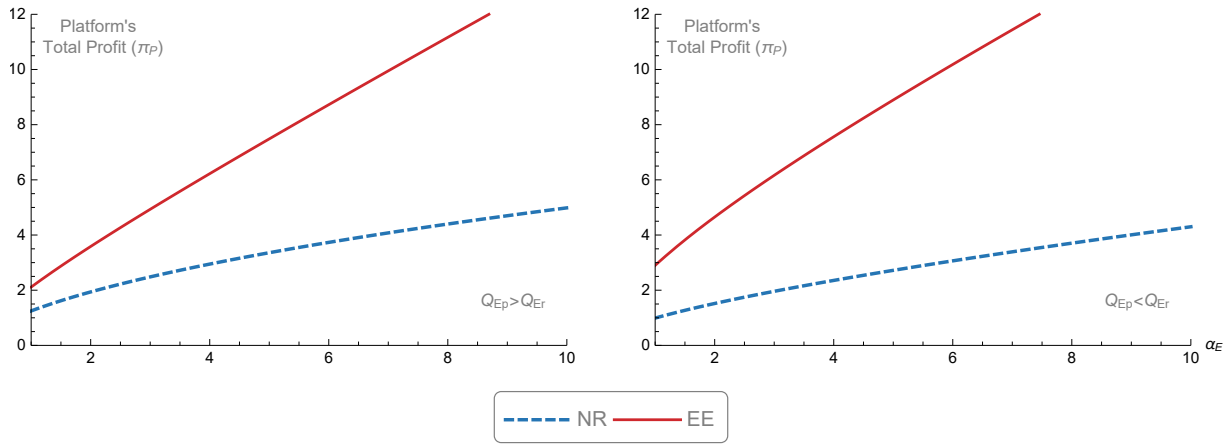
Lemma 2.12. *Advertisers' decision on ad expenditures under the no-restriction and equal-exposure policy follows the same relationship as in the base model:*

$$\frac{e_{Ep}^{NR}}{e_{Rp}^{NR}} = \frac{\alpha_E}{\alpha_{Rp}},\ \frac{e_{Er}^{NR}}{e_{Rr}^{NR}} = \frac{\alpha_E}{\alpha_{Rr}},$$
$$\frac{e_{Er}^{EE}}{e_{Rr}^{EE}} = \frac{\phi_E}{\phi_R}.$$

The platform's profit under two policies are:

$$\pi_P^{NR} = \frac{1}{2}\Bigg[(\alpha_E + \alpha_{Rp}) N_p \left(\frac{Q_{Ep}\alpha_E}{Q_{Rp}\alpha_{Rp}} ln\left(1 + \frac{Q_{Rp}\alpha_{Rp}}{Q_{Ep}\alpha_E}\right) + \frac{Q_{Rp}\alpha_{Rp}}{Q_{Ep}\alpha_E} ln\left(1 + \frac{Q_{Ep}\alpha_E}{Q_{Rp}\alpha_{Rp}}\right)\right)$$
$$+ (\alpha_E + \alpha_{Rr}) N_r \left(\frac{Q_{Er}\alpha_E}{Q_{Rr}\alpha_{Rr}} ln\left(1 + \frac{Q_{Rr}\alpha_{Rr}}{Q_{Er}\alpha_E}\right) + \frac{Q_{Rr}\alpha_{Rr}}{Q_{Er}\alpha_E} ln\left(1 + \frac{Q_{Er}\alpha_E}{Q_{Rr}\alpha_{Rr}}\right)\right) - \phi_E - \phi_R \Bigg],$$
$$\pi_P^{EE} = \frac{\phi_E + \phi_R}{2} \left[\frac{\phi_R}{K\phi_E} ln\left(1 + \frac{K\phi_E}{\phi_R}\right) + \frac{K\phi_E}{\phi_R} ln\left(1 + \frac{\phi_R}{K\phi_E}\right) - 1 \right].$$

Comparing the above two profits, in Figure 16, we numerically show that $\pi_P^{EE} > \pi_P^{NR}$ for both cases, i.e., (i) when $Q_{Ep} < Q_{Er}$ and (ii) $Q_{Ep} > Q_{Er}$.



This plot shows the platform's total profit as advertiser E becomes more competitive (α_E increases), with the same standardized parameters ($N_p = N_r = \alpha_{Rr} = 1$).

Figure 16: Comparison of Platform's Profits (NR vs EE)

2.6 Discussion & Conclusion

A fair distribution of any scarce economic resource requires that all groups are equally informed about the existence of these resources before the allocation decisions are made. For example, a fair hiring process requires that every potential hire be informed upfront about the availability of the job position. There are several channels through which this information is disseminated among potential hires. For example, for a hiring manager to advertise a job position, he can share the job position on his social media account or tell people through his personal connections. All these channels of information dissemination are skewed in favor of advantaged groups. In this work, we focus on information dissemination through the advertising channel and analyze achieving fair information dissemination through advertising. However, many of the ideas discussed in this paper are very generalizable to other information dissemination channels.

Fairness is a subjective notion; therefore, there are several methods of achieving fairness. Some methods appearing fair on the surface might result in a very unfair outcome. Similarly,

some methods that appear costly might be profit-enhancing for firms. In this paper, we analyzed three notions of fair advertising in a setting in which advertisers and the platform are strategic. The primary focus of the paper is on the fairness notion of equal-exposure, which ensures that all groups are equally exposed to the ads of an economic-opportunity advertiser (E). The platform achieves equal-exposure by giving some free ad impressions of the protected group to E. Due to these free impressions, one may suspect that the platform's profit might be lower under this policy. However, our analysis suggests exactly the opposite. The platform's profit can be higher under the equal-exposure policy. The driving force behind this result is that the equal-exposure policy makes advertisers compete more fiercely. This increased competition leads to higher spending by advertisers on the platform and increases the platform's profit.

We conclude by providing some guidance about how the equal-exposure policy can be implemented in practice:

- **Implementing Centralized-Equal-Exposure (CEE):** With the rapid rise of targeted digital advertising, advertising platforms have been actively collecting demographic information about their users. These platforms also keep track of who is seeing whose ads. To help advertisers further, platforms sometimes provide a dashboard and allow advertisers to track the demographic attributes of the users seeing their ads[12]. Using this monitoring infrastructure, the platforms can obtain the total exposure level of protected and regular users to the ad of an economic-opportunity advertiser (E), over a period of time (e.g., a day). Suppose there is a gap in the exposure level, e.g., if fewer protected users are exposed to E's ads. In the next period, the platform can allocate more impressions to protected users (e.g., by setting aside some impressions for the protected users) to close the gap. Equivalently, the platform can also artificially inflate the ad budget allocated by the economic opportunity advertiser and then use the same ad-allocation function to achieve the equal-exposure.
- **Implementing Decentralized-Equal-Exposure (DEE):** In decentralized equal-exposure, the platform allows an economic-opportunity advertiser to set different ad budgets for the protected and regular users in order to achieve equal-exposure for these two groups.

[12]https://www.facebook.com/business/m/one-sheeters/facebook-bid-strategy-guide

Then, the platform only needs to monitor that the budget of E for protected and regular groups is leading to these groups being equally exposed to E's ad.

- **Implementing Equal-Exposure with Equal-Treatment (EET):** One concern with CEE and DEE is that economic opportunity advertisers allocate different advertising budgets between user groups. Even though they are acting with a fairness intention, such disparity treatment could raise regulatory concerns. The equal-exposure with equal-treatment policy, therefore, can address this concern and ensure individual and group fairness at the same time.

Our paper shows that three equal-exposure policies are identical regarding their welfare implications for the platform, advertisers, and users. However, in centralized equal-exposure and equal-exposure with equal-treatment, the platform has to actively intervene and ensure that user groups are equally exposed to the ads economic opportunity advertisers. On the other hand, in the decentralized equal-exposure policy, the platform can delegate the responsibility of ensuring equal-exposure to advertisers. In real-world complex situations, decentralized equal-exposure can relieve platforms from the responsibility of managing fairness in a potentially large number of ad campaigns.

One of the limitations of this work is the lack of empirical support for its main findings. Advertising technology is evolving rapidly, and the literature on fairness in advertising is relatively new. Thus, regulation is yet to catch up with the societal side effects of these technologies. In the absence of such regulations, no data is available to empirically test this paper's finding. Nevertheless, as our results suggest, implementing the equal-exposure fairness policy is in the interest of the advertising platform and might increase its profit. Thus, some platforms can volunteer to implement such policies and share their insights. Apart from this, our paper can lay the foundation to motivate some future empirical research, which can potentially test the results in this paper.

3.0 AI, Salary & Productivity

Seeking value from artificial intelligence (AI) technologies, firms are rapidly deploying them for augmenting employees and improving business performance. The diffusion of AI into firms' business processes affords the firms to track tangible task actions undertaken by high-performing employees and codify best practices into recommender systems or training programs. Such AI-induced knowledge transfer has the potential to elevate overall firm performance. However, given a firm's heterogeneous workforce and its extant human resource policies, it is unclear how AI-induced knowledge transfer would impact employees' incentives and consequent performance outcomes. In this paper, using a game-theoretic model, we examine the deployment of AI in pay-for-performance regimes, in which employees are paid in proportion to their output performance (rather than a fixed salary). Our results suggest that AI deployments may backfire if firms do not account for the impact of AI on employees' incentives. This can happen because AI is good at learning tangible skills compared to intangible skills. Thus, the tangible skills of employees might be quickly learned by AI and transferred to all other employees. In an environment where employees compete with each other, this might disincentivize employees with strong tangible skills and lead to a decrease in the firm's output after AI adoption. We find that the composition of such employees in total workforce and the ability gap in employee skills, along with the knowledge transfer efficacy of AI, play a key role in impacting the overall payoffs from AI deployment in pay-for-performance regimes. Specifically, we identify conditions, such as lower ability gaps among employees and a high proportion of employees with tangible skills, where AI deployment disincentivizes employees and reduces the firm's overall profit. Based on our results, we develop policy recommendations for avoiding such pitfalls and maximizing the return on investments from AI deployments.

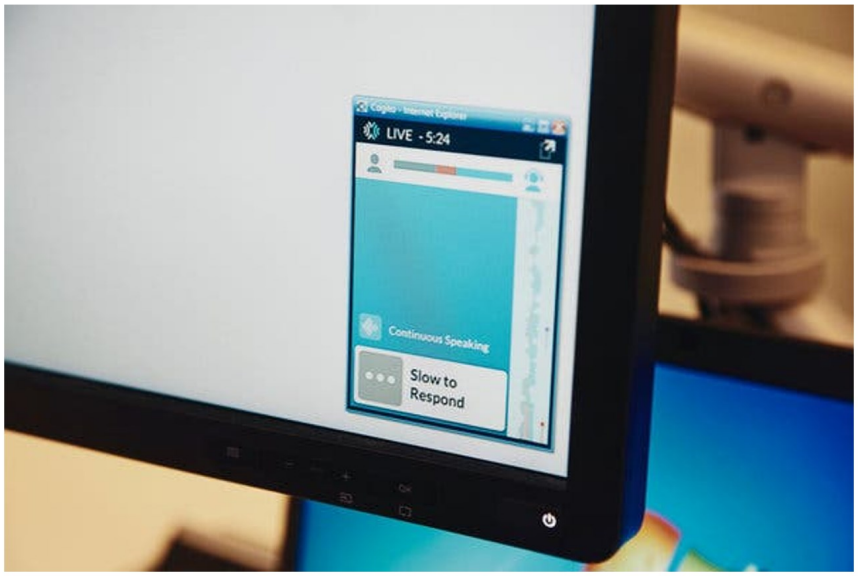
3.1 Introduction

Artificial Intelligence (AI) technologies are increasingly augmenting human workers in various tasks (Rai et al. 2019, Athey et al. 2020, Jain et al. 2021). Modern AI systems are designed to accomplish task automation using both symbolic and data-driven software models (Brynjolfsson and Mitchell 2017, Garnelo and Shanahan 2019). Symbolic models use rule-based, logical, and deductive instructions to precisely describe tasks' information-processing structure. They can be utilized to automate well-defined and repetitive task components, such as credit card processing. In contrast, data-driven models utilize statistical and machine learning algorithms for estimating outputs with inputs when tasks lack precise information-processing structures, such as making a medical diagnosis based on a variety of health data (Levy 2018). Recent advancements facilitate the combination of symbolic and data-driven models as well as the automatic and continuous improvement of performance utilizing large volumes of real-world data, which increase the affordances of AI systems and their applications to a broader array of tasks (Mitchener et al. 2022). Increasingly, such AI systems are considered general-purpose technology that catalyzes business performance improvements and innovation (Brynjolfsson et al. 2018, Furman and Seamans 2019).

Firm-level investments in AI have seen explosive growth, and a wide variety of occupations have been identified as having significant exposure to AI (Felten et al. 2018, Webb 2020, Felten et al. 2021). This has spurred discussions in the literature about the potential impact of AI on employees and their productivity (Autor 2015, Acemoglu and Restrepo 2019, Groshen and Holzer 2019). Similar to any other general-purpose technology, AI-induced automation may substitute labor. Market analysts have indeed predicted that about 15% of the global workforce, or about 400 million workers in roles such as financial advisors, medical transcriptionists, legal assistants, and customer service representatives, may be potentially impacted[1]. At the same time, AI adoption is also expected to enhance labor productivity in a variety of ways, including freeing up employees from monotonous and repetitive task components, speeding up information processing, and providing decision support for avoiding type 1 and type 2 errors (Agrawal et al. 2019, Rai et al. 2019). AI systems, especially those

[1]McKinsey Global Institute's estimates, accessed from: `https://tinyurl.com/bdewbdwm`.

with explainable predictions, have been reported to help employees enhance their learning, decision-making, and quality management activities (Senoner et al. 2021, Mele et al. 2022).



An example of AI that provides real-time feedback and helps improve employee performance.

Figure 17: AI application for employees in customer service center

Amidst this background, we raise the issue of whether the deployment of AI in firms would induce differential effects on incentives of employees in a heterogeneous workforce. Prior literature suggests that when organizations do not have the ability to precisely define and monitor requisite employee behaviors, outcome-based and pay-for-performance (PFP) compensation schemes perform better than fixed compensation schemes (Lazear 2000, Cadsby et al. 2007, Lazear 2018, He et al. 2021). Indeed, PFP is the dominant form of compensation scheme across industry sectors in the U.S., with more than 80% of the firms utilizing PFP schemes[2] (Gerhart and Fang 2014).

The impacts of PFP, however, are also contingent on other organizational and individual factors, such as competition, peer performance, and risk aversion (Stroh et al. 1996, Chan et al. 2014, Rubel and Prasad 2016, Abernethy et al. 2020). Since the deployment of AI can have profound impacts on the operating environment of organizations and the task

[2]Groysberg et al. (2021) report recent estimates across industrial sectors: `https://tinyurl.com/77cd4vbw`.

environments of employees, it is not clear if extant PFP schemes would offer the same level of economic incentive for employees to boost their performance. For instance, AI systems deployed at a firm can track and observe tangible, task-related actions of high-performing employees and codify these practices in organizational memory, which may then be utilized for training other employees. While such AI-induced transfer of know-how would benefit low-skilled employees, it may also increase the competition for performance-based rewards and dampen the incentives offered by PFP. Other factors, such as the workforce composition, ability gaps between employees, and the efficacy of AI-induced knowledge transfer, are also likely to influence the payoffs from AI deployment. In this context, we raise the research question of whether and how AI deployments in organizations that have instituted PFP schemes would improve overall employee and organizational performance. We answer the question by analytically modeling the deployment of AI in a PFP regime, and we describe the conditions under which AI deployments would be beneficial and harmful.

3.2 Model Setup

We build a game-theoretic model to understand the impact of AI adoption in a competitive environment where employees in a firm are competing with each other. Firms observe their employees and collect data about their activities. Then, this data is used to train AI algorithms. Since AI learns from employees and makes this knowledge available to all other employees, it affects the competition among employees. For example, some star employees might lose their competitive advantage. Thus, AI deployment has the potential to affect employees' incentives and productivity, which consequently impacts overall firm performance. Therefore, a firm's workforce composition, employee skill levels, and incentives are expected to influence the successful adoption of AI in the workplace. These features are at the core of our model setup.

Employees: We model that there are two types of skills that an employee needs to accomplish daily operational tasks: (i) tangible skills and (ii) intangible skills. An example of a tangible skill is knowing when to solicit a customer through email (and when not to).

For instance, for an employee in a sales team, it may be a bad idea to solicit customer leads through email on Friday afternoons because the response rates are typically lower. Similarly, a more advanced tangible skill would be to differentiate customer groups according to their preferred modes of contact and time slots and organize tasks accordingly. In the context of our paper, the defining feature of a tangible skill is that AI can learn these tangible skills from task-level data generated by all firm employees.

In contrast, we define intangible skills as those skills that are harder to be learned by AI. For example, the skill of talking a customer into buying a product or pacifying an angry customer. These are tasks with imprecise information structures, where tacit knowledge of employees may elude accurate detection by AI. We do note that AI technologies are rapidly advancing, and the current gap in AI's accuracy in discovering tangible and intangible task activities is poised to decrease in the future. Nevertheless, current AI deployments lack the affordances of general artificial intelligence, and there is a persistent gap in AI's ability to detect and leverage data related to tangible (or explicit) and intangible (or tacit) employee skills (Fjelland 2020, Heaven 2020). For simplicity of notations, we use subscripts 't' and 'i' to denote the tangible and intangible skills, respectively.

Each skill type has two levels, high and low. The high-type tangible skill is denoted by a_t^H, and the low-type tangible skill is denoted by a_t^L. Similarly, the high-type intangible skill is denoted by a_i^H, and the low-type intangible skill is denoted by a_i^L. We assume that a p_t proportion of employees have high-type tangible skills. Thus, a $1 - p_t$ proportion of employees have low-type tangible skills. Similarly, a p_i proportion of employees have high-type intangible skills, and a $1 - p_i$ proportion of employees have low-type intangible skills. Thus, there are four types of employees as listed in Table 4 (or Figure 18).

Firm's Reward Policy: Consistent with the practice, we model that the employees are paid at a salary rate per unit output. Thus, the total salary earned by an employee is the salary rate multiplied by the total output of the employee. In pay-for-performance regimes, firms usually implement a multi-tier salary (i.e., bonus or promotion for top performers) as incentive. We introduce a similar setup into our model with two reward levels – base salary-rate $\underline{\gamma}$ and bonus salary-rate $\gamma_{\scriptscriptstyle B}$, where $\gamma_{\scriptscriptstyle B} > \underline{\gamma}$. We use γ for the general notation to

Table 4: Employees

Type (Tangible, Intangible)	Ability level	Proportion
(H, H)	$(a_t^H,\ a_i^H)$	$p_{HH} = p_t p_i$
(H, L)	$(a_t^H,\ a_i^L)$	$p_{HL} = p_t(1 - p_i)$
(L, H)	$(a_t^L,\ a_i^H)$	$p_{LH} = (1 - p_t)p_i$
(L, L)	$(a_t^L,\ a_i^L)$	$p_{LL} = (1 - p_t)(1 - p_i)$

represent the salary rate. Thus,

$$\gamma = \begin{cases} \underline{\gamma}, & \text{base salary-rate,} \\ \gamma_B, & \text{bonus salary-rate.} \end{cases} \tag{15}$$

The firm cannot observe employee types, and therefore, the promotion or bonus decision – which type(s) of employees can receive the bonus salary rate – is based on the final output only. This assumption is consistent with the studies in the pay-for-performance literature (Lazear 2000, 2018). In the base model, we model the salary rates as exogenously given and unchanged after AI adoption. Later, we endogenize the firm's decision of salary rates in Section 3.6.

Employees Decisions & Utility: Each employee devotes two types of input into a task activity, tangible skill input and intangible skill input. The tangible type input is denoted by w_t, and the intangible type input is denoted by w_i. Each type of employee decides how much input to expend of each type. Employees' utility function has two components - benefit and cost. The benefit is determined by the final total output and the firm's reward policy. We assume that the final total output is a weighted sum of intangible and tangible labor inputs of employees. Let o denote the final output from an individual employee. Then, we have $o = \beta_t w_t + \beta_i w_i$ for each employee's output, where β_t and β_i are the relative importance of tangible and intangible skills in production. We also introduce O to represent the total product output, or the productivity level, of all the employees.

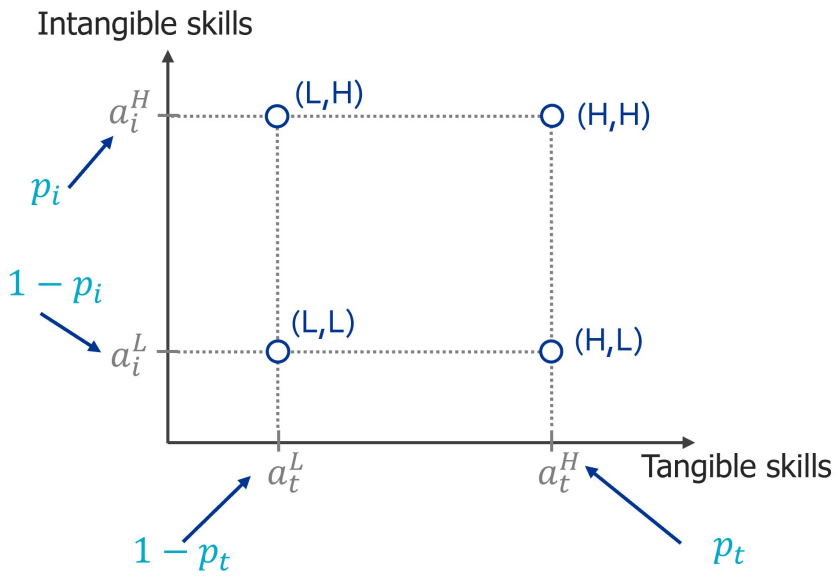


Figure 18: Illustration of four employee types by skill levels

We assume that an employee with tangible ability a_t incurs $\frac{w_t^2}{a_t}$ cost to produce w_t amount of tangible input. That is, to produce the same level of input, the employee with higher ability would incur a lower cost. Similarly, an employee with intangible ability a_i incurs $\frac{w_i^2}{a_i}$ cost to produce w_i amount of intangible input. Thus, the net utility of an employee with abilities (a_t, a_i), is

$$u(a_t, a_i) = \begin{cases} \underline{\gamma}\left(\beta_t w_t + \beta_i w_i\right) - \left(\frac{w_t^2}{a_t} + \frac{w_i^2}{a_i}\right), & \text{with base salary rate } \underline{\gamma}. \\ \gamma_{\scriptscriptstyle B}\left(\beta_t w_t + \beta_i w_i\right) - \left(\frac{w_t^2}{a_t} + \frac{w_i^2}{a_i}\right), & \text{with bonus salary rate } \gamma_{\scriptscriptstyle B}. \end{cases} \tag{16}$$

Employees are strategic players who choose the optimal labor levels (i.e., w_t and w_i) that maximize their net utility.

AI-Assisted Abilities: AI can observe employees' actions through data collection and then provide suggestions to employees on how to complete tasks with improved efficiency. It enables automatic knowledge-sharing and learning processes in daily business operations. We make a few assumptions about the learning process. First, the knowledge transferring via AI is only for the tangible skills (i.e., explicit knowledge) and from high-type employees

to low-type ones. With the help of AI, employees with low tangible skills could improve their natural ability (a_t^L). We define the new ability level with AI assistant as *effective ability* (denoted by a_{te}^L). Second, the effectiveness of AI, denoted by $f \in (0,1)$, decides the learning rate at which employees with low skill level could improve their tangible abilities. Specifically, we model the effective ability a_{te}^L as a function of the AI effectiveness rate, f, and the original ability levels:

$$a_{te}^L = a_t^L + f(a_t^H - a_t^L). \tag{17}$$

Note that a perfect AI (i.e., when $f = 1$) will improve the ability of a low-skill employee all way equal to the ability of a high-type employee. That is, when $f = 1$, we have $a_{te}^L = a_t^H$. Similarly, a completely ineffective AI (i.e., when $f = 0$) will have no effect on the ability of a low-type worker. That is, we have $a_{te}^L = a_t^L$, when $f = 0$.

Firm: The firm's revenue comes from the final output produced by its employees. The cost of the firm is the salary paid to these employees. Thus, the firm's profit can be written as

$$\pi = \sum_k \left(1 - \gamma^k\right) o^k p^k, \ k \in \{HH, \ HL, \ LH, \ LL\}, \tag{18}$$

where o^k and p^k are the output and the proportion of employee type k, and γ^k is the reward received by the employee type k. All notations are summarized in Table 5.

Our analysis assumes that employees know their own type and also know the other parameters of the model, e.g., the proportion of different types of employees, their ability level, etc. Employees also know how AI works and transfers knowledge. The firm knows the proportion of different types of employees, but it doesn't know which employee is of what type.

Table 5: AI & Productivity: Notations

Notation	Description
a_t^L	Low (L) tangible (t) ability level.
a_t^H	High (H) tangible (t) ability level.
a_i^L	Low (L) intangible (i) ability level.
a_i^H	High (H) intangible (i) ability level.
a_{te}^L	Effective ability level of employees with low tangible skills, after AI adoption.
p_t	The % proportion of employees with high tangible abilities.
p_i	The % proportion of employees with high intangible abilities.
β_t	Weightage of tangible input in the final output.
β_i	Weightage of intangible input in the final output.
o^k	Final output from labor of an employee in type k, for $k \in \{HH, HL, LH, LL\}$.
O	The total production output (or the productivity level).
u^k	Net utility of an employee in type k, for $k \in \{HH, HL, LH, LL\}$.
$\underline{\gamma}$	The base salary rate.
$\gamma_{_B}$	Bonus salary-rate, $\gamma_{_B} > \underline{\gamma}$.
f	The efficacy of AI, $0 \leqslant f \leqslant 1$.
$\pi_{_{noAI}}$	Firm's profit in the absence of AI.
$\pi_{_{AI}}$	Firm's profit in the presence of AI.

3.3 Analyses & Results

We solve the model before and after the adoption of AI and compare the result to assess the impact of AI on the welfare of different stakeholders. The sequence of the game is that the firm first announces its base and bonus salary rates. Each employee then decides how much labor of each skill type (w_t and w_i) they would devote to the production. Once the production is complete, the firm will rank employees according to their total output. Then,

the top two groups of employees will be paid a higher salary rate γ_{B}, and the remaining two groups will be paid a lower rate $\underline{\gamma}$. We model the firm's reward scheme as promoting the top two groups of employees because this is the most interesting case. Later, in the appendix, we also analyze the other case and derive the condition under which promoting the top two groups is optimal.

Using the backward induction method, we solve for the employee's decisions given the salary rates and then derive the profit for employees and the firm. We obtain the equilibrium outcome both before and after the adoption of AI and then compare these equilibria to understand the impact of AI's adoption.

3.3.1 Model Solution

Each employee decides the labor inputs based on their ability level and the expected salary. Because employees are heterogenous in their abilities, naturally, their performance also varies. Under the two-tier reward policy, the ranking of employees' output levels determines the salary rate each employee type receives. To solve for the equilibrium, we first solve the employees' problem for the case when the firm offers only one salary rate, γ (no bonus rate). This analysis will reveal how the employee output will be ranked if the firm doesn't offer a bonus. Using this analysis, we will then proceed to the case when the firm offers a bonus salary to the top two employee groups.

We now obtain employees' decisions of labor inputs (w_t and w_i). An individual employee's net utility is the total salary minus the cost of working. Thus, an employee with endowed natural ability ($a_t,\ a_i$) a salary rate γ will get the net utility of

$$u = \max_{w_t, w_i} \gamma \left(\beta_t w_t + \beta_i w_i\right) - \left(\frac{w_t^2}{a_t} + \frac{w_i^2}{a_i}\right), \tag{19}$$

where $a_t \in \{a_t^H, a_t^L, a_{te}^L\},\ a_i \in \{a_i^H, a_i^L\}$ and $\gamma \in \{\underline{\gamma}, \gamma_{B}\}$.

Solving the above optimization problem of an employee in (19), we obtain the following results:

Lemma 3.1. *The employees' optimal choices of labor, equilibrium final output, and net utility are*

$$w_t^* = \frac{1}{2}\beta_t a_t \gamma, \quad w_i^* = \frac{1}{2}\beta_i a_i \gamma, \quad u^* = \frac{1}{4}\left(\beta_t^2 a_t + \beta_i^2 a_i\right)\gamma^2.$$
$$o^* = \beta_t w_t^* + \beta_i w_i^* = \frac{1}{2}\left(\beta_t^2 a_t + \beta_i^2 a_i\right)\gamma.$$

In later sections, we refer to values from Lemma 3.1 as employees' 'natural' equilibrium decisions and output.

3.3.1.1 Prior to AI:

In Lemma 3.1, we can see that the natural output (o^*) increases with both ability levels (i.e., a_t and a_i). Thus, it is straightforward to see that the (H, H) type is always the top-performing employee group, while the (L, L) type has the lowest output level. Note that we assume the firm would only promote members of the top two performing employee groups (out of four). The uncertainty in the ranking of employee output lies with the (H, L) and (L, H) type - which type of employee can deliver the second-highest output? Comparing the outputs of (H, L) and (L, H), we find that the output of (H, L) is higher than (L, H) when $(a_t^H - a_t^L)\beta_t^2 \geqslant (a_i^H - a_i^L)\beta_i^2$. This is because when $(a_t^H - a_t^L)\beta_t^2 \geqslant (a_i^H - a_i^L)\beta_i^2$, the (H, L) employees' advantage in tangible skills (i.e., the left-hand side of the inequality) can fully make up for their low intangible skills. Thus, the (H, L) type will be the second-ranked employee group and will receive a bonus salary. An example of a production environment in a firm that satisfies this condition is an engineering task where the tangible skills of an employee play a more important role in achieving higher task performance. For instance, highly skilled engineers and programmers are often highly valued, even if they are not good at oral communication. In contrast, when the advantage in soft skills is more important, i.e., when $(a_t^H - a_t^L)\beta_t^2 \leqslant (a_i^H - a_i^L)\beta_i^2$, the (L, H) type would take the second place in performance ranking. An example of such a situation could be in business development, where the ability to court and convert new customers or resolve customer grievances may be highly valued even if the employee lacks strong technical skills to examine and analyze revenue data. In this

paper, we focus on the production environment where the (H, L) type employees produce a higher level of output, i.e., we assume that $(a_t^H - a_t^L)\beta_t^2 \geqslant (a_i^H - a_i^L)\beta_i^2$.

When the firm introduces the reward policy of promoting the top two groups of employees with a bonus rate, γ_B ($\gamma_B > \underline{\gamma}$), the (H, H) and (H, L) types will receive the bonus under their natural output (i.e., the output without bonus). However, the competition will intensify as the bonus rate incentivizes the lower-ranking employees to work beyond their natural labor output decisions. It makes sense for them to do so because if they advance themselves to the second place, the net utility could be larger than the utility when they settle with the third position. Under the same incentives, the employees who originally ranked in the top two positions would also devote more effort to maintain their top positions. Employees can improve their overall ranking by producing more output. All employees would produce more output at the bonus salary rate. However, no employee would produce so much output at the bonus salary rate that their net utility is lower than their net utility with a lower salary rate. To obtain the equilibrium output of different employee groups under a reward policy with a bonus, we first obtain the maximum possible output produced by an employee. To this end, we solve the following optimization problem.

$$\begin{aligned} o = \max_{w_t, w_i} \quad & \beta_t w_t + \beta_i w_i \\ s.t. \quad & \gamma_B \left(\beta_t w_t + \beta_i w_i\right) - \left(\frac{w_t^2}{a_t} + \frac{w_i^2}{a_i}\right) \geqslant \frac{1}{4}\left(\beta_t^2 a_t + \beta_i^2 a_i\right)\underline{\gamma}^2 \end{aligned}$$

The output of the above problem represents the maximum output an employee will produce to get the bonus salary rate. The constraint in the above optimization problem ensures that the net utility of the employee at the bonus salary rate (γ_B) is higher than the net utility at the base salary rate ($\underline{\gamma}$). This is because otherwise, there is no benefit of aiming for the bonus salary rate.

Solving the above optimization problem, we have the following result:

Lemma 3.2. *The highest possible output from an employee with ability level* $(a_t,\ a_i)$ *is*

$$o_{max}^* = \frac{1}{2}\left(\beta_t^2 a_t + \beta_i^2 a_i\right)\left(\gamma_B + \sqrt{\gamma_B^2 - \underline{\gamma}^2}\right).$$

From the results in Lemma 3.1 and 3.2, we have all possible production output for each employee type summarized in Table 6. We use subscripts and superscripts to denote output from a specific employee type at a specific salary rate. For example, $o_{\gamma_B}^{HH}$ is the natural output from an (H, H) employee under salary rate γ_B and o_{max}^{LH} is the maximum output from an (L, H) employee.

Table 6: Employees' output under different salary rates

Type	Natural output at rate γ	Maximum output
(H, H)	$\frac{1}{2}\left(\beta_t^2 a_t^H + \beta_i^2 a_i^H\right)\gamma$	$\frac{1}{2}\left(\beta_t^2 a_t^H + \beta_i^2 a_i^H\right)\left(\gamma_B + \sqrt{\gamma_B^2 - \underline{\gamma}^2}\right)$
(H, L)	$\frac{1}{2}\left(\beta_t^2 a_t^H + \beta_i^2 a_i^L\right)\gamma$	$\frac{1}{2}\left(\beta_t^2 a_t^H + \beta_i^2 a_i^L\right)\left(\gamma_B + \sqrt{\gamma_B^2 - \underline{\gamma}^2}\right)$
(L, H)	$\frac{1}{2}\left(\beta_t^2 a_t^L + \beta_i^2 a_i^H\right)\gamma$	$\frac{1}{2}\left(\beta_t^2 a_t^L + \beta_i^2 a_i^H\right)\left(\gamma_B + \sqrt{\gamma_B^2 - \underline{\gamma}^2}\right)$
(L, L)	$\frac{1}{2}\left(\beta_t^2 a_t^L + \beta_i^2 a_i^L\right)\gamma$	$\frac{1}{2}\left(\beta_t^2 a_t^L + \beta_i^2 a_i^L\right)\left(\gamma_B + \sqrt{\gamma_B^2 - \underline{\gamma}^2}\right)$

We now need to find out each employee's effort decisions at equilibrium, i.e., whether they produce at the natural output level or at the maximum level. There are three possible cases:

- **Case 1:** $o_{\gamma_B}^{HH}, o_{max}^{HL} > o_{max}^{LH} > o_{\gamma_B}^{HL}$. It occurs when the (L, H) type's maximum output is higher than (H, L) employees' ordinary output, imposing a challenge to (H, L)'s second place. However, the (H, L) employees can keep their second place in the output ranking and get rewarded with the bonus if they keep the output level at o_{max}^{LH} to deter the (L, H) type from overtaking attempt. The other three types, (H, H), (L, H) and (L, L), choose to maintain their natural labor decisions. This case is also the focus of our paper.
- **Case 2:** $o_{max}^{LH} > o_{\gamma_B}^{HH} > o_{\gamma_B}^{HL}$. This case follows a similar logic as that of Case 1. The (L, H) type becomes a challenge to both the (H, H) and (H, L) employees. Hence, at equilibrium both (H, H) and (H, L) keep their output level at o_{max}^{LH}.
- **Case 3:** $o_{\gamma_B}^{HH} > o_{\gamma_B}^{HL} > o_{max}^{LH}$. Such a parameter condition implies that even if the (L,H) employees worked hardest and produced their maximum output, they do not pose a challenge to the top two groups. Thus, every employee type produces their natural output.

In this paper, we focus on Case 1, where the (L, H) type's maximum output is higher than (H, L) employees' natural output under the bonus rate ($o_{max}^{LH} > o_{\gamma_B}^{HL}$) and poses a challenge to (H, L)'s second place in the output ranking. Thus, the (H, L) employees produce o_{max}^{LH} to keep their second output rank. The other three types, (H, H), (L, H) and (L, L), choose to maintain their natural labor decisions. Under this scenario, we can write the firm's profit. Let π_{noAI} and π_{AI} represent the firm's profit before and after AI. Substituting the output values from Table 6 into the firm's profit equation in 18, we get

$$\begin{aligned}\pi_{noAI} =& \frac{(1-\gamma_B)\,\gamma_B}{2}\left(\beta_t^2 a_t^H + \beta_i^2 a_i^H\right) p_t p_i + \frac{(1-\gamma_B)\left(\gamma_B + \sqrt{\gamma_B^2 - \underline{\gamma}^2}\right)}{2}\left(\beta_t^2 a_t^L + \beta_i^2 a_i^H\right) p_t(1-p_i) \\ &+ \frac{\left(1-\underline{\gamma}\right)\underline{\gamma}}{2}\left[\left(\beta_t^2 a_t^L + \beta_i^2 a_i^H\right)(1-p_t)p_i + \left(\beta_t^2 a_t^L + \beta_i^2 a_i^L\right)(1-p_t)(1-p_i)\right].\end{aligned} \tag{20}$$

We now proceed to the analysis in the presence of AI.

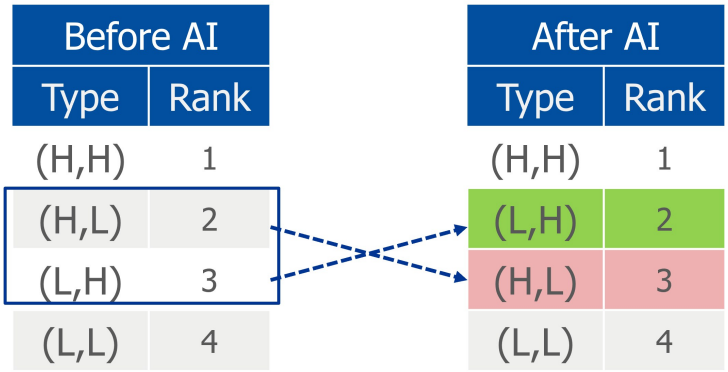
3.3.1.2 Post AI:

After introducing AI assistant into the daily operation, employees with low tangible skills improve their ability to $a_{te}^L = a_t^L + f(a_t^H - a_t^L)$ as given in equation (17). With the enhanced tangible ability, the ranking of employees' output can differ from the prior-AI case. We focus on the most interesting scenario where (L, H) can outperform (H, L) with the help of AI as portrayed by Figure 19. That is, the natural output from (L, H) is higher than the natural output from (H, L) in an AI-assisted environment. Thus, we assume that $(a_t^H - a_t^L)(1-f)\beta_t^2 < (a_i^H - a_i^L)\beta_i^2$.

The competition for the second place follows a similar logic that (L, H) employees need to devote extra effort to ensure the (H, L) type employees stay in the third place. Hence, we have the post-AI output ranking in Table 7.

Substituting the output values from Table 7 into the firm's profit equation in 18, we get

$$\begin{aligned}\pi_{AI} =& \frac{(1-\gamma_B)\,\gamma_B}{2}\left(\beta_t^2 a_t^H + \beta_i^2 a_i^H\right) p_t p_i + \frac{(1-\gamma_B)\left(\gamma_B + \sqrt{\gamma_B^2 - \underline{\gamma}^2}\right)}{2}\left(\beta_t^2 a_t^H + \beta_i^2 a_i^L\right)(1-p_t)p_i \\ &+ \frac{\left(1-\underline{\gamma}\right)\underline{\gamma}}{2}\left[\left(\beta_t^2 a_t^H + \beta_i^2 a_i^L\right) p_t(1-p_i) + \left(\beta_t^2 a_{te}^L + \beta_i^2 a_i^L\right)(1-p_t)(1-p_i)\right].\end{aligned} \tag{21}$$



AI could help (L, H) type employees overtake (H, L) type employees.

Figure 19: Employee ranking Before AI vs. After AI

Table 7: Employees' optimal decisions - prior to & after AI adoption

Type	Before AI		After AI	
	Ranking	Output	Ranking	Output
(H, H)	1	$\frac{1}{2}\left(\beta_t^2 a_t^H + \beta_i^2 a_i^H\right)\gamma_B$	1	$\frac{1}{2}\left(\beta_t^2 a_t^H + \beta_i^2 a_i^H\right)\gamma_B$
(H, L)	2	$\frac{1}{2}\left(\beta_t^2 a_t^L + \beta_i^2 a_i^H\right)\left(\gamma_B + \sqrt{\gamma_B^2 - \underline{\gamma}^2}\right)$	3	$\frac{1}{2}\left(\beta_t^2 a_t^H + \beta_i^2 a_i^L\right)\underline{\gamma}$
(L, H)	3	$\frac{1}{2}\left(\beta_t^2 a_t^L + \beta_i^2 a_i^H\right)\underline{\gamma}$	2	$\frac{1}{2}\left(\beta_t^2 a_t^H + \beta_i^2 a_i^L\right)\left(\gamma_B + \sqrt{\gamma_B^2 - \underline{\gamma}^2}\right)$
(L, L)	4	$\frac{1}{2}\left(\beta_t^2 a_t^L + \beta_i^2 a_i^L\right)\underline{\gamma}$	4	$\frac{1}{2}\left(\beta_t^2 a_{te}^L + \beta_i^2 a_i^L\right)\underline{\gamma}$

We now proceed to compare the results before and after AI to assess the impact of AI.

3.3.2 Impact on Firm's Output and Profit

To obtain the impact of AI on the firm's output and profit, we first analyze AI's impact on output and profit from each employee type. Then, we will aggregate this to understand the impact on the firm.

Impact on Output of Employee Groups: Using the notations for equilibrium output

levels specified in Table 6, we have the total output as:

$$O_{noAI} = o^{HH}_{\gamma_B} p_{HH} + o^{LH}_{max} p_{HL} + o^{LH}_{\underline{\gamma}} p_{LH} + o^{LL}_{\underline{\gamma}} p_{LL}.$$
$$O_{AI} = o^{HH}_{\gamma_B} p_{HH} + o^{HL}_{max} p_{LH} + o^{HL}_{\underline{\gamma}} p_{HL} + o^{LL_{AI}}_{\underline{\gamma}} p_{LL}.$$

Let Δo^{XY}, defined as the post-AI output subtract the prior-to-AI value, represent the productivity change for an individual employee of type (X, Y), with $X,\ Y \in \{L, H\}$. By comparing the production output of each employee type in Table 7, we get the following:

$$\begin{aligned}
\Delta o^{HH} =& 0, \\
\Delta o^{HL} =& \underbrace{o^{HL}_{\underline{\gamma}} - o^{LH}_{max}}_{Competition\ Effect}, \\
\Delta o^{LH} =& o^{HL}_{max} - o^{LH}_{\underline{\gamma}}, \\
=& \underbrace{\left(o^{HL}_{max} - o^{LH_{AI}}_{\underline{\gamma}}\right)}_{Competition\ Effect} + \underbrace{\left(o^{LH_{AI}}_{\underline{\gamma}} - o^{LH}_{\underline{\gamma}}\right)}_{Learning\ Effect}, \\
\Delta o^{LL} =& \underbrace{o^{LL_{AI}}_{\underline{\gamma}} - o^{LL}_{\underline{\gamma}}}_{Learning\ Effect}.
\end{aligned}$$

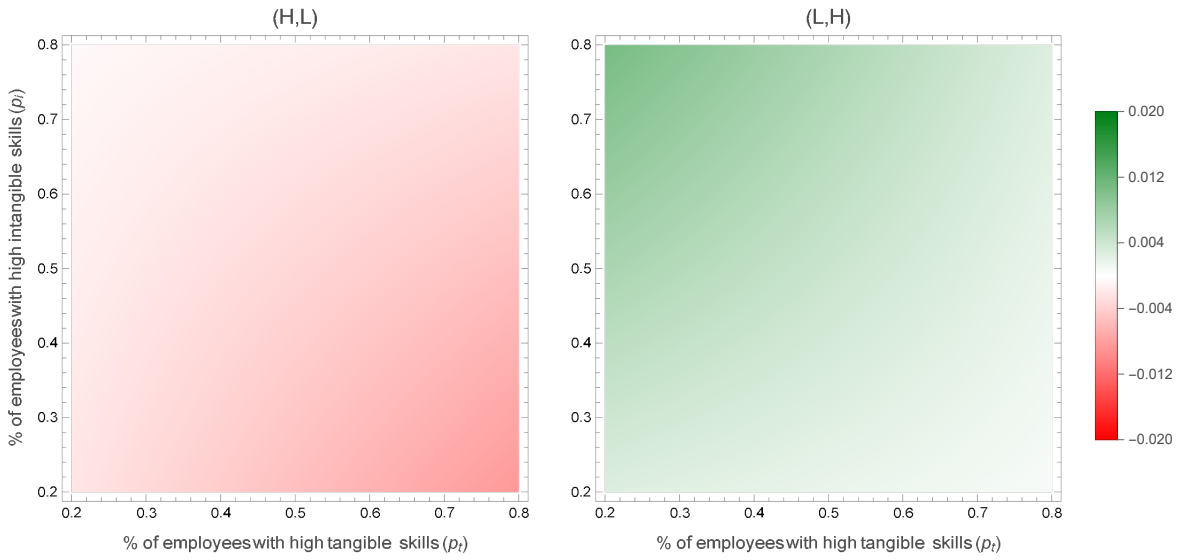
From the above breakdown of output, we observe the following: (i) productivity of the (H, H) type is not affected by AI adoption, (ii) AI has direct and indirect impact on the productivity of employees. AI's direct impact on productivity is through boosting the tangible abilities of the (L, H) and (L, L) types (we refer to this as the "learning effect" of AI). (iii) AI also has an indirect impact on productivity as a result of the change in employees' performance ranking (we refer to this as the "competition effect" of AI): This competition effect is negative on (H, L) type and they produce less because they are pushed down to the third place and don't receive bonus; on the other hand, this impact is positive on the (L, H) type employees. Thus, on top of their increased contribution due to ability improvement (via "learning effect"), they are further incentivized by the bonus rate after AI adoption due to the improvement in their output ranking. Formally, we note the AI's impact on the output of different employee types as follows:

Proposition 3.1. *The output from the* (H, L) *type decreases, and the output of* (L, H) *and* (L, L) *employees increases. The output of* (H, H) *remains the same. That is,* $\Delta o^{HL} < 0, \Delta o^{LH} > 0$ *&*$\Delta o^{HH} = 0$.

Impact on Profit from Employees Groups: We now analyze AI's impact on the firm's profit. We define $\Delta\pi^{XY} = \pi_{AI}^{XY} - \pi_{noAI}^{XY}$ as the change in profit contributed by an individual employee of type (X, Y), with $X,\ Y \in \{L, H\}$. We compare the expression of π_{noAI} and π_{AI} from equations (20) and (21) and break down the total profit change by employee types and get the following:

$$
\begin{aligned}
\Delta\pi^{HH} =& 0, \\
\Delta\pi^{HL} =& \underbrace{(1-\underline{\gamma})o_{\underline{\gamma}}^{HL} - (1-\gamma_B)o_{max}^{LH}}_{Competition\ Effect}, \\
\Delta\pi^{LH} =& (1-\gamma_B)o_{max}^{HL} - (1-\underline{\gamma})o_{\underline{\gamma}}^{LH} \\
=& \underbrace{\left((1-\gamma_B)o_{max}^{HL} - (1-\underline{\gamma})o_{\underline{\gamma}}^{LH_{AI}}\right)}_{Competition\ Effect} + \underbrace{(1-\underline{\gamma})\left(o_{\underline{\gamma}}^{LH_{AI}} - o_{\underline{\gamma}}^{LH}\right)}_{Learning\ Effect}, \\
\Delta\pi^{LL} =& \underbrace{(1-\underline{\gamma})\left(o_{\underline{\gamma}}^{LL_{AI}} - o_{\underline{\gamma}}^{LL}\right)}_{Learning\ Effect}.
\end{aligned}
\tag{22}
$$

It is easy to see that the profit from (H, H) has no impact from AI, because the output and salary of this employee group remain unchanged after AI adoption. Similarly, it is also easy to see that the profit from (L, L) type employees increases, because their output increases but the salary remains the same. Thus, the firm gets more output from these employees at the same salary. However, the direction of impact on the output from (H, L) and (L, H) is not clear because the output of (H, L) decreases, but the firm has to pay a lower salary rate. Similarly, the output of (L, H), increases, but the firm also has to pay a bonus salary. In Figure 20, we illustrate how AI affects the output of (H, L) and (L, H) employees differently. The color represents the change in profit contribution after AI (i.e., $\pi_{AI} - \pi_{noAI}$). The green color means that the profit attributed to an employee type increased after AI, and the red color represents a decrease in the profit after AI. We can see that the profit from (H, L) type employees decreases after AI and the profit from (L, H) type employees increases after AI.



The color represents the change in profit after AI (i.e., π_{AI} minus π_{noAI}). The green color means that the profit increased after AI from that employee type. The profit from (H, L) type employees decreases after AI, and the profit from (L, H) type employees increases after AI.

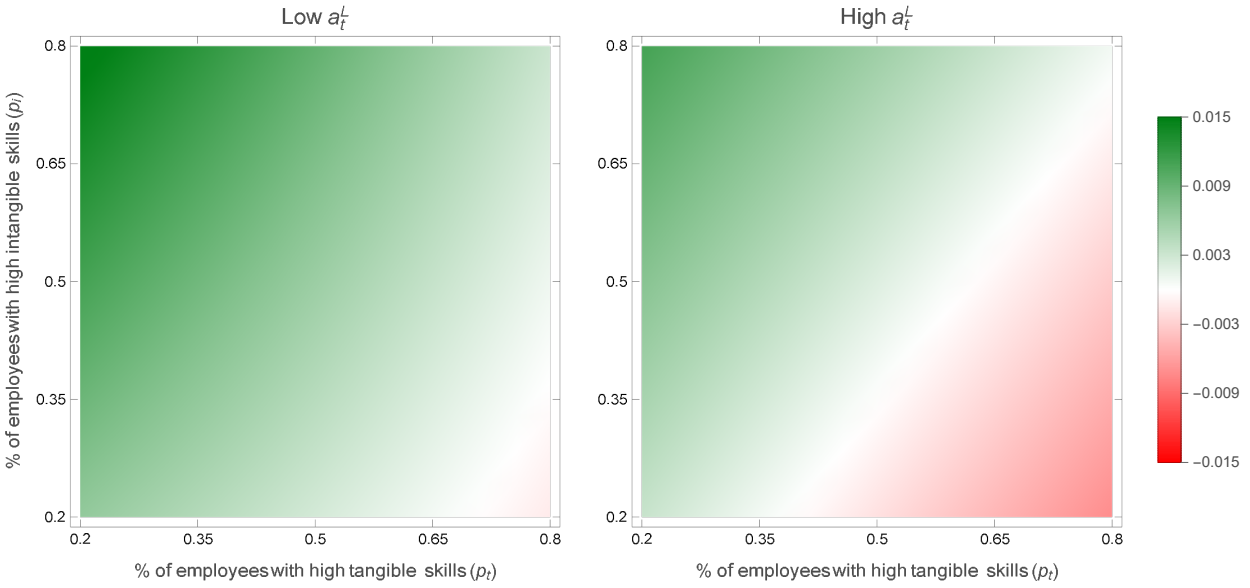
Figure 20: Change in the profit contribution (after AI minus before AI) by employee types

Combining the analyses of individual employee's output and profit contribution, we now compare the total output and profit for the firm and reach the following conclusion. Define $K = \gamma_B + \sqrt{\gamma_B^2 - \underline{\gamma}^2}$, $\hat{a}^{(o)} = 2\frac{(o_{max}^{HH} - o_{\underline{\gamma}}^{HL})p_{HL} - (o_{max}^{HL} - o_{\underline{\gamma}}^{HH})p_{LH}}{\beta_t^2(Kp_{HL} + \underline{\gamma}p_{LH} + f\underline{\gamma}p_{LL})}$, $\hat{p}_{HL}^{(o)} = \frac{\Delta o^{LH}p_{LH} + \Delta o^{LL}p_{LL}}{\Delta o^{HL}}$, $\hat{a}^{(\pi)} = 2\frac{(M_2 o_{max}^{HH} - M_1 o_{\underline{\gamma}}^{HL})p_{HL} - (M_2 o_{max}^{HL} - M_1 o_{\underline{\gamma}}^{HH})p_{LH}}{\beta_t^2(M_2 Kp_{HL} + M_1\underline{\gamma}p_{LH} + M_1 f\underline{\gamma}p_{LL})}$, $\hat{p}_{HL}^{(\pi)} = \frac{\Delta\pi^{LH}p_{LH} + \Delta\pi^{LL}p_{LL}}{\Delta\pi^{HL}}$, $\hat{p}_{HL} = \max\{\hat{p}_{HL}^{(o)}, \hat{p}_{HL}^{(\pi)}\}$, and $\hat{a} = \min\{\hat{a}^{(o)}, \hat{a}^{(\pi)}\}$.

Theorem 3.1. *The firm's output and profit decrease after adopting AI when the tangible ability gap between employees is low and there are a large number of (H, L) type employees. That is, $O_{AI} < O_{noAI}$ and $\pi_{AI} < \pi_{noAI}$, when $\Delta a_t = a_t^H - a_t^L \leqslant \hat{a}$ and $p_{HL} \geqslant \hat{p}_{HL}$.*

From Theorem 3.1, we can see that the parameter conditions for the firm's productivity and profit follow similar patterns. We focus on providing the intuition on the profit drop–the second part of Theorem 3.1. First, the parameter condition $\Delta a_t = a_t^H - a_t^L \leqslant \hat{a}$ indicates that if the low-type employee's ability level of tangible skills is not very different from that of the high-type, then the adoption of AI can lead to a decrease in profit for the firm. Intuitively, when the tangible ability of low-type employees is very different from the high-type (i.e., a_t^L

is small), there is considerable potential for employees to improve with AI. Therefore, the Learning Effect of AI, which is always positive, can be more significant. On the other hand, if the tangible ability gap between high-type and low-type employees is not much, then there isn't much for AI to improve. The second factor is employee composition. If the composition of labor in a firm has a high percentage of (H, L) employees, i.e., $p_{HL} \geqslant \hat{p}_{HL}$, then the firm is more likely to experience profit decrease after adopting AI. Intuitively, when the (H, L) type accounts for a large proportion of the workforce, the reduction in the firm's profit is mainly due to the demotivating effect experienced by the (H, L) type employees. With AI hurting the motivation of the main body of employees, we expect the firm's profit to decrease.



This plot visually shows Theorem 3.1–how the firm's profit after AI adoption changes, with respect to the ability of low-type (a_t^L) and employee composition (p_t & p_i).

Figure 21: Firm's profit change (after AI minus before AI)

Figure 21 illustrates the results in Theorem 3.1 and shows how AI affects a firm's profit in various task and organizational environments, with green areas indicating an increase in profit after AI implementation and red gradient areas representing a profit drop. The comparison between the plots reveals that when the ability gap between employees is small, i.e., a large a_t^L, profit decrease after AI adoption is more likely to occur. From the individual

plot, we observe that AI can backfire when the workforce falls into scenarios in the lower right corner, i.e., a large proportion of employees are the (H, L) type.

We now proceed to analyze the impact of AI adoption on the welfare of employees.

3.3.3 Impact on Employee Welfare

In the previous section, we saw that AI can negatively impact a firm's profit. We now analyze the impact of AI on the welfare of employees. We measure the employees' welfare by their net utility in equilibrium. Using employees' decisions as given in Table 7, we find the following result:

Proposition 3.2. *After the adoption of AI:*

- *The* (H, H) *employees are unaffected.*
- *The* (H, L) *employees are worse-off.*
- *The* (L, H) *employees are better-off.*
- *The* (L, L) *employees are better-off.*

The conclusion on the welfare of individual employee types is consistent with the breakdown analyses for productivity in Section 3.3.2. The (H, H) type employees remain unaffected by AI because AI doesn't improve their ability and they continue to maintain the top position in the output ranking. The (H, L) type employees suffer a welfare loss because of the lower output ranking after AI. As AI improves the abilities of (L, H) and (L, L) employees, they become more productive and competitive and achieve a higher net utility after AI.

From Proposition 3.2, we also observe that employees with high-tangible skills $\big($i.e., (H, H) and $(H, L)\big)$ will be worse off, and the low-tangible skills $\big($i.e., (L, H) and $(L, L)\big)$ are better off after AI. Similarly, employees with high-intangible skills $\big($i.e., (H, H) and $(L, H)\big)$ will be better off. Interestingly, AI benefits employees who are better at intangible skills. Intuitively, this is because as AI evens out the differences in tangible skills, those who are strong in intangible skills emerge as the more competitive employees and become the top performers.

In the paper thus far, we showed that adopting AI could lead to a decrease in the firm's profit. We now analyze some remedies that can be used to avoid a profit loss for the firm

due to the demotivating effect of AI adoption.

3.4 Remedies

We know that the main reason behind the profit loss of a firm is the decrease in output from (H, L) type employees. This happens because the output ranking of (H, L) employees goes down, and they are paid a lower salary rate. We analyze the following two policies to mitigate the problem:

- **Guaranteed Salary:** In this policy, the firm guarantees that no employee's salary rate goes down after the adoption of AI. Thus, if an employee type was getting a bonus salary rate before the AI adoption, they would continue to get that salary rate even after the AI adoption.
- **Choosing Optimal AI Level:** In our model, the reason (H, L) type employees' output ranking goes down is that AI transfers their tangible knowledge to their competing employees. This knowledge transfer makes the competitors of (H, L) more productive, and they overtake (H, L) in the output ranking. The firm can mitigate this problem by deliberately choosing a less effective AI. We model this by allowing the firm to choose the optimal value of AI efficacy f.

  We now analyze these policies in detail below.

3.4.1 Guaranteed Salary

In this policy, the firm guarantees that no employee's salary rate will decrease after AI adoption. We name this reward scheme the 'guaranteed-salary' policy and use π^{g}_{AI} to denote the firm's post-AI profit under this reward policy (here, the superscript 'g' stands for 'guarantee'). We update the notations and use π^{ng}_{noAI} and π^{ng}_{AI} for the profits under the base model reward scheme to emphasize that it doesn't provide any salary guarantee ('ng' for 'no-guarantee').

Since the guaranteed-salary policy 'guarantees' that the no employee's salary rate will go down after AI adoption, it is rational for the (H, L) employees to expect to be rewarded at the bonus rate because this is what they were getting before AI adoption. This guarantee of bonus rate has an anti-competitive element because now the (H, L) employees do not have to produce extra output to maintain their second position and get the bonus (the bonus is guaranteed now). Hence, the (H, L) type simply produces the 'natural' output according to Lemma 3.1.

Similarly, (L, H) type employees also produce their 'natural' output because (H, L) employees are no longer giving them competition and the natural output is enough to place them in the second position in the output ranking to get the bonus rate. Overall, all employee types produce their 'natural' output as given the Lemma 3.1, and nobody produces any extra output (or competitive output). Therefore, we can write the profit of the firm under the guaranteed-salary policy as:

$$\pi^{g}_{AI} = (1-\gamma_B)(o^{HH}_{\gamma_B} p_{HH} + o^{LH_{AI}}_{\gamma_B} p_{LH} + o^{HL}_{\gamma_B} p_{HL}) + (1-\underline{\gamma}) o^{LL_{AI}}_{\underline{\gamma}} p_{LL}.$$

To find out whether the guaranteed salary could rectify AI's backfiring phenomenon, we compare the firm's post-AI profit under the guaranteed and no-guaranteed policies (π^{g}_{AI} vs π^{ng}_{AI}). We find that the profit under the guaranteed policy can only perform better than the no-guaranteed policy with stringent conditions. Denote $\hat{a}^{(g)} = 2\frac{M_2(o^{HH}_{\gamma_B} - o^{HL}_{max})p_{LH} - (M_1 o^{HL}_{\underline{\gamma}} - M_2 o^{HL}_{\gamma_B})p_{HL}}{\beta_t^2 M_2 \gamma_B (1-f) p_{LH}}$ and $\hat{p}^{(g)}_{HL} = \frac{M_2(o^{HL}_{max} - o^{LH_{AI}}_{\gamma_B})p_{LH}}{M_2 o^{HL}_{\gamma_B} - M_1 o^{HL}_{\underline{\gamma}}}$, we formally note the result as follows:

Proposition 3.3. *The firm's profit under the guaranteed policy could be higher than that under the no-guaranteed policy, i.e.,* $\pi^{g}_{AI} \geqslant \pi^{ng}_{AI}$ *when* $\Delta a_t = a^{H}_t - a^{L}_t \leqslant \hat{a}^{(g)}$, $p_{HL} \geqslant \hat{p}^{(g)}_{HL}$ *and* $(1-\gamma_B)\gamma_B > (1-\underline{\gamma})\underline{\gamma}$.

Figure 22 visually depicts parameter conditions in Proposition 3.3 and compares that with the main results. The first plot shows the cases where the guaranteed-salary policy outperforms the no-guaranteed policy. At the same time, the plot in the middle portrays the scenarios in Theorem 3.1, where AI is backfiring. Comparing these two subplots, we observe that when AI's demotivating effect dominates (i.e., $\pi^{ng}_{AI} < \pi_{noAI}$), the guaranteed policy can only be helpful for certain extreme cases (i.e., the percentage of the (H, L) type is very high

and the tangible skill gap is minimal). Additionally, even when the guaranteed-salary policy could mitigate AI's negative impact, it still leads to a lower profit level than the prior-to-AI case ($\pi^g_{AI} < \pi_{noAI}$). This can be seen by comparing the first and the third plots that in the parameter region where π^g_{AI} outperforms π^{ng}_{AI}, the solid shading, π^g_{AI} is still lower than the pre-AI profit, as shown in the third plot.

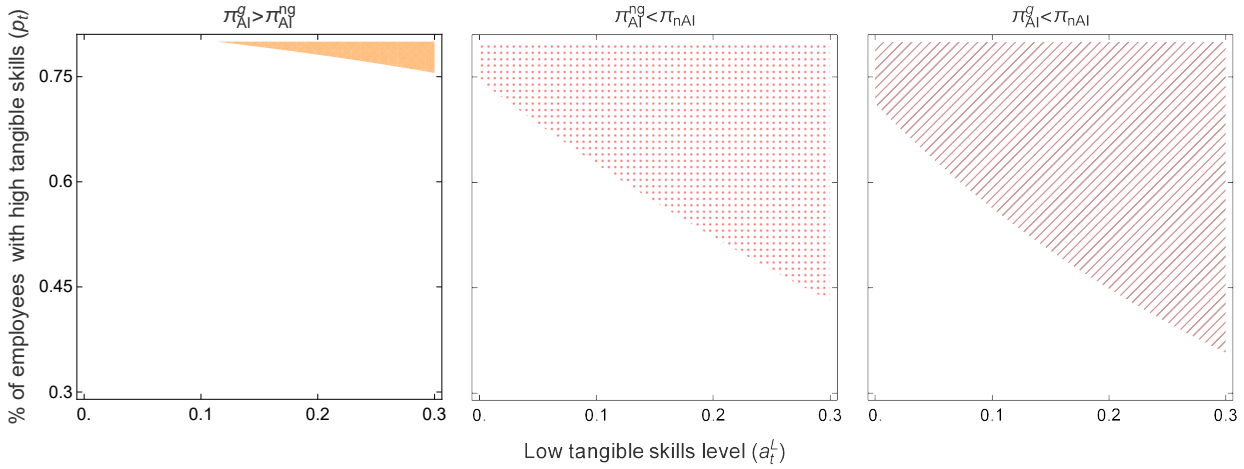


Figure 22: Profit comparison among three cases: before AI, after AI, and after AI with guaranteed salary

To investigate why the guaranteed salary cannot fully mitigate AI's demotivating effect, we break down the difference between the profits under two post-AI reward policies:

$$\pi^g_{AI} - \pi^{ng}_{AI} = M_2(o^{LH_{AI}}_{\gamma_B} - o^{HL}_{max})p_{LH} + (M_2 o^{HL}_{\gamma_B} - M_1 o^{HL}_{\underline{\gamma}})p_{HL}.$$

The main differences lie with the profit attributed to (H, L) and (L, H) employees. First, because the salary guarantee eliminates competition between employees, the (L, H) employees are no longer motivated to produce beyond their natural output. Hence, the (L, H) employees' post-AI profit contribution shrinks (i.e., $o^{LH_{AI}}_{\gamma_B} - o^{HL}_{max}$ is negative). Second, compared to the no-guaranteed policy, the (H, L) type increases productivity because they expect a higher salary rate. However, the firm will have to pay higher wages. It is possible that the total profit generated by (H, L) employees becomes even lower (i.e., when $(1-\gamma_B)\gamma_B < (1-\underline{\gamma})\underline{\gamma}$, the profit generated by (H, L) employees $M_2 o^{HL}_{\gamma_B} - M_1 o^{HL}_{\underline{\gamma}}$ is also negative). Therefore, only when

the skill gap, the size of (H, L) employees, and the salary rates satisfy the restricted conditions can the guaranteed policy mitigate AI's demotivating effect. Otherwise, the salary guarantee would hurt the firm, and eventually, the firm's profit would be lower than the no-guarantee policy.

We now proceed to analyze the policy where the firm chooses the optimal level of AI.

3.4.2 Choosing Optimal AI Level

Thus far, we have focused on the case when AI changes the productivity ranking between (H, L) and (L, H) employees. Therefore, the (H, L) employees reduce their output, resulting in a total productivity and profit drop. From our analysis of Section 3.3.1, we know that the flipping in performance ranking does not always occur. Specifically, the choice of AI (f) greatly affects the competition dynamic among employees. We first consider the case when AI is not very effective, that is when AI efficacy f is lower than $1 - \frac{\beta_i^2\left(a_i^H - a_i^L\right)}{\beta_t^2\left(a_t^H - a_t^L\right)}$, the (H, L) employees can keep their lead in performance ranking over the (L, H) type. Hence, employees' output decisions will be similar to the before-AI scenario in Table 7, with the (L, H) and (L, L) producing at their new skill levels. Another scenario we consider is when AI becomes very effective, and it could boost the (L, H) type to be so productive that the other employees cannot compete with them for the bonus. Under this scenario, all employees will produce at the natural output level.

Denote $\hat{f}^{(l)} = 1 - \frac{\beta_i^2\left(a_i^H - a_i^L\right)}{\beta_t^2\left(a_t^H - a_t^L\right)}$ and $\hat{f}^{(u)} = 1 - \frac{\beta_i^2\left(a_i^H - a_i^L\right)}{\beta_t^2\left(a_t^H - a_t^L\right)} + \frac{\sqrt{\gamma_B^2 - \underline{\gamma}^2}(\beta_t^2 a_t^H + \beta_i^2 a_i^L)}{\gamma_B \beta_t^2\left(a_t^H - a_t^L\right)}$, we combine three cases of employee competition and derive the firm's profit as a function of f.

$$\pi_{AI}(f) = \begin{cases} (1-\gamma_B)(o_{\gamma_B}^{HH} p_{HH} + o_{max}^{LH_{AI}} p_{HL}) + (1-\underline{\gamma})(o_{\underline{\gamma}}^{LH_{AI}} p_{LH} + o_{\underline{\gamma}}^{LL_{AI}} p_{LL}), & f \leqslant \hat{f}^{(l)} \\ (1-\gamma_B)(o_{\gamma_B}^{HH} p_{HH} + o_{max}^{HL} p_{LH}) + (1-\underline{\gamma})(o_{\underline{\gamma}}^{HL} p_{HL} + o_{\underline{\gamma}}^{LL_{AI}} p_{LL}), & \hat{f}^{(l)} < f < \hat{f}^{(u)} \\ (1-\gamma_B)(o_{\gamma_B}^{HH} p_{HH} + o_{\gamma_B}^{LH_{AI}} p_{LH}) + (1-\underline{\gamma})(o_{\underline{\gamma}}^{HL} p_{HL} + o_{\underline{\gamma}}^{LL_{AI}} p_{LL}), & f \geqslant \hat{f}^{(u)} \end{cases} \quad (23)$$

Figure 23 plots the above function $\pi_{AI}(f)$, which has two kinks at thresholds $\hat{f}^{(l)}$ and $\hat{f}^{(u)}$, and linearly increases otherwise. We compare the firm's profit under three different cases. Denote $\hat{p}_{HL}^{(f)} = \frac{(M_2 o_{\gamma_B}^{HH} - M_1 o_{\underline{\gamma}}^{HL}) p_{LH} + M_1 \frac{\gamma}{2} \beta_i^2 (a_i^H - a_i^L) p_{LL}}{M_2 o_{max}^{HL} - M_1 o_{\underline{\gamma}}^{HL}}$, we reach the following conclusion on the maximum level of the post-AI profit:

Proposition 3.4. *The firm can achieve the highest profit by choosing an AI that is just about to avoid the flipping in employee performance. That is, the maximum profit is* $\pi_{AI}(\hat{f}^{(l)})$ *when* $p_{HL} \geqslant \hat{p}_{HL}^{(f)}$.

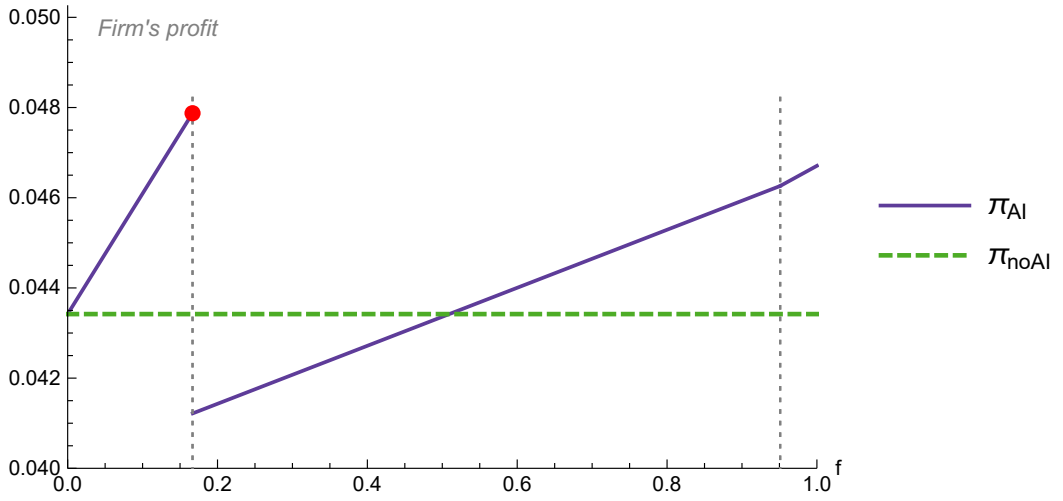


Figure 23: How the firm's profit changes with the choice of AI

In Figure 23, we illustrate the profit trend as the firm chooses different levels of AI f. The vertical dotted line on the left capture the flipping threshold $\hat{f}^{(l)}$ above which the AI-induced knowledge transfer leads to a different competition paradigm among employee types. We can see that if the firm picks an AI application that is less effective than $\hat{f}^{(l)}$, it can reach the profitable AI deployment because no flipping in productivity happens between (H, L) and (L, H) employee types. In this scenario, there is no demotivating effect on the (H, L) employee types; therefore, AI is beneficial as the post-AI profit (i.e., purple solid line) is always above the amount without AI (i.e., green dashed line).

When the firm picks an AI with more capability than $\hat{f}^{(l)}$, the profit could suddenly drop to a level that AI implementation becomes an unwise investment. Even though the profit level recovers as the firm chooses an even more effective AI product, the profit generated by a perfect AI could still be lower than the amount without flipping. This is illustrated with Figure 23: in each case, the profit increases as f increases; and the firm's profit is at the highest level when AI is just about to trigger performance flipping (i.e., the red dot).

Up to this point in the paper, we demonstrate that implementing AI may lead to a reduction in the company's earnings. We find that the firm can circumvent the profit decrease by optimizing the choice of AI. Next, we introduce two model extensions as robustness checks.

3.5 Generalized Model of AI

In real-world business use cases, AI relies on data and even human feedback to improve its capability to help employees. In this section, we extend the model on how AI effectiveness f is realized by learning from employee activity data. We model that two factors would decide the overall AI efficacy: the first factor is the base effectiveness (f_0) when the firm purchases an off-the-shelf solution or builds an AI assistant; the second part comes from what AI observes and learns from employees' data. Mathematically, we formulate AI effectiveness as

$$f = f_0 + (1 - f_0)\theta \frac{W_t^H}{W_t^{max}}. \tag{24}$$

In the second component of AI efficacy, we use θ, which ranges between 0 to 1, to represent the quality of the monitoring infrastructure to collect employee data (i.e., $\theta = 1$ means that the firm has data collection infrastructure in place and AI has the potential to become perfect by feeding enough data.) W_t^H is the total tangible labor from high-type employees. The normalization factor W_t^{max} is the maximum possible total tangible labor produced by the high type. Without loss of generality, we set $W_t^{max} = \frac{\bar{\gamma}}{2}\beta_t a_t^H(p_{HH} + p_{HL})$ with $\bar{\gamma}$ as the maximum possible salary rate. W_t^H comes from the labor decisions of employees with high-type tangible skills. The realized AI efficacy will, in turn, affect employees' ability and their competition. Therefore, the (H, L) type could strategically choose the tangible labor input to minimize AI's impact on improving the (L, H) type. The choice of f_0 and θ will also decide if the competition dynamic falls into one of the following three scenarios:

- (H, L) type's labor decision when they rank the 2nd is not enough for flipping;
- (H, L) type's natural labor decision at the base rate is enough to trigger the flipping;
- (H, L) employees reduce the tangible labor to maintain their 2nd place

Next, we analyze how employees make decisions with f_0 and θ as given parameters under each scenario and derive firms' profit accordingly.

Case 1: the (H, L) type maintains its lead. The high-type tangible labor from the (H, L) employees is affected by the (L, H) type's devotion to challenging the top performers. Under this scenario, the (L, H) type at lower rank has a strong motivation as they can benefit from the increased high-type tangible labor. Hence, they are a creditable threat to those in the 2nd place, and the (H, L) employees would have to produce at the maximum level as shown in Table 7. The overall effectiveness becomes:

$$\begin{aligned} f =& f_0 + (1 - f_0)\theta \frac{\frac{\gamma_B}{2}\beta_t a_t^H p_{HH} + \frac{K}{2}\beta_t a_t^H \frac{A_5}{A_2} p_{HL}}{\frac{\bar{\gamma}}{2}\beta_t a_t^H (p_{HH} + p_{HL})} \\ =& f_0 + (1 - f_0)\frac{\theta}{\bar{\gamma}}\left[\gamma_B p_i + K\frac{\beta_t^2 a_{te}^L + \beta_i^2 a_i^H}{\beta_t^2 a_t^H + \beta_i^2 a_i^L}(1 - p_i)\right] \end{aligned}$$

Solving for the rational expectation equilibrium, we obtain the final AI efficacy as a function of other parameters, which we denote with $F_1(f_0, \theta)$:

$$F_1(f_0, \theta) = \frac{\bar{\gamma}A_2 f_0 + (1 - f_0)[\gamma_B A_2 p_i + K A_3(1 - p_i)]\theta}{\bar{\gamma}A_2 - (1 - f_0)K\beta_t^2(a_t^H - a_t^L)(1 - p_i)\theta}.$$

Last, we identify the conditions under which case 1 will hold. That is, when $F_1(f_0, \theta) \leqslant \hat{f}$, the high-type tangible labor when (H, L) is at the 2nd place is still not enough to make (L, H) more productive than (H, L); hence, flipping would never occur. The performance ranking and employee incentives remain the same as in the prior-AI period. As a result, the firm's profit follows the output decisions of the 'Before AI' scenario in Table 7, with AI improving the low-type employees' tangible ability by rate $F_1(f_0, \theta)$.

Case 2: the flipping occurs and the (H, L) type becomes the 3rd in performance ranking. Under this scenario, (H, L) employees would be willing to settle at the 3rd place because they know even their tangible labor at the natural level with the expected salary at $\underline{\gamma}$ would trigger the flipping. Therefore, we have the tangible labor from the (H, L) to be $\frac{\gamma}{2}\beta_t a_t^H$. The (H, H) employees would also produce at the natural level with the expected salary γ_B (their tangible labor is $\frac{\gamma_B}{2}\beta_t a_t^H$.) The overall AI effectiveness at equilibrium can be derived as a function of the firm's choice over f_0 and θ, which we denote with $F_2(f_0, \theta)$:

$$
\begin{aligned}
F_2(f_0,\theta) =& f_0 + (1-f_0)\theta \frac{\frac{\gamma_B}{2}\beta_t a_t^H p_{HH} + \frac{\underline{\gamma}}{2}\beta_t a_t^H p_{HL}}{\frac{\bar{\gamma}}{2}\beta_t a_t^H (p_{HH}+p_{HL})} \\
=& f_0 + (1-f_0)\theta \frac{\gamma_B p_i + \underline{\gamma}(1-p_i)}{\bar{\gamma}}.
\end{aligned}
$$

To obtain the firm's profit, we need to determine employee decisions. Intriguingly, employee awareness about how AI works would affect the competitive dynamic for the 2nd place. When the (H,L) knows that their labor benefits the (L,H) type, they have little incentive to challenge the (L,H) employees because the more they produce, the more capable (L,H) becomes. If the (H,L) type produces at its maximum output level, it will make the (L,H) type even more productive, and even their maximum output level would not be a threat to the (L,H) type. As a result, it is not in the (H,L) employees' interests to challenge those promoted. The (L,H) type only needs to produce at the natural level with the expectation of receiving the bonus rate.

Next, we identify the conditions that if the firm's decisions fall within the range, the overall AI efficacy with the (H,L) employees in the 3rd place is equal to or higher than the threshold value for flipping ($\hat{f}$). From Section 3.3.1, we know that $\hat{f} = 1 - \frac{\beta_i^2(a_i^H - a_i^L)}{\beta_t^2(a_t^H - a_t^L)}$. Therefore, when $F_2(f_0,\theta) \geqslant \hat{f}$, the high-type tangible labor when (H,L) is at rank 3 is enough to promote (L,H) and we have an equilibrium performance ranking of $(L,H) > (H,L)$.

Case 3: we find that the conditions for Case 1 and Case 2 are different, indicating that there is a gap region where the overall AI efficacy won't be enough to trigger the flipping with the (H,L) type's labor decision at the 3rd place. Specifically, we find that when $F_1(f_0,\theta) < F_2(f_0,\theta)$, the (H,L) employees would strategically reduce their effort so that they maintain their performance lead. (H,L) employees reduce their tangible labor to a level where both (H,L) and (L,H) have the same output level and the firm will only promote the (H,L) type. With the equal output condition at the equilibrium, we can solve for individual (H,L) employee's labor decision and obtain the function for the final AI efficacy:

$$
F_3(f_0,\theta) = \frac{\bar{\gamma}\beta_t^2 f_0 a_t^H + (1-f_0)[\underline{\gamma}A_3(1-p_i) + \gamma_B(A_2 p_i - \beta_i^2 a_i^L)]\theta}{\beta_t^2[\bar{\gamma}a_t^H - (1-f_0)(a_t^H - a_t^L)(1-p_i)\underline{\gamma}\theta]}.
$$

Next, we put together three equilibria under different values of f_0 and θ and focus on how the choice of base AI effectiveness f_0 affects the final AI efficacy and the overall profit.

With two boundary values of f_0 as $\hat{f}_0^{(l)} = \frac{\hat{f}\bar{\gamma}-[\gamma_B p_i+K(1-p_i)]\theta}{\bar{\gamma}-[\gamma_B p_i+K(1-p_i)]\theta}$ and $\hat{f}_0^{(u)} = \frac{\hat{f}\bar{\gamma}-\left[\gamma_B p_i+\underline{\gamma}(1-p_i)\right]\theta}{\bar{\gamma}-\left[\gamma_B p_i+\underline{\gamma}(1-p_i)\right]\theta}$ we have the firm profit as:

$$\pi_{AI} = \begin{cases} M_2(o_{\gamma_B}^{HH} p_{HH} + o_{max}^{LH_{AI}} p_{HL}) + M_1(o_{\underline{\gamma}}^{LH_{AI}} p_{LH} + o_{\underline{\gamma}}^{LL_{AI}} p_{LL}), & f_0 \leqslant \hat{f}_0^{(l)} \\ M_2(o_{\gamma_B}^{HH} p_{HH} + o_{\underline{\gamma}}^{LH_{AI}} p_{HL}) + M_1(o_{\underline{\gamma}}^{LH_{AI}} p_{LH} + o_{\underline{\gamma}}^{LL_{AI}} p_{LL}), & \hat{f}_0^{(l)} < f_0 < \hat{f}_0^{(u)} \\ M_2(o_{\gamma_B}^{HH} p_{HH} + o_{\gamma_B}^{LH_{AI}} p_{LH}) + M_1(o_{\underline{\gamma}}^{HL} p_{HL} + o_{\underline{\gamma}}^{LL_{AI}} p_{LL}), & f_0 \geqslant \hat{f}_0^{(u)}. \end{cases} \tag{25}$$

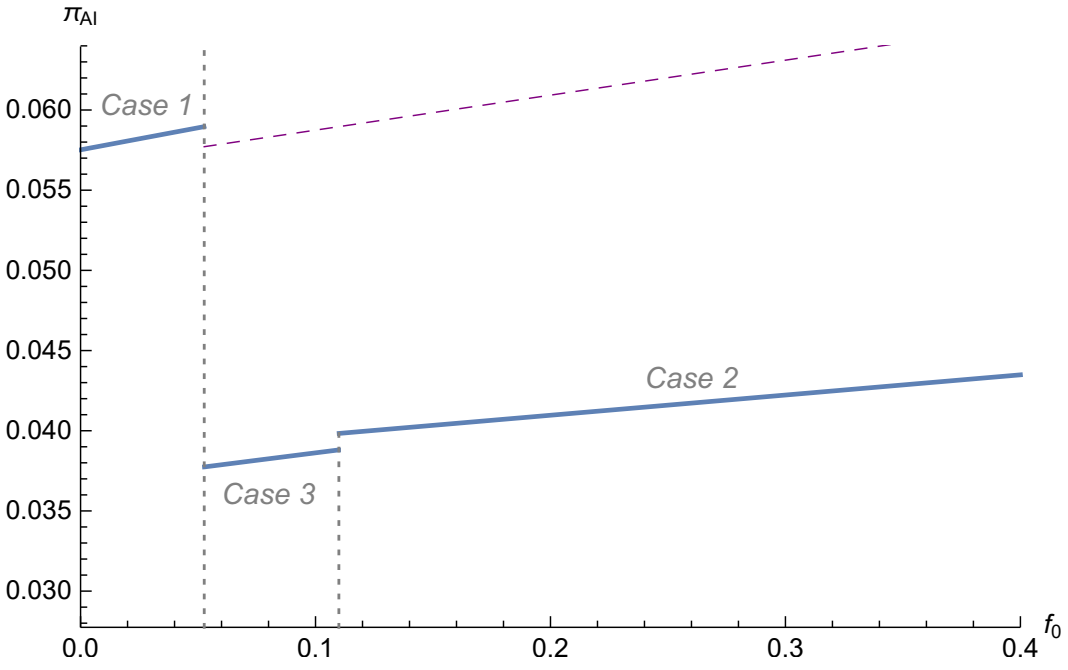


Figure 24: Trend in firm's post-AI profit with the choice of base AI efficacy.

Figure 24 depicts how the choice of f_0 affects the profit. We can see a significant drop at the boundary value between Case 1 and Case 3 ($f_0 = \hat{f}_0^{(l)}$). If the firm can pick the value of f_0, it would be better off choosing the base AI capability at $\hat{f}_0^{(l)}$. This observation is consistent with our finding in Section 3.4.2 that the firm can prevent AI's negative impact by picking a less effective AI product. When the firm decides to choose a base AI value that does not lead to performance ranking flipping, we find the following pattern:

Proposition 3.5. *The firm will pick a higher f_0 and reach a higher overall AI efficacy level when the tangible skill gap is more important in determining the performance ranking. That is, $\hat{f}_0^{(l)}$ and $\hat{f}$ increase when a_t^L decreases or a_i^L increases.*

When the tangible skill gap plays a more important role in performance ranking, that is, $\beta_t^2(a_t^H - a_t^L)$ is way larger than $\beta_i^2(a_i^H - a_i^L)$, the required AI efficacy for flipping to happen $(\hat{f})$ is larger, indicating that the firm can afford a better AI without triggering the incentive mismatch between AI and employees. From the equation for the optimal choice of f_0, or $\hat{f}_0^{(l)}$, we can easily see that this value is increasing in the flipping threshold $\hat{f}$. Therefore, when a_t^L is low, or a_i^L is high, both the overall AI efficacy and the selected base AI effectiveness are high.

Another important observation we note is that when AI relies on internal data, it will take the competition effect away when choosing the base AI that is higher than $\hat{f}_0^{(l)}$. In Figure 24, the purple dashed line is the hypothetical profit if the intensified competition for bonus salary continues. The vertical gap between the hypothetical and realized profit reveals additional demotivating effects. Under the region of Case 2, AI could make the profit with flipping even worse because (H, L) has no incentive to compete for the bonus rate. Therefore, (L, H) employees do not need extra effort to maintain their promotion status. In Case 3, it is due to the strategic choice of the (H, L) type to lower their tangible labor to keep their leading position. Overall, under the generalized model of AI, the firm's best option is still to choose a base AI level that is low enough to ensure employees' incentives remain the same as the without-AI scenario.

3.6 Endogenize Salary Rates

In this section, we extend our analysis to the firm's decision over salary rates as a profit-optimizing company would choose the reward level that maximizes the overall return. To ensure results are analytically trackable, we introduce an exogenous parameter δ to describe the relationship between the base rate $\underline{\gamma}$ and the bonus rate γ_B - that is $\underline{\gamma} = \delta\gamma_B$. The firm optimizes the profit - Equation (20) and (21) - with respect to γ_B. We use γ_B^{noAI} and γ_B^{AI} to represent the firm's equilibrium decision on the bonus salary rate, prior-to-AI, and post-AI, respectively.

To simplify the equations, we denote $D = \left(1 + \sqrt{1 - \delta^2}\right)$, $A_1 = \beta_t^2 a_t^H + \beta_i^2 a_i^H$, $A_2 =$

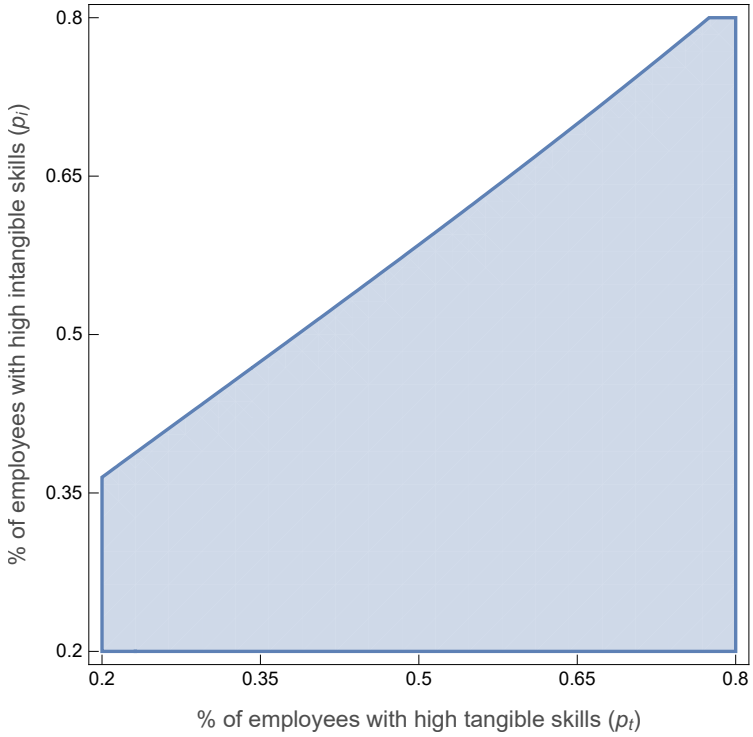
$\beta_t^2 a_t^H + \beta_i^2 a_i^L$, $A_3 = \beta_t^2 a_t^L + \beta_i^2 a_i^H$, $A_4 = \beta_t^2 a_t^L + \beta_i^2 a_i^L$, $A_5 = \beta_t^2 a_{te}^L + \beta_i^2 a_i^H$ and $A_6 = \beta_t^2 a_{te}^L + \beta_i^2 a_i^L$.

Lemma 3.3. *Before AI adoption, the firm's decision on the bonus rate is*

$$\gamma_B^{noAI} = \frac{A_1 p_{HH} + A_3 \left[D p_{HL} + \delta p_{LH}\right] + A_4 \delta p_{LL}}{2\left(A_1 p_{HH} + A_3 \left[D p_{HL} + \delta^2 p_{LH}\right] + A_4 \delta^2 p_{LL}\right)};$$

With AI, the firm's choice of salary rate becomes

$$\gamma_B^{AI} = \frac{A_1 p_{HH} + A_2 \left[D p_{LH} + \delta p_{HL}\right] + A_6 \delta p_{LL}}{2\left(A_1 p_{HH} + A_2 \left[D p_{LH} + \delta^2 p_{HL}\right] + A_6 \delta^2 p_{LL}\right)}.$$



The blue shading shows the parameter conditions under which the firm is willing to pay higher salary rates after AI.

Figure 25: Compare the optimal bonus rates (before vs after AI)

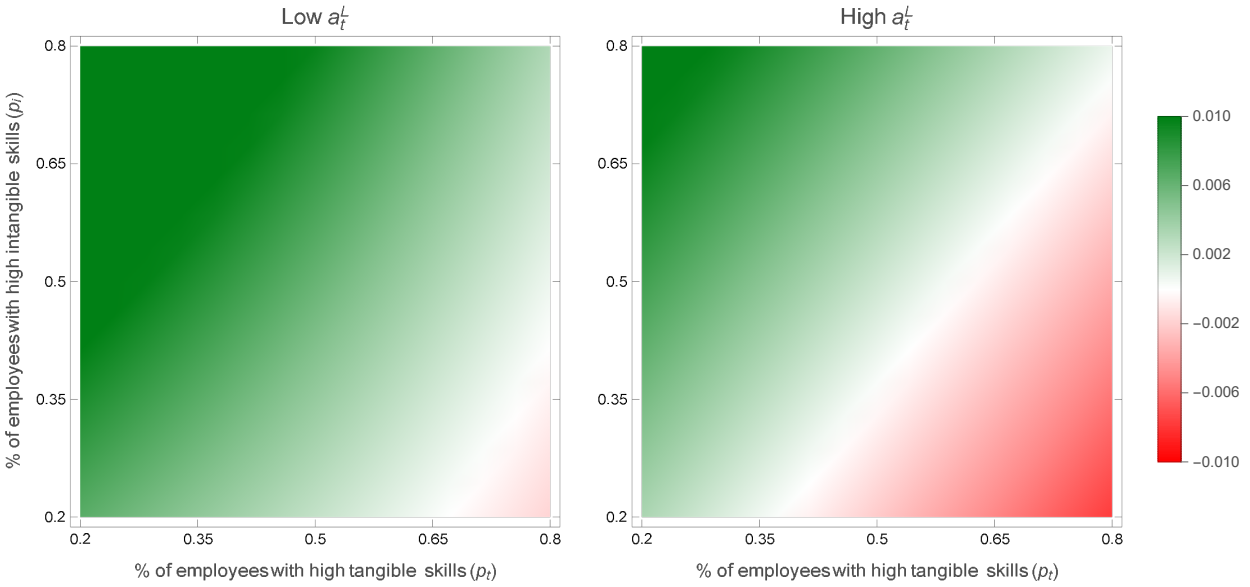
Figure 25 illustrates the employee compositions where the firm would be willing to pay a higher salary rate with AI deployment. Thereby, we make the following observation:

Observation 3.1. *The firm is willing to pay a higher rate after AI when there is a high percentage of employees in the* (H, L) *group.*

The intuition is that if there are more employees of (H, L) type (or fewer (L, H) ones), the firm ends up paying a smaller number of employees at the promoted rate as it is the (H, H) and (L, H) types that are receiving the bonus rate.

Numerically, we show that our main result (Theorem 3.1) continues to hold with Figure 26. We also identify an intriguing relationship between overall productivity and profit. We obtain the firm's total productivity and profit from the equilibrium decisions of employees and the firm. By comparing the productivity and profits, we observe that the profit is always half of the productivity. Hence, we reach the following conclusion:

Proposition 3.6. *When the firm is worse off after AI adoption, the reduction of the firm's profit comes 'entirely' from the decrease in productivity.*



This plot visually shows the Theorem 3.1 still holds when salary rates are endogenously picked by the firm, with respect to the ability of low-type (a_t^L) and employee composition (p_t & p_i).

Figure 26: Firm's profit change (after AI - before AI)

Recall that the salary rates go up after AI; one might consider the salary raise as the reason for the profit drop. However, we show with Proposition 3.6 that the loss in profit is not due to the increase in salary rate. Instead, it's because of the productivity decrease. This conclusion is counter-intuitive because it seems that when the firm pays higher rates, the

total productivity and the profit could decrease. This is because AI changes the incentives of employees. When there are more employees of (H, L) type that contribute to knowledge transfer but are being demoted, the salary rates that the firm is willing to pay go up. The groups of employees benefiting from AI and the salary increases may not be able to make up for the loss of motivation in (H, L) employees. As a result, the firm could only achieve a lower total production and profit.

3.7 Discussion & Conclusion

This study demonstrates that AI can have adverse effects when organizational factors, such as the nature of the tasks, employee composition, and compensation scheme, do not align with AI. The findings highlight the importance of employee incentive design to maximize returns on AI deployments. As AI can observe task-level data and learn from high-performing employees, employees' tangible skills are susceptible to AI-induced knowledge transfer. In competitive pay-for-performance environments, AI would induce a demotivating effect on employees who 'involuntarily' share their knowledge but are not compensated for it. The greater the proportion of employees with high levels of tangible skills, the worse the motivation-dampening effect, which in turn negatively impacts overall firm performance.

Our results have significant managerial implications. Before jumping on the bandwagon of AI adoption, firms must carefully evaluate their organizations, identify the nature of the production process and the composition of their workforce, and assess whether tangible skills play a deterministic role in production. If it is the case where more employees are likely to be demotivated by AI, choosing an AI system judiciously can help the firm to avoid pitfalls.

In future work, the modeling framework we have developed in this paper can be utilized to examine AI deployments in more complex organizational and task environments that feature job functions of specialists and generalists, specific incentives for training AI, and the role of employee non-compliance to AI instructions.

We conclude by providing a qualitative discussion of alternative policies that firms could consider to ensure that AI interacts with employee competition in a positive manner.

- **Revamp performance evaluation:** Instead of evaluating employee performance solely based on output, the firm could track employee labor input, especially tangible labor, as a byproduct of AI implementation. This way, the firm can be aware of the contributors to AI and compensate them for automatic knowledge sharing.
- **Hiring and employee retention:** Given the complexity of employee incentives, firms may be motivated to focus on the 'perfect' employees, the best performers unaffected by AI.
- **Employee training:** The firm can introduce training programs to enhance intangible skills, helping the knowledge contributors to maintain their performance lead. This can be viewed as an alternative compensation for the knowledge learned and distributed by AI.

4.0 Conclusions

This dissertation delves into two critical crucial consequences of the rapid advancements in artificial intelligence: fairness and productivity. To explore these issues, we select two settings: online advertising and AI for human capital enhancement. Using stylized analytical models, we address two questions: What is the economic mechanism behind these undesired consequences, and how can we maximize the benefits of AI?

In the first essay, titled *"Is Fair Advertising Good for Platforms?"*, we uncover that the biased delivery of economic opportunity ads is not only caused by discriminative advertisers and unfair ad-auction algorithms but also by the difference in advertisers' incentives. Among the fair advertising policies that ad platforms could impose, we find that equal-exposure is the best for the platform and overall fairness level. Meanwhile, the equal-treatment policy, which has been partially implemented, can be even worse than doing nothing. In the second essay– *"Backfiring AI? Examining AI Deployment in Pay-For-Performance Regimes"*–we provide a possible explanation for the modern productivity paradox: AI-enabled knowledge transfer may lead to a demotivating effect on employee incentives, resulting in lower productivity and profit. To prevent such negative impacts, companies should carefully choose an AI system that fits the production environment.

These findings have broad implications for business owners: those who want to reap the benefits of AI while avoiding negative externalities must carefully consider the economic incentives that AI creates. Previous studies have shown that factors such as user perception, trust, and intention, which are inherent to user behavior, can significantly impact the intended adoption of AI. Findings from the two essays demonstrate that even in the absence of any obstacles to AI usage, economic factors can still hinder overall performance. This is because the use of AI to create new markets or automate business activities could dramatically change the competition dynamic among stakeholders, either by creating a new set of incentives or by removing existing competition. For instance, unlike offline marketing channels, online advertising introduces direct competition among advertisers owing to the exclusive nature of user impressions, which can lead to biased delivery for economic opportu-

nity ads. AI for performance enhancement can also have a demotivating effect on employees' competitive spirit, which can negatively impact performance.

Similarly, other AI applications that companies may consider, such as personalized recommendations, virtual assistants, and generative AI tools, may also introduce incentives that are not aligned with the overall goal of AI. Therefore, organizations must understand the incentives of stakeholders affected by AI and consider their strategic behavior in response to the updated competition dynamics. By doing so, they can avoid negative externalities while optimizing the benefits of AI.

As regulatory attention and research effort lag behind, there are no regulations on AI fairness and only limited guidelines on AI applications for human capital management. One limitation of this study is that it is too early to have empirical evidence validating our model results. Nevertheless, our models are robust in many real situations. In addition, this research provides a foundation for future empirical research in the area of AI fairness and productivity.

Appendix A Proofs for Chapter 2

A.1 Proof of Lemma 2.1

From (4) and (5), the profit maximization problem of E and R can be written as

$$\max_{e_{Ep},\, e_{Er}} \pi_{E} = \max_{e_{Ep},\, e_{Er}} \alpha_{E} \left(N_p \frac{e_{Ep}}{e_{Ep} + e_{Rp}} + N_r \frac{e_{Er}}{e_{Er} + e_{Rr}} \right) - (e_{Ep} + e_{Er}), \tag{26}$$

$$\max_{e_{Rp},\, e_{Rr}} \pi_{R} = \max_{e_{Rp},\, e_{Rr}} \left(\alpha_{Rp} N_p \frac{e_{Rp}}{e_{Ep} + e_{Rp}} + \alpha_{Rr} N_r \frac{e_{Rr}}{e_{Er} + e_{Rr}} \right) - (e_{Rp} + e_{Rr}). \tag{27}$$

First-order conditions are:

$$\frac{\partial \pi_{E}}{\partial e_{Ep}} = \alpha_{E} N_p \frac{e_{Rp}}{(e_{Ep} + e_{Rp})^2} - 1 = 0, \tag{28}$$

$$\frac{\partial \pi_{E}}{\partial e_{Er}} = \alpha_{E} N_r \frac{e_{Rr}}{(e_{Er} + e_{Rr})^2} - 1 = 0, \tag{29}$$

$$\frac{\partial \pi_{R}}{\partial e_{Rp}} = \frac{\alpha_{Rp} N_p e_{Ep}}{(e_{Ep} + e_{Rp})^2} - 1 = 0, \tag{30}$$

$$\frac{\partial \pi_{R}}{\partial e_{Rr}} = \frac{\alpha_{Rr} N_r e_{Er}}{(e_{Er} + e_{Rr})^2} - 1 = 0. \tag{31}$$

From the following first-order conditions, we can obtain the equilibrium levels of ad expense in Lemma 2.1 by substituting $e_{Rp} = -e_{Ep} + \sqrt{\alpha_{Rp} N_p e_{Ep}}$ (from equation (30)) into (28) and $e_{Rr} = -e_{Er} + \sqrt{\alpha_{Rr} N_r e_{Er}}$ (from (31)) into (29).

$$e_{Ep}^{NR} = \frac{\alpha_{E}^2 \alpha_{Rp} N_p}{(\alpha_{E} + \alpha_{Rp})^2},\ e_{Er}^{NR} = \frac{\alpha_{E}^2 \alpha_{Rr} N_r}{(\alpha_{E} + \alpha_{Rr})^2},$$

$$e_{Rp}^{NR} = \frac{\alpha_{E} \alpha_{Rp}^2 N_p}{(\alpha_{E} + \alpha_{Rp})^2},\ e_{Rr}^{NR} = \frac{\alpha_{E} \alpha_{Rr}^2 N_r}{(\alpha_{E} + \alpha_{Rr})^2}.$$

To make sure the solution is a global maximizer over the assumed parameter ranges, we check the definiteness of the Hessians:

$$H_E = \begin{pmatrix} -\frac{2\alpha_E N_p e_{Rp}}{(e_{Ep}+e_{Rp})^3} & 0 \\ 0 & -\frac{2\alpha_E N_r e_{Rr}}{(e_{Er}+e_{Rr})^3} \end{pmatrix}$$

$$H_R = \begin{pmatrix} -\frac{2\alpha_{Rp}N_p e_{Ep}}{(e_{Ep}+e_{Rp})^3} & 0 \\ 0 & -\frac{2\alpha_{Rr}N_r e_{Er}}{(e_{Er}+e_{Rr})^3} \end{pmatrix}$$

It is obvious that both Hessians are negative definite over the entire parameter ranges.

Following the definition of ad share f_{ij}, we have:

$$f_{Ep}^{NR} = \frac{e_{Ep}^{NR}}{e_{Ep}^{NR} + e_{Rp}^{NR}} = \frac{\alpha_E}{\alpha_E + \alpha_{Rp}}$$
$$f_{Er}^{NR} = \frac{e_{Er}^{NR}}{e_{Er}^{NR} + e_{Rr}^{NR}} = \frac{\alpha_E}{\alpha_E + \alpha_{Rr}}$$

■

A.2 Proof of Proposition 2.1

Comparison of ad expense levels on *protected* users between advertisers:

$$\begin{aligned} & e_{Rp}^{NR} - e_{Ep}^{NR} \\ =& \frac{\alpha_E \alpha_{Rp}^2 N_p}{(\alpha_E + \alpha_{Rp})^2} - \frac{\alpha_E^2 \alpha_{Rp} N_p}{(\alpha_E + \alpha_{Rp})^2} \\ =& \frac{\alpha_E \alpha_{Rp} N_p}{(\alpha_E + \alpha_{Rp})^2} (\alpha_{Rp} - \alpha_E) > 0 \text{ when } \alpha_{Rp} > \alpha_E \end{aligned}$$

Comparison of ad expense levels on *regular* users between advertisers:

$$\begin{aligned} & e_{Rr}^{NR} - e_{Er}^{NR} \\ =& \frac{\alpha_E \alpha_{Rr}^2 N_r}{(\alpha_E + \alpha_{Rr})^2} - \frac{\alpha_E^2 \alpha_{Rr} N_r}{(\alpha_E + \alpha_{Rr})^2} \\ =& \frac{\alpha_E \alpha_{Rr} N_r}{(\alpha_E + \alpha_{Rr})^2} (\alpha_{Rr} - \alpha_E) < 0 \text{ when } \alpha_{Rr} < \alpha_E \end{aligned}$$

■

A.3 Proof of Lemma 2.2

With the fairness constraint $\frac{e_{Ep}}{N_p} = \frac{e_{Er}}{N_r}$, we have $e_{Ep} = \frac{N_p}{N_r} e_{Er}$. To simplify the notations, we define $n = \frac{N_p}{N_r}$ and substitute the constraint into advertisers' profit maximization problem:

$$\max_{e_{Er}} \pi_E = \max_{e_{Er}} \alpha_E \left(N_p \frac{n e_{Er}}{n e_{Er} + e_{Rp}} + N_r \frac{e_{Er}}{e_{Er} + e_{Rr}} \right) - (\frac{N_p}{N_r} e_{Er} + e_{Er}),$$

$$\max_{e_{Rp},\, e_{Rr}} \pi_R = \max_{e_{Rp},\, e_{Rr}} \left(\alpha_{Rp} N_p \frac{e_{Rp}}{n e_{Er} + e_{Rp}} + \alpha_{Rr} N_r \frac{e_{Rr}}{e_{Er} + e_{Rr}} \right) - (e_{Rp} + e_{Rr}).$$

The FOCs are:

$$\frac{\partial \pi_E}{\partial e_{Er}} = \alpha_E \left[N_p \frac{n e_{Rp}}{(n e_{Er} + e_{Rp})^2} + N_r \frac{e_{Rr}}{(e_{Er} + e_{Rr})^2} \right] - (n+1) = 0 \tag{32}$$

$$\frac{\partial \pi_R}{\partial e_{Rp}} = \frac{\alpha_{Rp} N_p n e_{Er}}{(n e_{Er} + e_{Rp})^2} - 1 = 0 \tag{33}$$

$$\frac{\partial \pi_R}{\partial e_{Rr}} = \frac{\alpha_{Rr} N_r e_{Er}}{(e_{Er} + e_{Rr})^2} - 1 = 0 \tag{34}$$

From the conditions (33) and (34), we obtain that $e_{Rp} = -n e_{Er} + \sqrt{\alpha_{Rp} N_p n e_{Er}}$ and $e_{Rr} = -e_{Er} + \sqrt{\alpha_{Rr} N_r e_{Er}}$. Substituting these back into (28), is straightforward to solve for the solutions in Lemma 2.2.

To make sure the solution is a optimal over the assumed parameter ranges, we check the second-order conditions.

$$\frac{\partial^2 \pi_E}{\partial e_{Er}^2} = -2\alpha_E \left[N_p \frac{n^2 e_{Rp}}{(n e_{Er} + e_{Rp})^3} + N_r \frac{e_{Rr}}{(e_{Er} + e_{Rp})^3} \right] < 0$$

$$H_R = \begin{pmatrix} -\frac{2\alpha_{Rp} N_p n e_{Er}}{(n e_{Er} + e_{Rp})^3} & 0 \\ 0 & -\frac{2\alpha_{Rr} N_r e_{Er}}{(e_{Er} + e_{Rr})^3} \end{pmatrix}$$

It is straightforward to see that second-order conditions are satisfied for the optimal solution.

Follow the definition of ad share f_{ij} and the notation $B_1 = \sqrt{\alpha_{Rp}} \alpha_{Rr} N_p + \sqrt{\alpha_{Rr}} \alpha_{Rp} N_r$ and $B_2 = (\alpha_E + \alpha_{Rp}) \alpha_{Rr} N_p + \alpha_{Rp} (\alpha_E + \alpha_{Rr}) N_r$, we have:

$$f_{Ep}^{ER} = \frac{e_{Ep}^{ET}}{e_{Ep}^{ET} + e_{Rp}^{ET}} = \frac{\alpha_E B_1}{\sqrt{\alpha_{Rp}} B_2},$$

$$f_{Er}^{ET} = \frac{e_{Er}^{ET}}{e_{Er}^{ET} + e_{Rr}^{ET}} = \frac{\alpha_E B_1}{\sqrt{\alpha_{Rr}} B_2}.$$

■

A.4 Proof of Proposition 2.2

Compare two advertisers' expense on the protected users

$$\begin{aligned}
e_{Ep}^{ET} - e_{Rp}^{ET} =& N_p \left(\frac{B_1}{B_2}\alpha_E\right)^2 - \frac{B_1\left[\sqrt{\alpha_{Rp}}\alpha_{Rr}(N_p+N_r) + \alpha_E\left(\sqrt{\alpha_{Rp}} - \sqrt{\alpha_{Rr}}\right)N_r\right]}{B_2^2}\alpha_E\alpha_{Rp}N_p \\
=& \frac{\alpha_E N_p B_1}{B_2^2}\left(\alpha_E B_1 - \alpha_{Rp}\left[\sqrt{\alpha_{Rp}}\alpha_{Rr}(N_p+N_r) + \alpha_E\left(\sqrt{\alpha_{Rp}} - \sqrt{\alpha_{Rr}}\right)N_r\right]\right) \\
=& \frac{\alpha_E N_p B_1}{B_2^2}\left[\alpha_E\left(\sqrt{\alpha_{Rp}}\alpha_{Rr}N_p + 2\sqrt{\alpha_{Rr}}\alpha_{Rp}N_r - \sqrt{\alpha_{Rp}}\alpha_{Rp}N_r\right) - \sqrt{\alpha_{Rp}}\alpha_{Rp}\alpha_{Rr}(N_p+N_r)\right]
\end{aligned}$$

Therefore, we can see that $e_{Ep}^{ET} < e_{Rp}^{ET}$ when $\alpha_E < \frac{\alpha_{Rp}\alpha_{Rr}(N_p+N_r)}{2\sqrt{\alpha_{Rp}\alpha_{Rr}}N_r + \alpha_{Rr}N_p - \alpha_{Rp}N_r}$. We denote this upper threshold for α_E as $\hat{\alpha}_E^{ET-h}$. Notably, we find that $\hat{\alpha}_E^{ET-h}$ is larger than α_{Rp} with the following comparison:

$$\begin{aligned}
\hat{\alpha}_E^{ET-h} - \alpha_{Rp} =& \frac{\alpha_{Rp}\alpha_{Rr}(N_p+N_r)}{2\sqrt{\alpha_{Rp}\alpha_{Rr}}N_r + \alpha_{Rr}N_p - \alpha_{Rp}N_r} - \alpha_{Rp} \\
=& \alpha_{Rp}\left[\frac{\alpha_{Rr}(N_p+N_r) - 2\sqrt{\alpha_{Rp}\alpha_{Rr}}N_r - \alpha_{Rr}N_p + \alpha_{Rp}N_r}{2\sqrt{\alpha_{Rp}\alpha_{Rr}}N_r + \alpha_{Rr}N_p - \alpha_{Rp}N_r}\right] \\
=& \alpha_{Rp}\left[\frac{\left(\alpha_{Rp} - 2\sqrt{\alpha_{Rp}\alpha_{Rr}} + \alpha_{Rr}\right)N_r}{2\sqrt{\alpha_{Rp}\alpha_{Rr}}N_r + \alpha_{Rr}N_p - \alpha_{Rp}N_r}\right] > 0.
\end{aligned}$$

Compare two advertisers' expense on the regular users

$$\begin{aligned}
e_{Er}^{ET} - e_{Rr}^{ET} =& N_r \left(\frac{B_1}{B_2}\alpha_E\right)^2 - \frac{B_1\left[\sqrt{\alpha_{Rr}}\alpha_{Rp}(N_p+N_r) + \alpha_E\left(\sqrt{\alpha_{Rr}} - \sqrt{\alpha_{Rp}}\right)N_p\right]}{B_2^2}\alpha_E\alpha_{Rr}N_r \\
=& \frac{\alpha_E N_r B_1}{B_2^2}\left(\alpha_E B_1 - \alpha_{Rr}\left[\sqrt{\alpha_{Rr}}\alpha_{Rp}(N_p+N_r) + \alpha_E\left(\sqrt{\alpha_{Rr}} - \sqrt{\alpha_{Rp}}\right)N_p\right]\right) \\
=& \frac{\alpha_E N_r B_1}{B_2^2}\left[\alpha_E\left(\sqrt{\alpha_{Rr}}\alpha_{Rr}N_p + 2\sqrt{\alpha_{Rp}}\alpha_{Rr}N_p + \alpha_{Rp}\sqrt{\alpha_{Rr}}N_r\right) - \alpha_{Rp}\alpha_{Rr}\sqrt{\alpha_{Rr}}(N_p+N_r)\right]
\end{aligned}$$

Therefore, we can see that $e_{Er}^{ET} > e_{Rr}^{ET}$ when $\alpha_E < \frac{\alpha_{Rp}\alpha_{Rr}(N_p+N_r)}{2\sqrt{\alpha_{Rp}\alpha_{Rr}}N_p - \alpha_{Rr}N_p + \alpha_{Rp}N_r}$. We denote this upper threshold for α_E as $\hat{\alpha}_E^{ET-l}$.

Compare ad E's market share between user groups

With $f_{Ep}^{ET} - f_{Er}^{ET} = \frac{\alpha_E B_1}{\sqrt{\alpha_{Rp}}B_2} - \frac{\alpha_E B_1}{\sqrt{\alpha_{Rr}}B_2}$, we can easily see that $f_{Ep}^{ET} < f_{Er}^{ET}$ as long as $\alpha_{Rr} < \alpha_{Rp}$.

■

A.5 Proof of Lemma 2.3

First, we analyze the case $f_{Ep} \leqslant f_{Er}$. The equal-exposure constraint requires the platform to adjust advertiser E's ad spending on the protected users by Δe, so that the adjusted ad share satisfies the condition $f^a_{Ep} = f_{Er}$; that is:

$$\frac{e_{Ep} + \Delta e}{e_{Ep} + \Delta e + e_{Rp}} = \frac{e_{Er}}{e_{Er} + e_{Rr}}.$$

Obtain the expression of Δe in terms of advertisers' decisions and substitute it back into advertisers' profit functions. Next, we take the FOCs:

$$\begin{aligned}
&\frac{\partial \pi_{E}}{\partial e_{Ep}} = -1, \\
&\frac{\partial \pi_{R}}{\partial e_{Rp}} = -1, \\
&\frac{\partial \pi_{E}}{\partial e_{Er}} = \alpha_{E}(N_p + N_r)\frac{e_{Rr}}{\left(e_{Er} + e_{Rr}\right)^2} - 1 = 0, \\
&\frac{\partial \pi_{R}}{\partial e_{Rr}} = \left(\alpha_{Rp}N_p + \alpha_{Rr}N_r\right)\frac{e_{Er}}{\left(e_{Er} + e_{Rr}\right)^2} - 1 = 0.
\end{aligned}$$

From the FOCs, it is straightforward to find the solution that both advertisers now have no incentive to spend money on the protected users (i.e., $e_{Ep} = e_{Rp} = 0$) and the ad expenditures on the regular users are the equilibrium (i) shown in Lemma 2.3. We also obtain the ads share, and the profit for the platform and advertisers:

$$\begin{aligned}
f^{CEE}_{Ep} &= f^{CEE}_{Er} = \frac{\alpha_{E}\left(N_p + N_r\right)}{\left(\alpha_{E} + \alpha_{Rp}\right)N_p + \left(\alpha_{E} + \alpha_{Rr}\right)N_r} \\
\pi^{CEE}_{E} &= \frac{\alpha^3_{E}\left(N_p + N_r\right)^3}{\left[\left(\alpha_{E} + \alpha_{Rp}\right)N_p + \left(\alpha_{E} + \alpha_{Rr}\right)N_r\right]^2} \\
\pi^{CEE}_{R} &= \frac{\left(\alpha_{Rp}N_p + \alpha_{Rr}N_r\right)^3}{\left[\left(\alpha_{E} + \alpha_{Rp}\right)N_p + \left(\alpha_{E} + \alpha_{Rr}\right)N_r\right]^2} \\
\pi^{CEE}_{P} &= \frac{\alpha_{E}\left(N_p + N_r\right)\left(\alpha_{Rp}N_p + \alpha_{Rr}N_r\right)}{\left[\left(\alpha_{E} + \alpha_{Rp}\right)N_p + \left(\alpha_{E} + \alpha_{Rr}\right)N_r\right]}
\end{aligned}$$

To make sure the solution is a optimal over the assumed parameter ranges, we check the second order conditions (see the following) and show that the equilibrium value is indeed optimal:

$$\frac{\partial^2 \pi_{E}}{\partial e_{Er}^2} = -\frac{2e_{Rr}\alpha_{E}\left(N_p + N_r\right)}{\left(e_{Er} + e_{Rr}\right)^3} < 0,$$
$$\frac{\partial^2 \pi_{E}}{\partial e_{Er}^2} = -\frac{2e_{Er}\left(\alpha_{Rp}N_p + \alpha_{Rr}N_r\right)}{\left(e_{Er} + e_{Rr}\right)^3} < 0.$$

Next, we solve for the mirroring case of $f_{Ep} > f_{Er}$ by following the same method, with the platform's intervention Δe applied to advertiser E's ad spending on the regular group. The FOCs now become;

$$\frac{\partial \pi_{E}}{\partial e_{Ep}} = \alpha_{E}(N_p + N_r)\frac{e_{Rp}}{\left(e_{Ep} + e_{Rp}\right)^2} - 1 = 0,$$
$$\frac{\partial \pi_{R}}{\partial e_{Rp}} = \left(\alpha_{Rp}N_p + \alpha_{Rr}N_r\right)\frac{e_{Ep}}{\left(e_{Ep} + e_{Rp}\right)^2} - 1 = 0,$$
$$\frac{\partial \pi_{E}}{\partial e_{Er}} = -1,$$
$$\frac{\partial \pi_{R}}{\partial e_{Rr}} = -1.$$

We can see that both advertisers would allocate no ad expense on the regular and their decision on the protected (e_{Ep} and e_{Rp}) are shown in equilibrium (ii) of Lemma 2.3. Another interesting observation is that the welfare of every stakeholder remains the same as in equilibrium (i). ■

A.6 Proof of Proposition 2.3

We focus on equilibrium (i) in Lemma 2.3 and compare two advertisers' spending on the regular users:

$$\begin{aligned} & e_{Rr}^{CEE} - e_{Er}^{CEE} \\ =& \frac{\alpha_{E}\left(N_p + N_r\right)\left(\alpha_{Rp}N_p + \alpha_{Rr}N_r\right)}{2\left[\left(\alpha_{E} + \alpha_{Rp}\right)N_p + \left(\alpha_{E} + \alpha_{Rr}\right)N_r\right]^2}\left[\left(\alpha_{Rp}N_p + \alpha_{Rr}N_r\right) - \alpha_{E}\left(N_p + N_r\right)\right] \end{aligned}$$

For the main term $\left(\alpha_{Rp}N_p + \alpha_{Rr}N_r\right) - \alpha_{E}\left(N_p + N_r\right)$ to be positive, that is $e_{Rr}^{CEE} > e_{Er}^{CEE}$, we require α_{E} satisfies:

$$\alpha_{E} \leqslant \bar{\alpha}_{R} = \frac{\alpha_{Rp}N_p + \alpha_{Rr}N_r}{N_p + N_r}$$

■

A.7 Proof of Lemma 2.5

First, we analyze the case $f_{Ep} \leqslant f_{Er}$. By substituting platform's intervention constraint into R's profit function, we have the first order condition on the retailer's choice of e_{Rp}: $\frac{\partial \pi_R}{\partial e_{Rp}} < 0$, indicating that the retailer has no incentive to spend more effort than the minimum required on the protected group. Given that the condition range $f_{Ep} \leqslant f_{Er}$ and the equal-treatment constraint $\frac{e_{Ep}}{e_{Er}} = \frac{N_p}{N_r}$, we can see that the minimum budget on the protected group is $\frac{N_p}{N_r} e_{Rr}$. With $e_{Rp} = \frac{N_p}{N_r} e_{Rr}$ and $e_{Ep} = \frac{N_p}{N_r} e_{Er}$, we are able to solve the optimal decisions e_{Er} and e_{Rr} from the following conditions:

$$\frac{\partial \pi_E}{\partial e_{Er}} = (N_p + N_r) \left(\frac{e_{Rr} \alpha_E}{(e_{Er} + e_{Rr})^2} - \frac{1}{N_r} \right) = 0 \tag{35}$$

$$\frac{\partial \pi_R}{\partial e_{Rr}} = \frac{e_{Er} (\alpha_{Rp} N_p + \alpha_{Rr} N_r)}{2 (e_{Er} + e_{Rr})^2} - \frac{N_p + N_r}{N_r} = 0 \tag{36}$$

It is straightforward to solve for the optimized e_{Er} and e_{Rr} and obtain the results in Lemma 2.5 from the FOCs.

Following the same logic, one can easily see that in the mirroring case of $f_{Ep} \geqslant f_{Er}$ (implying $e_{Rp} \leqslant \frac{N_p}{N_r} e_{Rr}$), the platform's intervention would result in $f_{Rr} = f_{Rp} = \frac{e_{Rp}}{e_{Ep} + e_{Rp}}$. Thus, the first order condition on the retailer's choice of e_{Rr}: $\frac{\partial \pi_R}{\partial e_{Rr}} < 0$ and R would only spend minimum effort on the regular group, that is $e_{Rr} = \frac{N_r}{N_p} e_{Rp}$. Therefore, two cases lead to the same equilibrium result.

To ensure the solutions are local maximizers over the assumed parameter ranges, we check the second-order conditions. For example, for the case $f_{Ep} \leqslant f_{Er}$ we have

$$\frac{\partial^2 \pi_E}{\partial e_{Er}^2} = -\frac{2\alpha_E (N_p + N_r) e_{Rr}}{(e_{Er} + e_{Rr})^3} < 0$$

$$H_R = \begin{pmatrix} 0 & 0 \\ 0 & -\frac{2e_{Er}(\alpha_{Rp} N_p + \alpha_{Rr} N_r)}{(e_E + e_{Rr})^3} \end{pmatrix} \text{ is negative semidefinite.}$$

We can see that second-order conditions are satisfied for a local maximum. ■

A.8 Proof of Proposition 2.5

From Lemma 2.5, we can easily obtain that $\frac{e_{Ep}^{EET}}{e_{Er}^{EET}} = \frac{e_{Rp}^{EET}}{e_{Rr}^{EET}} = \frac{N_p}{N_r}$. Next, we compare the ad budget between two advertisers:

$$\begin{aligned}
&e_{Rp}^{EET} - e_{Ep}^{EET} \\
=&\frac{\alpha_{E}(N_p+N_r)\left(\alpha_{Rp}N_p+\alpha_{Rr}N_r\right)N_p}{\left[\left(\alpha_{E}+\alpha_{Rp}\right)N_p+\left(\alpha_{E}+\alpha_{Rr}\right)N_r\right]^2}\left[\left(\alpha_{Rp}N_p+\alpha_{Rr}N_r\right)-\alpha_{E}\left(N_p+N_r\right)\right] \\
&e_{Rr}^{EET} - e_{Er}^{EET} \\
=&\frac{\alpha_{E}(N_p+N_r)\left(\alpha_{Rp}N_p+\alpha_{Rr}N_r\right)N_r}{\left[\left(\alpha_{E}+\alpha_{Rp}\right)N_p+\left(\alpha_{E}+\alpha_{Rr}\right)N_r\right]^2}\left[\left(\alpha_{Rp}N_p+\alpha_{Rr}N_r\right)-\alpha_{E}\left(N_p+N_r\right)\right]
\end{aligned}$$

We can see that these two comparisons share the main term. Therefore, to have $e_{Ep}^{EET} < e_{Rp}^{EET}$ and $e_{Er}^{EET} < e_{Rr}^{EET}$, we require α_{E} satisfies:

$$\alpha_{E} \leqslant \bar{\alpha}_{R} = \frac{\alpha_{Rp}N_p+\alpha_{Rr}N_r}{N_p+N_r}$$

■

A.9 Proof of Proposition 2.6

From the equilibrium values under two equal-exposure policies, we obtain that

$$\begin{aligned}
f_{Ep}^{CEE} = f_{Ep}^{DEE} = f_{Ep}^{EET} = f_{Er}^{CEE} = f_{Er}^{DEE} = f_{Er}^{EET} &= \frac{\alpha_{E}\left(N_p+N_r\right)}{\left(\alpha_{E}+\alpha_{Rp}\right)N_p+\left(\alpha_{E}+\alpha_{Rr}\right)N_r} \\
\pi_{E}^{CEE} = \pi_{E}^{DEE} = \pi_{E}^{EET} &= \frac{\alpha_{E}^3\left(N_p+N_r\right)^3}{\left[\left(\alpha_{E}+\alpha_{Rp}\right)N_p+\left(\alpha_{E}+\alpha_{Rr}\right)N_r\right]^2} \\
\pi_{R}^{CEE} = \pi_{R}^{DEE} = \pi_{R}^{EET} &= \frac{\left(\alpha_{Rp}N_p+\alpha_{Rr}N_r\right)^3}{\left[\left(\alpha_{E}+\alpha_{Rp}\right)N_p+\left(\alpha_{E}+\alpha_{Rr}\right)N_r\right]^2} \\
\pi_{P}^{CEE} = \pi_{P}^{DEE} = \pi_{P}^{EET} &= \frac{\alpha_{E}\left(N_p+N_r\right)\left(\alpha_{Rp}N_p+\alpha_{Rr}N_r\right)}{\left[\left(\alpha_{E}+\alpha_{Rp}\right)N_p+\left(\alpha_{E}+\alpha_{Rr}\right)N_r\right]}
\end{aligned}$$

■

A.10 Proof of Theorem 2.1

EE vs. NR

$$
\begin{aligned}
&\pi_{P}^{EE}-\pi_{P}^{NR}\\
=&\frac{\alpha_{E}\left(N_{p}+N_{r}\right)\left(\alpha_{Rp}N_{p}+\alpha_{Rr}N_{r}\right)}{\left[\left(\alpha_{E}+\alpha_{Rp}\right)N_{p}+\left(\alpha_{E}+\alpha_{Rr}\right)N_{r}\right]}-\alpha_{E}\left(\frac{\alpha_{Rp}N_{p}}{\alpha_{E}+\alpha_{Rp}}+\frac{\alpha_{Rr}N_{r}}{\alpha_{E}+\alpha_{Rr}}\right)\\
=&\alpha_{E}\frac{\begin{array}{c}\left(N_{p}+N_{r}\right)\left(\alpha_{Rp}N_{p}+\alpha_{Rr}N_{r}\right)\left(\alpha_{E}+\alpha_{Rp}\right)\left(\alpha_{E}+\alpha_{Rr}\right)\\-\left[\alpha_{Rp}N_{p}\left(\alpha_{E}+\alpha_{Rr}\right)+\alpha_{Rr}N_{r}\left(\alpha_{E}+\alpha_{Rp}\right)\right]\left[\left(\alpha_{E}+\alpha_{Rp}\right)N_{p}+\left(\alpha_{E}+\alpha_{Rr}\right)N_{r}\right]\end{array}}{\left[\left(\alpha_{E}+\alpha_{Rp}\right)N_{p}+\left(\alpha_{E}+\alpha_{Rr}\right)N_{r}\right]\left(\alpha_{E}+\alpha_{Rp}\right)\left(\alpha_{E}+\alpha_{Rr}\right)}\\
=&\alpha_{E}\frac{\begin{array}{c}\left[\alpha_{Rp}N_{p}^{2}+\left(\alpha_{Rp}+\alpha_{Rr}\right)N_{p}N_{r}+\alpha_{Rr}N_{r}^{2}\right]\left(\alpha_{E}+\alpha_{Rp}\right)\left(\alpha_{E}+\alpha_{Rr}\right)-\alpha_{Rp}N_{p}^{2}\left(\alpha_{E}+\alpha_{Rp}\right)\left(\alpha_{E}+\alpha_{Rr}\right)\\-\alpha_{Rp}N_{p}N_{r}\left(\alpha_{E}+\alpha_{Rr}\right)^{2}-\alpha_{Rr}N_{p}N_{r}\left(\alpha_{E}+\alpha_{Rp}\right)^{2}-\alpha_{Rr}N_{r}^{2}\left(\alpha_{E}+\alpha_{Rp}\right)\left(\alpha_{E}+\alpha_{Rr}\right)\end{array}}{\left[\left(\alpha_{E}+\alpha_{Rp}\right)N_{p}+\left(\alpha_{E}+\alpha_{Rr}\right)N_{r}\right]\left(\alpha_{E}+\alpha_{Rp}\right)\left(\alpha_{E}+\alpha_{Rr}\right)}\\
=&\alpha_{E}\frac{N_{p}N_{r}\left[\left(\alpha_{Rp}+\alpha_{Rr}\right)\left(\alpha_{E}+\alpha_{Rp}\right)\left(\alpha_{E}+\alpha_{Rr}\right)-\alpha_{Rr}\left(\alpha_{E}+\alpha_{Rp}\right)^{2}-\alpha_{Rp}\left(\alpha_{E}+\alpha_{Rr}\right)^{2}\right]}{\left[\left(\alpha_{E}+\alpha_{Rp}\right)N_{p}+\left(\alpha_{E}+\alpha_{Rr}\right)N_{r}\right]\left(\alpha_{E}+\alpha_{Rp}\right)\left(\alpha_{E}+\alpha_{Rr}\right)}\\
=&\alpha_{E}\frac{N_{p}N_{r}\left(\alpha_{Rp}-\alpha_{Rr}\right)^{2}\alpha_{E}}{\left[\left(\alpha_{E}+\alpha_{Rp}\right)N_{p}+\left(\alpha_{E}+\alpha_{Rr}\right)N_{r}\right]\left(\alpha_{E}+\alpha_{Rp}\right)\left(\alpha_{E}+\alpha_{Rr}\right)}>0
\end{aligned}
$$

■

A.11 Proof of Theorem 2.2

EE vs. ET

$$
\pi_{P}^{EE}-\pi_{P}^{ET}=\frac{\alpha_{E}\left(N_{p}+N_{r}\right)\left(\alpha_{Rp}N_{p}+\alpha_{Rr}N_{r}\right)}{\left[\left(\alpha_{E}+\alpha_{Rp}\right)N_{p}+\left(\alpha_{E}+\alpha_{Rr}\right)N_{r}\right]}-\frac{\alpha_{E}\alpha_{Rp}\alpha_{Rr}(N_{p}^{2}+N_{r}^{2})+\alpha_{E}\sqrt{\alpha_{Rp}\alpha_{Rr}}N_{p}N_{r}(\alpha_{Rp}+\alpha_{Rr})}{(\alpha_{E}+\alpha_{Rp})\alpha_{Rr}N_{p}+\alpha_{Rp}(\alpha_{E}+\alpha_{Rr})N_{r}}
$$

The sign of the above equation is determined by numerators, thus we focus on:

$$
\begin{aligned}
&\left(N_{p}+N_{r}\right)\left(\alpha_{Rp}N_{p}+\alpha_{Rr}N_{r}\right)\left[(\alpha_{E}+\alpha_{Rp})\alpha_{Rr}N_{p}+\alpha_{Rp}(\alpha_{E}+\alpha_{Rr})N_{r}\right]\\
&\quad-\left[\alpha_{Rp}\alpha_{Rr}(N_{p}^{2}+N_{r}^{2})+\sqrt{\alpha_{Rp}\alpha_{Rr}}N_{p}N_{r}(\alpha_{Rp}+\alpha_{Rr})\right]\left[\left(\alpha_{E}+\alpha_{Rp}\right)N_{p}+\left(\alpha_{E}+\alpha_{Rr}\right)N_{r}\right]\\
=&\alpha_{E}N_{p}N_{r}(\sqrt{\alpha_{Rp}}-\sqrt{\alpha_{Rr}})^{2}\left[\alpha_{E}\left(\alpha_{Rp}+\alpha_{Rr}+\sqrt{\alpha_{Rp}\alpha_{Rr}}\right)\left(N_{p}+N_{r}\right)-\sqrt{\alpha_{Rp}\alpha_{Rr}}\left(\alpha_{Rp}N_{p}+\alpha_{Rr}N_{r}\right)\right]
\end{aligned}
$$

Therefore, when α_{E} satisfies the following condition, equal-exposure strategies dominate equal-treatment for platform's profit ($\pi_{P}^{EE}\geqslant\pi_{P}^{ET}$):

$$
\alpha_{E}\geqslant\hat{\alpha}_{E}^{\pi_{P}}=\frac{\sqrt{\alpha_{Rp}\alpha_{Rr}}\left(\alpha_{Rp}N_{p}+\alpha_{Rr}N_{r}\right)}{\left(\alpha_{Rp}+\alpha_{Rr}+\sqrt{\alpha_{Rp}\alpha_{Rr}}\right)\left(N_{p}+N_{r}\right)}
$$

■

A.12 Proof of Proposition 2.7

Compare the fairness level between NR and ET

$$\begin{aligned}\theta^{NR} - \theta^{ET} =& \frac{\alpha_E + \alpha_{Rr}}{\alpha_E + \alpha_{Rp}} - \frac{\sqrt{\alpha_{Rr}}}{\sqrt{\alpha_{Rp}}} \\ =& \frac{\left(\alpha_E + \alpha_{Rr}\right)\sqrt{\alpha_{Rp}} - \left(\alpha_E + \alpha_{Rp}\right)\sqrt{\alpha_{Rr}}}{\left(\alpha_E + \alpha_{Rp}\right)\sqrt{\alpha_{Rp}}} \\ =& \frac{\left(\sqrt{\alpha_{Rp}} - \sqrt{\alpha_{Rr}}\right)\left(\alpha_E - \sqrt{\alpha_{Rp}\alpha_{Rr}}\right)}{\left(\alpha_E + \alpha_{Rp}\right)\sqrt{\alpha_{Rp}}}\end{aligned}$$

When $\alpha_E > \hat{\alpha}_E^{\theta} = \sqrt{\alpha_{Rp}\alpha_{Rr}}$, $\theta^{NR} > \theta^{ET}$ and equal-treatment policy results in lower level of fairness than the baseline policy. ■

A.13 Proof of Proposition 2.8

EE vs. NR

We first calculate the market share changes between no-restriction and equal-exposure policies:

$$\begin{aligned}\Delta f_{Ep} =& f_{Ep}^{EE} - f_{Ep}^{NR} \\ =& \frac{\alpha_E\left(N_p + N_r\right)}{\left(\alpha_E + \alpha_{Rp}\right)N_p + \left(\alpha_E + \alpha_{Rr}\right)N_r} - \frac{\alpha_E}{\alpha_E + \alpha_{Rp}} \\ =& \alpha_E \frac{\left(N_p + N_r\right)\left(\alpha_E + \alpha_{Rp}\right) - \left(\alpha_E + \alpha_{Rp}\right)N_p - \left(\alpha_E + \alpha_{Rr}\right)N_r}{\left[\left(\alpha_E + \alpha_{Rp}\right)N_p + \left(\alpha_E + \alpha_{Rr}\right)N_r\right]\left(\alpha_E + \alpha_{Rp}\right)} \\ =& \alpha_E \frac{\left(\alpha_E + \alpha_{Rp}\right)N_r - \left(\alpha_E + \alpha_{Rr}\right)N_r}{\left[\left(\alpha_E + \alpha_{Rp}\right)N_p + \left(\alpha_E + \alpha_{Rr}\right)N_r\right]\left(\alpha_E + \alpha_{Rp}\right)} \\ =& \frac{\alpha_E\left(\alpha_{Rp} - \alpha_{Rr}\right)N_r}{\left[\left(\alpha_E + \alpha_{Rp}\right)N_p + \left(\alpha_E + \alpha_{Rr}\right)N_r\right]\left(\alpha_E + \alpha_{Rp}\right)} > 0;\end{aligned}$$

$$
\begin{aligned}
\Delta f_{Er} =& f_{Er}^{EE} - f_{Er}^{NR} \\
=& \frac{\alpha_{E}\left(N_p + N_r\right)}{\left(\alpha_{E} + \alpha_{Rp}\right) N_p + \left(\alpha_{E} + \alpha_{Rr}\right) N_r} - \frac{\alpha_{E}}{\alpha_{E} + \alpha_{Rr}} \\
=& \alpha_{E} \frac{\left(N_p + N_r\right)\left(\alpha_{E} + \alpha_{Rr}\right) - \left(\alpha_{E} + \alpha_{Rp}\right) N_p - \left(\alpha_{E} + \alpha_{Rr}\right) N_r}{\left[\left(\alpha_{E} + \alpha_{Rp}\right) N_p + \left(\alpha_{E} + \alpha_{Rr}\right) N_r\right]\left(\alpha_{E} + \alpha_{Rr}\right)} \\
=& \frac{\alpha_{E}\left(\alpha_{Rr} - \alpha_{Rp}\right) N_p}{\left[\left(\alpha_{E} + \alpha_{Rp}\right) N_p + \left(\alpha_{E} + \alpha_{Rr}\right) N_r\right]\left(\alpha_{E} + \alpha_{Rr}\right)} < 0.
\end{aligned}
$$

Therefore, the change in total number of users who saw ads E is:

$$
\begin{aligned}
& N_p \Delta f_{Ep} + N_r \Delta f_{Er} \\
=& \frac{\alpha_{E}\left(\alpha_{Rp} - \alpha_{Rr}\right) N_p N_r}{\left[\left(\alpha_{E} + \alpha_{Rp}\right) N_p + \left(\alpha_{E} + \alpha_{Rr}\right) N_r\right]\left(\alpha_{E} + \alpha_{Rp}\right)} + \frac{\alpha_{E}\left(\alpha_{Rr} - \alpha_{Rp}\right) N_p N_r}{\left[\left(\alpha_{E} + \alpha_{Rp}\right) N_p + \left(\alpha_{E} + \alpha_{Rr}\right) N_r\right]\left(\alpha_{E} + \alpha_{Rr}\right)} \\
=& \frac{\alpha_{E}\left(\alpha_{Rp} - \alpha_{Rr}\right) N_p N_r}{\left[\left(\alpha_{E} + \alpha_{Rp}\right) N_p + \left(\alpha_{E} + \alpha_{Rr}\right) N_r\right]} \left(\frac{1}{\alpha_{E} + \alpha_{Rp}} - \frac{1}{\alpha_{E} + \alpha_{Rr}}\right) < 0.
\end{aligned}
$$

EE vs. ET

Following the same progress, we have the market share changes between equal-exposure and equal-treatment policies as:

$$
\begin{aligned}
\Delta f_{Ep} =& f_{Ep}^{EE} - f_{Ep}^{ET} \\
=& \frac{\alpha_{E}(N_p + N_r)}{\left[\left(\alpha_{E} + \alpha_{Rp}\right) N_p + \left(\alpha_{E} + \alpha_{Rr}\right) N_r\right]} - \frac{\alpha_{E} B_1}{\sqrt{\alpha_{Rp}} B_2} \\
=& \frac{\alpha_{E}\left(\sqrt{\alpha_{Rp}} - \sqrt{\alpha_{Rr}}\right) N_r \left[\alpha_{E}(N_p + N_r)\sqrt{\alpha_{Rp}} + (N_p + N_r)\sqrt{\alpha_{Rp}}\alpha_{Rr} - \sqrt{\alpha_{Rr}} N_p(\alpha_{Rp} - \alpha_{Rr})\right]}{\left[\left(\alpha_{E} + \alpha_{Rp}\right) N_p + \left(\alpha_{E} + \alpha_{Rr}\right) N_r\right] B_2},
\end{aligned}
$$

$$
\begin{aligned}
\Delta f_{Er} =& f_{Er}^{EE} - f_{Er}^{ET} \\
=& \frac{\alpha_{E}(N_p + N_r)}{\left[\left(\alpha_{E} + \alpha_{Rp}\right) N_p + \left(\alpha_{E} + \alpha_{Rr}\right) N_r\right]} - \frac{\alpha_{E} B_1}{\sqrt{\alpha_{Rr}} B_2} \\
=& \frac{-\alpha_{E}\left(\sqrt{\alpha_{Rp}} - \sqrt{\alpha_{Rr}}\right) N_p \left[\alpha_{E}(N_p + N_r)\sqrt{\alpha_{Rr}} + (N_p + N_r)\alpha_{Rp}\sqrt{\alpha_{Rr}} + \sqrt{\alpha_{Rp}} N_r(\alpha_{Rp} - \alpha_{Rr})\right]}{\left[\left(\alpha_{E} + \alpha_{Rp}\right) N_p + \left(\alpha_{E} + \alpha_{Rr}\right) N_r\right] B_2}.
\end{aligned}
$$

It is straightforward to see that under the normal parameter condition of $\alpha_{Rr} < \alpha_{E} < \alpha_{Rp}$, we have $\Delta f_{Ep} > 0$ and $\Delta f_{Er} < 0$. Next, for the net change in the total number of users who are exposed ads E ($N_p \Delta f_{Ep} + N_r \Delta f_{Er}$), it is equivalent to:

$$
\begin{aligned}
&\left[\alpha_{E}(N_p+N_r)\sqrt{\alpha_{Rp}}+(N_p+N_r)\sqrt{\alpha_{Rp}}\alpha_{Rr}-\sqrt{\alpha_{Rr}}N_p(\alpha_{Rp}-\alpha_{Rr})\right]\\
&\quad-\left[\alpha_{E}(N_p+N_r)\sqrt{\alpha_{Rr}}+(N_p+N_r)\alpha_{Rp}\sqrt{\alpha_{Rr}}+\sqrt{\alpha_{Rp}}N_r(\alpha_{Rp}-\alpha_{Rr})\right]\\
&=\left(\alpha_{E}-\sqrt{\alpha_{Rp}\alpha_{Rr}}\right)(N_p+N_r)\left(\sqrt{\alpha_{Rp}}-\sqrt{\alpha_{Rr}}\right)-\left(\sqrt{\alpha_{Rr}}N_p+\sqrt{\alpha_{Rp}}N_r\right)(\alpha_{Rp}-\alpha_{Rr})
\end{aligned}
$$

Thus, to have the net impact on ad viewers be negative, we require α_{E} satisfies:

$$
\alpha_{E}<\hat{\alpha}_{E}^{f_E}=\sqrt{\alpha_{Rp}\alpha_{Rr}}+\frac{\left(\sqrt{\alpha_{Rp}}+\sqrt{\alpha_{Rr}}\right)\left(\sqrt{\alpha_{Rr}}N_p+\sqrt{\alpha_{Rp}}N_r\right)}{N_p+N_r}
$$

■

A.14 Proof of Lemma 2.6 & 2.7

No-restriction policy

Following the formulation of ad market share $f_{ij}^{(n)}$, we have the profit maximization for the n^{th} advertiser in each group:

$$
\begin{aligned}
\max_{e_{Ep}^{(n)},e_{Er}^{(n)}}\pi_{E}^{(n)}&=\max_{e_{Ep}^{(n)},e_{Er}^{(n)}}\alpha_{E}N_p\frac{e_{Ep}^{(n)}}{\sum_{i=1}^{N_E}e_{Ep}^{(i)}+\sum_{i=1}^{N_R}e_{Rp}^{(i)}}+\alpha_{E}N_r\frac{e_{Er}^{(n)}}{\sum_{i=1}^{N_E}e_{Er}^{(i)}+\sum_{i=1}^{N_R}e_{Rr}^{(i)}}-(e_{Ep}^{(n)}+e_{Er}^{(n)}),\\
\max_{e_{Rp}^{(n)},e_{Rr}^{(n)}}\pi_{R}^{(n)}&=\max_{e_{Rp}^{(n)},e_{Rr}^{(n)}}\alpha_{Rp}N_p\frac{e_{Rp}^{(n)}}{\sum_{i=1}^{N_E}e_{Ep}^{(i)}+\sum_{i=1}^{N_R}e_{Rp}^{(i)}}+\alpha_{Rr}N_r\frac{e_{Rr}^{(n)}}{\sum_{i=1}^{N_E}e_{Er}^{(i)}+\sum_{i=1}^{N_R}e_{Rr}^{(i)}}-(e_{Rp}^{(n)}+e_{Rr}^{(n)}).
\end{aligned}
$$

Under the no-restriction policy, FOCs are:

$$
\begin{aligned}
\frac{\partial\pi_{E}^{(n)}}{\partial e_{Ep}^{(n)}}&=\alpha_{E}N_p\frac{\sum_{i=1}^{N_E}e_{Ep}^{(i)}+\sum_{i=1}^{N_R}e_{Rp}^{(i)}-e_{Ep}^{(n)}}{\left(\sum_{i=1}^{N_E}e_{Ep}^{(i)}+\sum_{i=1}^{N_R}e_{Rp}^{(i)}\right)^2}-1=0,\\
\frac{\partial\pi_{E}^{(n)}}{\partial e_{Er}^{(n)}}&=\alpha_{E}N_r\frac{\sum_{i=1}^{N_E}e_{Er}^{(i)}+\sum_{i=1}^{N_R}e_{Rr}^{(i)}-e_{Er}^{(n)}}{\left(\sum_{i=1}^{N_E}e_{Er}^{(i)}+\sum_{i=1}^{N_R}e_{Rr}^{(i)}\right)^2}-1=0,\\
\frac{\partial\pi_{R}^{(n)}}{\partial e_{Rp}^{(n)}}&=\alpha_{Rp}N_p\frac{\sum_{i=1}^{N_E}e_{Ep}^{(i)}+\sum_{i=1}^{N_R}e_{Rp}^{(i)}-e_{Rp}^{(n)}}{\left(\sum_{i=1}^{N_E}e_{Ep}^{(i)}+\sum_{i=1}^{N_R}e_{Rp}^{(i)}\right)^2}-1=0,\\
\frac{\partial\pi_{R}^{(n)}}{\partial e_{Rr}^{(n)}}&=\alpha_{Rr}N_r\frac{\sum_{i=1}^{N_E}e_{Er}^{(i)}+\sum_{i=1}^{N_R}e_{Rr}^{(i)}-e_{Rr}^{(n)}}{\left(\sum_{i=1}^{N_E}e_{Er}^{(i)}+\sum_{i=1}^{N_R}e_{Rr}^{(i)}\right)^2}-1=0.
\end{aligned}
$$

From the best response functions, we can see that the equilibrium solutions are symmetric decisions for advertisers in the same type. We use e^*_{ij} to denote the optimal ad budget from any advertiser of type i in user group j and transform the FOCs into:

$$\begin{aligned}
\alpha_{E} N_p \frac{N_E e^*_{Ep} + N_R e^*_{Rp} - e^*_{Ep}}{\left(N_E e^*_{Ep} + N_R e^*_{Rp}\right)^2} &= 1, \\
\alpha_{E} N_r \frac{N_E e^*_{Er} + N_R e^*_{Rr} - e^*_{Er}}{\left(N_E e^*_{Er} + N_R e^*_{Rr}\right)^2} &= 1, \\
\alpha_{Rp} N_p \frac{N_E e^*_{Ep} + N_R e^*_{Rp} - e^*_{Rp}}{\left(N_E e^*_{Ep} + N_R e^*_{Rp}\right)^2} &= 1, \\
\alpha_{Rr} N_r \frac{N_E e^*_{Er} + N_R e^*_{Rr} - e^*_{Rr}}{\left(N_E e^*_{Er} + N_R e^*_{Rr}\right)^2} &= 1.
\end{aligned}$$

The solutions are:

$$\begin{aligned}
e^*_{Ep} &= \frac{\alpha_{E}\alpha_{Rp} N_p \left(N_E + N_R - 1\right)\left(N_R\left(\alpha_{E} - \alpha_{Rp}\right) + \alpha_{Rp}\right)}{\left(\alpha_{Rp} N_E + \alpha_{E} N_R\right)^2}, \\
e^*_{Er} &= \frac{\alpha_{E}\alpha_{Rr} N_r \left(N_E + N_R - 1\right)\left(N_R\left(\alpha_{E} - \alpha_{Rr}\right) + \alpha_{Rr}\right)}{\left(\alpha_{Rr} N_E + \alpha_{E} N_R\right)^2}, \\
e^*_{Rp} &= \frac{\alpha_{E}\alpha_{Rp} N_p \left(N_E + N_R - 1\right)\left(N_E\left(\alpha_{Rp} - \alpha_{E}\right) + \alpha_{E}\right)}{\left(\alpha_{Rp} N_E + \alpha_{E} N_R\right)^2}, \\
e^*_{Rr} &= \frac{\alpha_{E}\alpha_{Rr} N_r \left(N_E + N_R - 1\right)\left(N_E\left(\alpha_{Rr} - \alpha_{E}\right) + \alpha_{E}\right)}{\left(\alpha_{Rr} N_E + \alpha_{E} N_R\right)^2}.
\end{aligned}$$

By dividing both the numerator and denominator with factor $(N_E N_R)^2$, we can transform the above solution in terms of $\bar{\alpha}_{E} = \frac{\alpha_E}{N_E}$, $\bar{\alpha}_{Rp} = \frac{\alpha_{Rp}}{N_R}$ and $\bar{\alpha}_{Rr} = \frac{\alpha_{Rr}}{N_R}$ as shown in Lemma 2.6.

Next, we check the second order conditions for individual advertiser of type E and R:

$$H_E = \begin{pmatrix}
-2\alpha_{E} N_p \frac{\sum_{i=1}^{N_E} e^{(i)}_{Ep} + \sum_{i=1}^{N_R} e^{(i)}_{Rp} - e^{(n)}_{Ep}}{\left(\sum_{i=1}^{N_E} e^{(i)}_{Ep} + \sum_{i=1}^{N_R} e^{(i)}_{Rp}\right)^3} & 0 \\
0 & -2\alpha_{E} N_r \frac{\sum_{i=1}^{N_E} e^{(i)}_{Er} + \sum_{i=1}^{N_R} e^{(i)}_{Rr} - e^{(n)}_{Er}}{\left(\sum_{i=1}^{N_E} e^{(i)}_{Er} + \sum_{i=1}^{N_R} e^{(i)}_{Rr}\right)^3}
\end{pmatrix},$$

$$H_R = \begin{pmatrix}
-2\alpha_{Rp} N_p \frac{\sum_{i=1}^{N_E} e^{(i)}_{Ep} + \sum_{i=1}^{N_R} e^{(i)}_{Rp} - e^{(n)}_{Rp}}{\left(\sum_{i=1}^{N_E} e^{(i)}_{Ep} + \sum_{i=1}^{N_R} e^{(i)}_{Rp}\right)^3} & 0 \\
0 & -2\alpha_{Rr} N_r \frac{\sum_{i=1}^{N_E} e^{(i)}_{Er} + \sum_{i=1}^{N_R} e^{(i)}_{Rr} - e^{(n)}_{Rr}}{\left(\sum_{i=1}^{N_E} e^{(i)}_{Er} + \sum_{i=1}^{N_R} e^{(i)}_{Rr}\right)^3}
\end{pmatrix}.$$

It is obvious that both Hessians are negative definite over the entire parameter ranges.

With Equation 3, the platform's profit under the no-restriction policy is

$$\begin{aligned}\pi_{P}^{NR} &= (N_E + N_R - 1)\left[\frac{\alpha_E \alpha_{Rp} N_p}{\alpha_{Rp} N_E + \alpha_E N_R} + \frac{\alpha_E \alpha_{Rr} N_r}{\alpha_{Rr} N_E + \alpha_E N_R}\right] \\ &= (N_E + N_R - 1)\left[N_p \frac{\bar{\alpha}_E \bar{\alpha}_{Rp}}{\bar{\alpha}_E + \bar{\alpha}_{Rp}} + N_r \frac{\bar{\alpha}_E \bar{\alpha}_{Rr}}{\bar{\alpha}_E + \bar{\alpha}_{Rr}}\right].\end{aligned}$$

Equal-exposure policy

After adjusting for the platform's intervention (Δe) under equal-exposure policy, we have the profit maximization for the n^{th} advertiser in each group when $e_{Rp} \geqslant e_{Rr}$ as:

$$\begin{aligned}\max_{e_{Ep}^{(n)}, e_{Er}^{(n)}} \pi_E^n &= \max_{e_{Ep}^{(n)}, e_{Er}^{(n)}} \alpha_E N_p \frac{e_{Er}^{(n)}}{\sum_{i=1}^{N_E} e_{Er}^{(i)} + \sum_{i=1}^{N_R} e_{Rr}^{(i)}} + \alpha_E N_r \frac{e_{Er}^{(n)}}{\sum_{i=1}^{N_E} e_{Er}^{(i)} + \sum_{i=1}^{N_R} e_{Rr}^{(i)}} - (e_{Ep}^{(n)} + e_{Er}^{(n)}), \\ \max_{e_{Rp}^{(n)}, e_{Rr}^{(n)}} \pi_R^n &= \max_{e_{Rp}^{(n)}, e_{Rr}^{(n)}} \alpha_{Rp} N_p \frac{e_{Rr}^{(n)}}{\sum_{i=1}^{N_E} e_{Er}^{(i)} + \sum_{i=1}^{N_R} e_{Rr}^{(i)}} + \alpha_{Rr} N_r \frac{e_{Rr}^{(n)}}{\sum_{i=1}^{N_E} e_{Er}^{(i)} + \sum_{i=1}^{N_R} e_{Rr}^{(i)}} - (e_{Rp}^{(n)} + e_{Rr}^{(n)}).\end{aligned}$$

Following the same process of solving the base model equal-exposure, we know that both advertisers have no incentive to spend money on the protected users, and the FOCs of e_{Er} & e_{Rr} are:

$$\begin{aligned}\frac{\partial \pi_E^n}{\partial e_{Er}^{(n)}} &= \alpha_E (N_p + N_r) \frac{\sum_{i=1}^{N_E} e_{Er}^{(i)} + \sum_{i=1}^{N_R} e_{Rp}^{(i)} - e_{Er}^{(n)}}{\left(\sum_{i=1}^{N_E} e_E^{(i)} + \sum_{i=1}^{N_R} e_{Rp}^{(i)}\right)^2} - 1 = 0, \\ \frac{\partial \pi_E^n}{\partial e_{Rr}^{(n)}} &= (\alpha_{Rp} N_p + \alpha_{Rr} N_r) \frac{\sum_{i=1}^{N_E} e_{Er}^{(i)} + \sum_{i=1}^{N_R} e_{Rr}^{(i)} - e_{Rr}^{(n)}}{\left(\sum_{i=1}^{N_E} e_{Er}^{(i)} + \sum_{i=1}^{N_R} e_{Rr}^{(i)}\right)^2} - 1 = 0.\end{aligned}$$

With all advertisers in the same type being identical, we solve for the symmetric equilibrium:

$$\begin{aligned}e_{Er}^{EE} &= \frac{\Phi_E \Phi_R}{(\Phi_E + \Phi_R)^2} \frac{(N_E + N_R - 1)}{N_E} \left[(\alpha_E - \alpha_{Rp} + \bar{\alpha}_{Rp}) N_p + (\alpha_E - \alpha_{Rr} + \bar{\alpha}_{Rr}) N_r\right], \\ e_{Rr}^{EE} &= \frac{\Phi_E \Phi_R}{(\Phi_E + \Phi_R)^2} \frac{(N_E + N_R - 1)}{N_R} \left[(\alpha_{Rp} - \alpha_E + \bar{\alpha}_E) N_p + (\alpha_{Rr} - \alpha_E + \bar{\alpha}_E) N_r\right].\end{aligned}$$

The platform's profit is:

$$\begin{aligned}\pi_P^{EE} &= (N_E + N_R - 1) \frac{\alpha_E (N_p + N_r)(\alpha_{Rp} N_p + \alpha_{Rr} N_r)}{N_R \alpha_E (N_p + N_r) + N_E (\alpha_{Rp} N_p + \alpha_{Rr} N_r)} \\ &= (N_E + N_R - 1) \frac{\Phi_E \Phi_R}{\Phi_E + \Phi_R}.\end{aligned}$$

Next, we compare the platform's profit under two policies to check if the main result holds.

$$\pi_{P}^{EE} - \pi_{P}^{NR} = (N_E + N_R - 1)\left[\frac{\bar{\alpha}_E(N_p + N_r)(\bar{\alpha}_{Rp}N_p + \bar{\alpha}_{Rr}N_r)}{\bar{\alpha}_E(N_p + N_r) + (\bar{\alpha}_{Rp}N_p + \bar{\alpha}_{Rr}N_r)} - N_p\frac{\bar{\alpha}_E\bar{\alpha}_{Rp}}{\bar{\alpha}_E + \bar{\alpha}_{Rp}} - N_r\frac{\bar{\alpha}_E\bar{\alpha}_{Rr}}{\bar{\alpha}_E + \bar{\alpha}_{Rr}}\right]$$

We can see that this comparison is almost identical to that in the proof of Theorem 2.1 in Appendix , with the original user valuation α_E, α_{Rp} & α_{Rr} replaced with $\bar{\alpha}_E, \bar{\alpha}_{Rp}$ & $\bar{\alpha}_{Rr}$. Therefore, it is straightforward to see that $\pi_{P}^{EE} > \pi_{P}^{NR}$ for all possible values of N_E & N_R. ■

A.15 Proof of Lemma 2.8

Based on Equation (9) & (10) and the notation $\phi_E = \alpha_E(N_p + N_r)$, $\phi_R = \alpha_{Rp}N_p + \alpha_{Rr}N_r$, we can summarize the advertisers' best response functions with respect to the probability of staying (p_E & p_R).

E's best response:

$$\begin{cases} \text{Prefer to stay: } \phi_E\left(\frac{\phi_E}{\phi_E+\phi_R}\right)^2 p_R + \phi_E(1-p_R) > u_E, \\ \text{Indifferent: } \phi_E\left(\frac{\phi_E}{\phi_E+\phi_R}\right)^2 p_R + \phi_E(1-p_R) = u_E, \\ \text{Prefer to leave: } \phi_E\left(\frac{\phi_E}{\phi_E+\phi_R}\right)^2 p_R + \phi_E(1-p_R) < u_E. \end{cases}$$

R's best response:

$$\begin{cases} \text{Prefer to stay: } \phi_R\left(\frac{\phi_R}{\phi_E+\phi_R}\right)^2 p_E + \phi_R(1-p_E) > u_R, \\ \text{Indifferent: } \phi_R\left(\frac{\phi_R}{\phi_E+\phi_R}\right)^2 p_E + \phi_R(1-p_E) = u_R, \\ \text{Prefer to leave: } \phi_R\left(\frac{\phi_R}{\phi_E+\phi_R}\right)^2 p_E + \phi_R(1-p_E) < u_R. \end{cases}$$

The mixed strategy nash equilibrium requires both advertisers are indifferent in changing their strategies. Hence, we have the probabilities of staying on with the focal platform as:

$$p_E = \frac{1 - u_R/\phi_R}{1 - (\frac{\phi_R}{\phi_E+\phi_R})^2},$$
$$p_R = \frac{1 - u_E/\phi_E}{1 - (\frac{\phi_E}{\phi_E+\phi_R})^2}.$$

The optimal profit levels for advertisers and the platform in the second part of Lemma 2.8 can be easily calculated by substituting p_E & p_R into profit equations. ■

A.16 Proof of Proposition 2.9

Comparing the platform's profit under EE with that under NR:

$$\pi_{P}^{EE} - \pi_{P}^{NR} = \frac{1 - u_R/\phi_R}{1 - (\frac{\phi_R}{\phi_E+\phi_R})^2} \times \frac{1 - u_E/\phi_E}{1 - (\frac{\phi_E}{\phi_E+\phi_R})^2} \times \frac{\phi_E\phi_R}{\phi_E + \phi_R} - \left(N_p \frac{\alpha_E \alpha_{Rp}}{\alpha_E + \alpha_{Rp}} + N_r \frac{\alpha_E \alpha_{Rr}}{\alpha_E + \alpha_{Rr}} \right)$$

Therefore, when either of the following inequalities is met, the platform still benefits from equal-exposure (even with the presence of an outside option):

$$u_E < \phi_E \left[1 - \frac{1}{1 - u_R/\phi_R} \frac{(\phi_E + 2\phi_R)(2\phi_E + \phi_R)}{(\phi_E + \phi_R)^3} \pi_{P}^{NR} \right], \text{ or}$$
$$u_R < \phi_R \left[1 - \frac{1}{1 - u_E/\phi_E} \frac{(\phi_E + 2\phi_R)(2\phi_E + \phi_R)}{(\phi_E + \phi_R)^3} \pi_{P}^{NR} \right],$$

■

A.17 Proof of Lemma 2.9 & 2.10

No-restriction policy

Take FOCs of Equation (11) & (12):

$$\frac{\partial \pi_E}{\partial e_{Ep}} = \alpha_E \gamma_p \overline{N_p} \frac{2 e_{Ep} e_{Rp}}{(e_{Ep} + e_{Rp})^3} - 1 = 0,$$
$$\frac{\partial \pi_E}{\partial e_{Er}} = \alpha_E \gamma_r \overline{N_r} \frac{2 e_{Er} e_{Rr}}{(e_{Er} + e_{Rr})^3} - 1 = 0,$$
$$\frac{\partial \pi_R}{\partial e_{Rp}} = \alpha_{Rp} \gamma_p \overline{N_p} \frac{e_{Ep}(e_{Ep} - e_{Rp})}{(e_{Ep} + e_{Rp})^3} - 1 = 0,$$
$$\frac{\partial \pi_R}{\partial e_{Rr}} = \alpha_{Rr} \gamma_r \overline{N_r} \frac{e_{Er}(e_{Er} - e_{Rr})}{(e_{Er} + e_{Rr})^3} - 1 = 0.$$

It is straightforward to obtain the optimal ad budget levels (as listed in Lemma 2.9).

$$
\begin{aligned}
e_{Ep}^{NR} &= \frac{\alpha_E \alpha_{Rp} \left(2\alpha_E + \alpha_{Rp}\right)^2}{4\left(\alpha_E + \alpha_{Rp}\right)^3} \gamma_p \overline{N_p}, \\
e_{Er}^{NR} &= \frac{\alpha_E \alpha_{Rr} \left(2\alpha_E + \alpha_{Rr}\right)^2}{4\left(\alpha_E + \alpha_{Rr}\right)^3} \gamma_r \overline{N_r}, \\
e_{Rp}^{NR} &= \frac{\alpha_E \alpha_{Rp}^2 \left(2\alpha_E + \alpha_{Rp}\right)}{4\left(\alpha_E + \alpha_{Rp}\right)^3} \gamma_p \overline{N_p}, \\
e_{Rr}^{NR} &= \frac{\alpha_E \alpha_{Rr}^2 \left(2\alpha_E + \alpha_{Rr}\right)}{4\left(\alpha_E + \alpha_{Rr}\right)^3} \gamma_r \overline{N_r}.
\end{aligned}
$$

Next, we check the second order conditions:

$$
H_E = \begin{pmatrix} \frac{2\alpha_E \gamma_p \overline{N_p} e_{Ep}(e_{Rp} - 2e_{Ep})}{(e_{Ep}+e_{Rp})^4} & 0 \\ 0 & \frac{2\alpha_E \gamma_r \overline{N_r} e_{Er}(e_{Rr} - 2e_{Er})}{(e_{Er}+e_{Rr})^4} \end{pmatrix},
$$

$$
H_R = \begin{pmatrix} \frac{2\alpha_{Rp} \gamma_p \overline{N_p} e_{Rp}(e_{Rp} - 2e_{Ep})}{(e_{Ep}+e_{Rp})^4} & 0 \\ 0 & \frac{2\alpha_{Rr} \gamma_r \overline{N_r} e_{Rr}(e_{Rr} - 2e_{Er})}{(e_{Er}+e_{Rr})^4} \end{pmatrix}.
$$

At the optimal values, we can see that the Hessians are negatively definite.

Equal-exposure policy

After adjusting for the platform's intervention (Δe) under equal-exposure policy, we have the profit maximization when $e_{Rp} \geqslant e_{Rr}$ as:

$$
\begin{aligned}
\max_{e_{Ep}, e_{Er}} \pi_E &= \max_{e_{Ep}, e_{Er}} \alpha_E \left(\gamma_p \overline{N_p} + \gamma_r \overline{N_r}\right) \frac{e_{Er}^2}{\left(e_{Er} + e_{Rr}\right)^2} - (e_{Ep} + e_{Er}), \\
\max_{e_{Rp}, e_{Rr}} \pi_R &= \max_{e_{Rp}, e_{Rr}} \left(\alpha_{Rp} \gamma_p \overline{N_p} + \alpha_{Rr} \gamma_r \overline{N_r}\right) \frac{e_{Er} e_{Rr}}{\left(e_{Er} + e_{Rr}\right)^2} - (e_{Rp} + e_{Rr}).
\end{aligned}
$$

Following the same process of solving the base model equal-exposure, we know that both advertisers have no incentive to spend money on the protected users, and the FOCs of e_E & e_{Rr} are:

$$
\begin{aligned}
\frac{\partial \pi_E}{\partial e_{Er}} &= \frac{2 e_{Er} e_{Rr} \alpha_E \left(\gamma_p \overline{N_p} + \gamma_r \overline{N_r}\right)}{\left(e_{Er} + e_{Rr}\right)^3} - 1 = 0, \\
\frac{\partial \pi_R}{\partial e_{Rr}} &= \frac{e_{Er}(e_{Er} - e_{Rr}) \left(\alpha_{Rp} \gamma_p \overline{N_p} + \alpha_{Rr} \gamma_r \overline{N_r}\right)}{\left(e_{Er} + e_{Rr}\right)^2} - 1 = 0.
\end{aligned}
$$

Solving for these equations, we have the advertisers' decisions to be:

$$e_{Er}^{EE} = \frac{\alpha_E\left(\gamma_p\overline{N_p}+\gamma_r\overline{N_r}\right)\left(\alpha_{Rp}\gamma_p\overline{N_p}+\alpha_{Rr}\gamma_r\overline{N_r}\right)\left[\left(2\alpha_E+\alpha_{Rp}\right)\gamma_p\overline{N_p}+\left(2\alpha_E+\alpha_{Rr}\right)\gamma_r\overline{N_r}\right]^2}{4\left[\left(\alpha_E+\alpha_{Rp}\right)\gamma_p\overline{N_p}+\left(\alpha_E+\alpha_{Rr}\right)\gamma_r\overline{N_r}\right]^3},$$
$$e_{Rr}^{EE} = \frac{\alpha_E\left(\gamma_p\overline{N_p}+\gamma_r\overline{N_r}\right)\left(\alpha_{Rp}\gamma_p\overline{N_p}+\alpha_{Rr}\gamma_r\overline{N_r}\right)^2\left[\left(2\alpha_E+\alpha_{Rp}\right)\gamma_p\overline{N_p}+\left(2\alpha_E+\alpha_{Rr}\right)\gamma_r\overline{N_r}\right]}{4\left[\left(\alpha_E+\alpha_{Rp}\right)\gamma_p\overline{N_p}+\left(\alpha_E+\alpha_{Rr}\right)\gamma_r\overline{N_r}\right]^3}.$$

Next, we check the second order conditions for maximum points:

$$\frac{\partial^2\pi_E}{\partial e_{Er}^2} = -\frac{2e_{Rr}(2e_{Er}-e_{Rr})\alpha_E\left(\gamma_p\overline{N_p}+\gamma_r\overline{N_r}\right)}{\left(e_{Er}+e_{Rr}\right)^4} < 0,$$
$$\frac{\partial^2\pi_R}{\partial e_{Rr}^2} = -\frac{2e_E(2e_{Er}-e_{Rr})\left(\alpha_{Rp}\gamma_p\overline{N_p}+\alpha_{Rr}\gamma_r\overline{N_r}\right)}{\left(e_{Er}+e_{Rr}\right)^4}$$

It is straightforward to see that second order conditions are satisfied for the optimal decisions of e_{Er}^{EE} and e_{Rr}^{EE} obtained before. With the advertisers' decisions, we can easily obtain the platform's profit (as summarized in Lemma 2.10).

To check if our main results hold, we simplify the gap between π_P^{NR} and π_P^{EE} and show that $\pi_P^{EE} > \pi_P^{NR}$:

$$\pi_P^{EE}-\pi_P^{NR} = \frac{\begin{array}{r}\left[(\alpha_E+\alpha_{Rp})(3\alpha_E+\alpha_{Rp}+2\alpha_{Rr})\gamma_p\overline{N_p}+(\alpha_E+\alpha_{Rr})(3\alpha_E+2\alpha_{Rp}+\alpha_{Rr})\gamma_r\overline{N_r}\right]\\ \times(\alpha_{Rp}-\alpha_{Rr})^2\alpha_E^2\gamma_p\gamma_r\overline{N_p}\overline{N_r}\end{array}}{(\alpha_E+\alpha_{Rp})^2(\alpha_E+\alpha_{Rr})^2\left[(\alpha_E+\alpha_{Rp})\gamma_p\overline{N_p}+(\alpha_E+\alpha_{Rr})\gamma_R\overline{N_r}\right]} > 0$$

■

A.18 Proof of Lemma 2.11

No-restriction policy

Take FOCs of Equation (13) & (14):

$$\frac{\partial\pi_E}{\partial e_{Ep}} = \frac{\alpha_E N_p}{2}\left[\frac{1}{e_{Rp}}\ln\left(\frac{e_{Ep}+e_{Rp}}{e_{Ep}}\right)+\frac{e_{Rp}}{e_{Ep}^2}\ln\left(\frac{e_{Ep}+e_{Rp}}{e_{Rp}}\right)-\frac{1}{e_{Ep}}\right]-1=0,$$
$$\frac{\partial\pi_E}{\partial e_{Er}} = \frac{\alpha_E N_r}{2}\left[\frac{1}{e_{Rr}}\ln\left(\frac{e_{Er}+e_{Rr}}{e_{Er}}\right)+\frac{e_{Rr}}{e_{Er}^2}\ln\left(\frac{e_{Er}+e_{Rr}}{e_{Rr}}\right)-\frac{1}{e_{Er}}\right]-1=0,$$
$$\frac{\partial\pi_R}{\partial e_{Rp}} = \frac{\alpha_{Rp} N_p}{2}\left[\frac{1}{e_{Ep}}\ln\left(\frac{e_{Ep}+e_{Rp}}{e_{Rp}}\right)+\frac{e_{Ep}}{e_{Rp}^2}\ln\left(\frac{e_{Ep}+e_{Rp}}{e_{Ep}}\right)-\frac{1}{e_{Rp}}\right]-1=0,$$
$$\frac{\partial\pi_R}{\partial e_{Rr}} = \frac{\alpha_{Rr} N_r}{2}\left[\frac{1}{e_{Er}}\ln\left(\frac{e_{Er}+e_{Rr}}{e_{Rr}}\right)+\frac{e_{Er}}{e_{Rr}^2}\ln\left(\frac{e_{Er}+e_{Rr}}{e_{Er}}\right)-\frac{1}{e_{Rr}}\right]-1=0.$$

By dividing $\frac{\partial \pi_E}{\partial e_{Ep}} = 0$ with $\frac{\partial \pi_R}{\partial e_{Rp}} = 0$, we obtain the relationship between advertisers ad budget: $\frac{e_{Ep}^{NR}}{e_{Rp}^{NR}} = \frac{\alpha_E}{\alpha_{Rp}}$, $\frac{e_{Er}^{NR}}{e_{Rr}^{NR}} = \frac{\alpha_E}{\alpha_{Rr}}$. Advertisers' optimal decisions are:

$$
\begin{aligned}
e_{Ep}^{NR} &= \frac{\alpha_E N_p}{2}\left[\frac{\alpha_E}{\alpha_{Rp}}\ln\left(\frac{\alpha_E+\alpha_{Rp}}{\alpha_E}\right)+\frac{\alpha_{Rp}}{\alpha_E}\ln\left(\frac{\alpha_E+\alpha_{Rp}}{\alpha_{Rp}}\right)-1\right],\\
e_{Er}^{NR} &= \frac{\alpha_E N_r}{2}\left[\frac{\alpha_E}{\alpha_{Rr}}\ln\left(\frac{\alpha_E+\alpha_{Rr}}{\alpha_E}\right)+\frac{\alpha_{Rr}}{\alpha_E}\ln\left(\frac{\alpha_E+\alpha_{Rr}}{\alpha_{Rr}}\right)-1\right],\\
e_{Rp}^{NR} &= \frac{\alpha_{Rp} N_p}{2}\left[\frac{\alpha_E}{\alpha_{Rp}}\ln\left(\frac{\alpha_E+\alpha_{Rp}}{\alpha_E}\right)+\frac{\alpha_{Rp}}{\alpha_E}\ln\left(\frac{\alpha_E+\alpha_{Rp}}{\alpha_{Rp}}\right)-1\right],\\
e_{Rr}^{NR} &= \frac{\alpha_{Rr} N_r}{2}\left[\frac{\alpha_E}{\alpha_{Rr}}\ln\left(\frac{\alpha_E+\alpha_{Rr}}{\alpha_E}\right)+\frac{\alpha_{Rr}}{\alpha_E}\ln\left(\frac{\alpha_E+\alpha_{Rr}}{\alpha_{Rr}}\right)-1\right].
\end{aligned}
$$

The platform's profit is:

$$
\begin{aligned}
\pi_P^{NR} =& e_{Ep}^{NR}+e_{Er}^{NR}+e_{Rp}^{NR}+e_{Rr}^{NR}\\
=& \frac{(\alpha_E+\alpha_{Rp})N_p}{2}\left[\frac{\alpha_E}{\alpha_{Rp}}\ln\left(\frac{\alpha_E+\alpha_{Rp}}{\alpha_E}\right)+\frac{\alpha_{Rp}}{\alpha_E}\ln\left(\frac{\alpha_E+\alpha_{Rp}}{\alpha_{Rp}}\right)-1\right]\\
&+\frac{(\alpha_E+\alpha_{Rr})N_r}{2}\left[\frac{\alpha_E}{\alpha_{Rr}}\ln\left(\frac{\alpha_E+\alpha_{Rr}}{\alpha_E}\right)+\frac{\alpha_{Rr}}{\alpha_E}\ln\left(\frac{\alpha_E+\alpha_{Rr}}{\alpha_{Rr}}\right)-1\right].
\end{aligned}
$$

Next, we check the second order conditions at the equilibrium values:

$$
H_E = \begin{pmatrix}
\frac{\alpha_{Rp}N_p}{e_{Ep}^2}\left[\frac{\alpha_E}{\alpha_E+\alpha_{Rp}}-\ln\left(\frac{\alpha_E+\alpha_{Rp}}{\alpha_{Rp}}\right)\right] & 0\\
0 & \frac{\alpha_{Rr}N_r}{e_{Er}^2}\left[\frac{\alpha_E}{\alpha_E+\alpha_{Rr}}-\ln\left(\frac{\alpha_E+\alpha_{Rr}}{\alpha_{Rr}}\right)\right]
\end{pmatrix},
$$

$$
H_R = \begin{pmatrix}
\frac{\alpha_E N_p}{e_{Rp}^2}\left[\frac{\alpha_{Rp}}{\alpha_E+\alpha_{Rp}}-\ln\left(\frac{\alpha_E+\alpha_{Rp}}{\alpha_E}\right)\right] & 0\\
0 & \frac{\alpha_E N_r}{e_{Rr}^2}\left[\frac{\alpha_{Rr}}{\alpha_E+\alpha_{Rr}}-\ln\left(\frac{\alpha_E+\alpha_{Rr}}{\alpha_E}\right)\right]
\end{pmatrix}.
$$

The factor that decides to the definiteness of Hessians can be symbolized with the function $h(x) = \frac{x}{1+x} - \ln(1+x)$, with x being the valuation ratio between advertisers. It is not difficult to see that $h(x) < 0 \ \forall \ x > 0$. Therefore, the Hessians are negatively definite, and the FOC solution is the maximum point.

Equal-exposure policy

Using f_{Ep} as an example, we first show that the ad share f is monotonically increasing in the ad budget ratio:

$$
\begin{aligned}
&With\ f_{Ep} = \frac{1}{2}\left[1 + \frac{e_{Ep}}{e_{Rp}}\ln\left(1 + \frac{e_{Rp}}{e_{Ep}}\right) - \frac{e_{Rp}}{e_{Ep}}\ln\left(1 + \frac{e_{Ep}}{e_{Rp}}\right)\right],\\
&denote\ g(x) = x\ln\left(1 + \frac{1}{x}\right) - \frac{1}{x}\ln(1+x)\\
\Rightarrow &g'(x) = \ln\left(1 + \frac{1}{x}\right) + \frac{1}{x^2}\ln(1+x) - \frac{1}{x}\\
&g''(x) = \frac{2\left[x - (1+x)\ln(1+x)\right]}{x^3(1+x)}\\
&\because g'(x) \to 0\ when\ x \to \infty\ \&\ g''(x) < 0\\
&\therefore g'(x) > 0\ for\ x > 0
\end{aligned}
$$

Therefore, for the platform to ensure equal-exposure of ad E (that is, $f_{Ep} = f_{Er}$), the adjustment Δe satisfies $\frac{e_{Ep}+\Delta e}{e_{Ep}+\Delta e+e_{Rp}} = \frac{e_{Er}}{e_{Er}+e_{Rr}}$. We adjust advertisers' profit functions for the platform's intervention (Δe) under equal-exposure policy when $f_{Ep} \leqslant f_{Er}$ to be:

$$
\begin{aligned}
\pi_E =& \frac{\alpha_E(N_p + N_r)}{2}\left[1 + \frac{e_{Rr}}{e_{Er}}\ln\left(\frac{e_{Rr}}{e_{Er}+e_{Rr}}\right) + \frac{e_{Er}}{e_{Rr}}\ln\left(\frac{e_{Er}+e_{Rr}}{e_{Er}}\right)\right] - (e_{Ep} + e_{Er}),\\
\pi_R =& \frac{(\alpha_{Rp}N_p + \alpha_{Rr}N_r}{2}\left[1 + \frac{e_{Er}}{e_{Rr}}\ln\left(\frac{e_{Er}}{e_{Er}+e_{Rr}}\right) + \frac{e_{Rr}}{e_{Er}}\ln\left(\frac{e_{Er}+e_{Rr}}{e_{Rr}}\right)\right] - (e_{Rp} + e_{Rr}).
\end{aligned}
$$

Following the same process of solving the profit maximization with $e_{Rp} \geqslant e_{Rr}$, the FOCs are:

$$
\begin{aligned}
\frac{\partial \pi_E}{\partial e_{Er}} &= \frac{\alpha_E(N_p + N_r)}{2}\left[\frac{1}{e_{Rr}}\ln\left(\frac{e_{Er}+e_{Rr}}{e_{Er}}\right) + \frac{e_{Rr}}{e_{Er}^2}\ln\left(\frac{e_{Er}+e_{Rr}}{e_{Rr}}\right) - \frac{1}{e_{Er}}\right] - 1 = 0,\\
\frac{\partial \pi_R}{\partial e_{Rr}} &= \frac{(\alpha_{Rp}N_p + \alpha_{Rr}N_r)}{2}\left[\frac{1}{e_{Er}}\ln\left(\frac{e_{Er}+e_{Rr}}{e_{Rr}}\right) + \frac{e_{Er}}{e_{Rr}^2}\ln\left(\frac{e_{Er}+e_{Rr}}{e_{Er}}\right) - \frac{1}{e_{Rr}}\right] - 1 = 0.
\end{aligned}
$$

Same as the proof for the NR policy, we obtain the relationship between advertisers' ad budget to be $\frac{e_{Er}^{EE}}{e_{Rr}^{EE}} = \frac{\alpha_E(N_p+N_r)}{\alpha_{Rp}N_p+\alpha_{Rr}N_r}$.

The equilibrium values are:

$$
\begin{aligned}
e_{Er}^{EE} &= \frac{\phi_E}{2}\left[\frac{\phi_E}{\phi_R}\ln\left(1 + \frac{\phi_R}{\phi_E}\right) + \frac{\phi_R}{\phi_E}\ln\left(1 + \frac{\phi_E}{\phi_R}\right) - 1\right],\\
e_{Rr}^{EE} &= \frac{\phi_R}{2}\left[\frac{\phi_E}{\phi_R}\ln\left(1 + \frac{\phi_R}{\phi_E}\right) + \frac{\phi_R}{\phi_E}\ln\left(1 + \frac{\phi_E}{\phi_R}\right) - 1\right].
\end{aligned}
$$

The platform's profit is:

$$\begin{aligned}\pi_{P}^{EE} =& e_{Er}^{EE} + e_{Rr}^{EE} \\ =& \frac{\phi_E + \phi_R}{2}\left[\frac{\phi_E}{\phi_R}\ln\left(1+\frac{\phi_R}{\phi_E}\right) + \frac{\phi_R}{\phi_E}\ln\left(1+\frac{\phi_E}{\phi_R}\right) - 1\right].\end{aligned}$$

Last, we check the value of second order conditions at the optimal values to make sure the solutions are indeed maximum points:

$$\begin{aligned}\frac{\partial^2 \pi_E}{\partial e_{Er}^2} &= \frac{e_{Rr}}{e_{Er}^3}\phi_E\left[\frac{e_{Er}}{e_{Er}+e_{Rr}} + \ln\left(\frac{e_{Rr}}{e_{Er}+e_{Rr}}\right)\right], \\ \frac{\partial^2 \pi_R}{\partial e_{Rr}^2} &= \frac{e_{Er}}{e_{Rr}^3}\phi_R\left[\frac{e_{Rr}}{e_{Er}+e_{Rr}} + \ln\left(\frac{e_{Er}}{e_{Er}+e_{Rr}}\right)\right].\end{aligned}$$

At equilibrium values (e_{Er}^{EE} and e_{Rr}^{EE}), the second order condition becomes:

$$\begin{aligned}&\frac{\phi_E}{\phi_E+\phi_R} - \ln\left(\frac{\phi_E+\phi_R}{\phi_R}\right), \\ &\frac{\phi_R}{\phi_E+\phi_R} - \ln\left(\frac{\phi_E+\phi_R}{\phi_E}\right).\end{aligned}$$

Following the same logic as the proof for the NR policy, we use function $h(x) = \frac{x}{1+x} - \ln(1+x)$ to represent the above two expressions, with x being the ratio of advertisers' maximum profit (i.e., $x \in \{\frac{\phi_E}{\phi_R}, \frac{\phi_R}{\phi_E}\}$). Since $h(x) < 0 \ \forall \ x > 0$, the solution we obtain is a maximum point. ■

A.19 Proof of Lemma 2.12

With heterogeneous user quality Q, advertisers profit functions can be updated (with notations $A_p = \frac{e_{Ep}Q_{Ep}}{e_{Rp}Q_{Rp}}$ and $A_r = \frac{e_{Er}Q_{Er}}{e_{Rr}Q_{Rr}}$):

$$\begin{aligned}
\pi_{E} =& \frac{\alpha_{E}N_p}{2}\left[1 + A_p \ln\left(1+\frac{1}{A_p}\right) - \frac{1}{A_p}\ln\left(1+A_p\right)\right] \\
&+ \frac{\alpha_{E}N_r}{2}\left[1 + A_r \ln\left(1+\frac{1}{A_r}\right) - \frac{1}{A_r}\ln\left(1+A_r\right)\right] - (e_{Ep}+e_{Er}), \\
\pi_{R} =& \frac{\alpha_{Rp}N_p}{2}\left[1 + \frac{1}{A_p}\ln\left(1+A_p\right) - A_p \ln\left(1+\frac{1}{A_p}\right)\right] \\
&+ \frac{\alpha_{Rr}N_r}{2}\left[1 + \frac{1}{A_r}\ln\left(1+A_r\right) - A_r \ln\left(1+\frac{1}{A_r}\right)\right] - (e_{Rp}+e_{Rr}).
\end{aligned}$$

No-restriction policy

Take FOCs of advertisers' profit functions with heterogeneous user quality:

$$\begin{aligned}
\frac{\partial \pi_{E}}{\partial e_{Ep}} &= \frac{\alpha_{E}N_p}{2e_{Ep}}\left[A_p \ln\left(1+\frac{1}{A_p}\right) + \frac{1}{A_p}\ln\left(1+A_p\right) - 1\right] - 1 = 0, \\
\frac{\partial \pi_{E}}{\partial e_{Er}} &= \frac{\alpha_{E}N_r}{2e_{Er}}\left[A_r \ln\left(1+\frac{1}{A_r}\right) + \frac{1}{A_r}\ln\left(1+A_r\right) - 1\right] - 1 = 0, \\
\frac{\partial \pi_{R}}{\partial e_{Rp}} &= \frac{\alpha_{Rp}N_p}{2e_{Ep}}\left[A_p \ln\left(1+\frac{1}{A_p}\right) + \frac{1}{A_p}\ln\left(1+A_p\right) - 1\right] - 1 = 0, \\
\frac{\partial \pi_{R}}{\partial e_{Rr}} &= \frac{\alpha_{Rr}N_r}{2e_{Rr}}\left[A_r \ln\left(1+\frac{1}{A_r}\right) + \frac{1}{A_r}\ln\left(1+A_r\right) - 1\right] - 1 = 0.
\end{aligned}$$

We can derive the first part of the result in Lemma 2.12 easily from the above conditions:

$$\begin{aligned}
e_{Ep}^{NR} &= \frac{\alpha_{E}N_p}{2}\left[\frac{\alpha_{E}Q_{Ep}}{\alpha_{Rp}Q_{Rp}}\ln\left(1+\frac{\alpha_{Rp}Q_{Rp}}{\alpha_{E}Q_{Ep}}\right) + \frac{\alpha_{Rp}Q_{Rp}}{\alpha_{E}Q_{Ep}}\ln\left(1+\frac{\alpha_{E}Q_{Ep}}{\alpha_{Rp}Q_{Rp}}\right) - 1\right], \\
e_{Er}^{NR} &= \frac{\alpha_{E}N_r}{2}\left[\frac{\alpha_{E}Q_{Er}}{\alpha_{Rr}Q_{Rr}}\ln\left(1+\frac{\alpha_{Rr}Q_{Rr}}{\alpha_{E}Q_{Er}}\right) + \frac{\alpha_{Rr}Q_{Rr}}{\alpha_{E}Q_{Er}}\ln\left(1+\frac{\alpha_{E}Q_{Er}}{\alpha_{Rr}Q_{Rr}}\right) - 1\right], \\
e_{Rp}^{NR} &= \frac{\alpha_{Rp}N_p}{2}\left[\frac{\alpha_{E}Q_{Ep}}{\alpha_{Rp}Q_{Rp}}\ln\left(1+\frac{\alpha_{Rp}Q_{Rp}}{\alpha_{E}Q_{Ep}}\right) + \frac{\alpha_{Rp}Q_{Rp}}{\alpha_{E}Q_{Ep}}\ln\left(1+\frac{\alpha_{E}Q_{Ep}}{\alpha_{Rp}Q_{Rp}}\right) - 1\right], \\
e_{Rr}^{NR} &= \frac{\alpha_{Rr}N_r}{2}\left[\frac{\alpha_{E}Q_{Er}}{\alpha_{Rr}Q_{Rr}}\ln\left(1+\frac{\alpha_{Rr}Q_{Rr}}{\alpha_{E}Q_{Er}}\right) + \frac{\alpha_{Rr}Q_{Rr}}{\alpha_{E}Q_{Er}}\ln\left(1+\frac{\alpha_{E}Q_{Er}}{\alpha_{Rr}Q_{Rr}}\right) - 1\right].
\end{aligned}$$

We obtain the relationships among advertisers' ad expenses as $\frac{e_{Ep}^{NR}}{e_{Rp}^{NR}} = \frac{\alpha_E}{\alpha_{Rp}}$ and $\frac{e_{Er}^{NR}}{e_{Rr}^{NR}} = \frac{\alpha_E}{\alpha_{Rr}}$. Next, we obtain the platform's profit under no-restriction:

$$
\begin{aligned}
\pi_{P}^{NR} =& e_{Ep}^{NR} + e_{Er}^{NR} + e_{Rp}^{NR} + e_{Rr}^{NR} \\
=& \frac{1}{2}\Bigg[\left(\alpha_{E} + \alpha_{Rp}\right) N_p \left(\frac{Q_{Ep}\alpha_{E}}{Q_{Rp}\alpha_{Rp}} \ln\left(1 + \frac{Q_{Rp}\alpha_{Rp}}{Q_{Ep}\alpha_{E}}\right) + \frac{Q_{Rp}\alpha_{Rp}}{Q_{Ep}\alpha_{E}} \ln\left(1 + \frac{Q_{Ep}\alpha_{E}}{Q_{Rp}\alpha_{Rp}}\right)\right) \\
& + \left(\alpha_{E} + \alpha_{Rr}\right) N_r \left(\frac{Q_{Er}\alpha_{E}}{Q_{Rr}\alpha_{Rr}} \ln\left(1 + \frac{Q_{Rp}\alpha_{Rp}}{Q_{Ep}\alpha_{E}}\right) + \frac{Q_{Rr}\alpha_{Rr}}{Q_{Er}\alpha_{E}} \ln\left(1 + \frac{Q_{Ep}\alpha_{E}}{Q_{Rp}\alpha_{Rp}}\right)\right) - \phi_{E} - \phi_{R} \Bigg].
\end{aligned}
$$

Last, we check the second order conditions at the equilibrium values:

$$
H_E = \begin{pmatrix} \frac{\alpha_E N_p}{A_p e_{Ep}^2}\left[\frac{A_p}{1+A_p} - \ln\left(1+A_p\right)\right] & 0 \\ 0 & \frac{\alpha_E N_r}{A_r e_{Er}^2}\left[\frac{A_r}{1+A_r} - \ln\left(1+A_r\right)\right] \end{pmatrix},
$$

$$
H_R = \begin{pmatrix} \frac{\alpha_{Rp} N_p}{A_p e_{Rp}^2}\left[\frac{1}{1+A_p} - \ln\left(1+\frac{1}{A_p}\right)\right] & 0 \\ 0 & \frac{\alpha_E N_r}{A_r e_{Er}^2}\left[\frac{1}{1+A_r} - \ln\left(1+\frac{1}{A_r}\right)\right] \end{pmatrix}.
$$

Using the same method as in the proof of second order conditions in Appendix A.18, we can see the Hessian conditions for a maximum point are met.

Equal-exposure policy

Following the same method of solving for equal-exposure as in the proof of Appendix A.18, we have the profit functions after adjusting for the platform's intervention Δe when $f_{Ep} \leqslant f_{Er}$ to be:

$$
\begin{aligned}
\pi_{E} =& \frac{\alpha_{E}(N_p + N_r)}{2}\left[1 + A_r \ln\left(1 + \frac{1}{A_r}\right) - \frac{1}{A_r}\ln\left(1 + A_r\right)\right] - \left(e_{Ep} + e_{Er}\right), \\
\pi_{R} =& \frac{\left(\alpha_{Rp} N_p + \alpha_{Rr} N_r\right.}{2}\left[1 + \frac{1}{A_r}\ln\left(1 + A_r\right) - A_r \ln\left(1 + \frac{1}{A_r}\right)\right] - \left(e_{Rp} + e_{Rr}\right).
\end{aligned}
$$

Using the same process of solving for the base model equal-exposure, we know that both advertisers have no incentive to spend money on the protected users, and the FOCs of e_{Er} & e_{Rr} are:

$$
\begin{aligned}
\frac{\partial \pi_{E}}{\partial e_{Er}} &= \frac{\alpha_{E}(N_p + N_r)}{2e_{E}}\left[A_r \ln\left(1 + \frac{1}{A_r}\right) + \frac{1}{A_r}\ln\left(1 + A_r\right) - 1\right] - 1 = 0, \\
\frac{\partial \pi_{R}}{\partial e_{Rr}} &= \frac{\left(\alpha_{Rp} N_p + \alpha_{Rr} N_r\right)}{2e_{Rr}}\left[A_r \ln\left(1 + \frac{1}{A_r}\right) + \frac{1}{A_r}\ln\left(1 + A_r\right) - 1\right] - 1 = 0.
\end{aligned}
$$

We can see easily see that the ratio between two advertisers' ad budget is $\frac{e_{Er}^{EE}}{e_{Rr}^{EE}} = \frac{\phi_E}{\phi_R}$. The equilibrium values are:

$$e_{Er}^{EE} = \frac{\phi_E}{2}\left[\frac{K\phi_E}{\phi_R} ln\left(1+\frac{\phi_R}{K\phi_E}\right) + \frac{\phi_R}{K\phi_E} ln\left(1+\frac{K\phi_E}{\phi_R}\right) - 1\right],$$
$$e_{Rr}^{EE} = \frac{\phi_R}{2}\left[\frac{K\phi_E}{\phi_R} ln\left(1+\frac{\phi_R}{K\phi_E}\right) + \frac{\phi_R}{K\phi_E} ln\left(1+\frac{K\phi_E}{\phi_R}\right) - 1\right].$$

The platform's profit is:

$$\begin{aligned}\pi_P^{EE} =& e_{Er}^{EE} + e_{Rr}^{EE} \\ =& \frac{\phi_E+\phi_R}{2}\left[\frac{K\phi_E}{\phi_R} ln\left(1+\frac{\phi_R}{K\phi_E}\right) + \frac{\phi_R}{K\phi_E} ln\left(1+\frac{K\phi_E}{\phi_R}\right) - 1\right].\end{aligned}$$

Last, we check the value of the Hessian matrix at the optimal values to make sure the solutions are maximum points.

$$\frac{\partial^2 \pi_E}{\partial e_{Er}^2} = \frac{\alpha_E (N_p + N_r) e_{Rr} Q_{Rr}}{e_{Er}^3 Q_{Er}}\left[\frac{A_r}{1+A_r} - \ln(1+A_r)\right],$$
$$\frac{\partial^2 \pi_R}{\partial e_{Rr}^2} = \frac{(\alpha_{Rp} N_p + \alpha_{Rr} N_r) e_{Er} Q_{Er}}{e_{Rr}^3 Q_{Rr}}\left[\frac{1}{1+A_r} - \ln\left(1+\frac{1}{A_r}\right)\right].$$

Following the same method as in the proof of second-order conditions in Appendix A.18, we can see that the solutions are optimal. ■

Appendix B Proofs for Chapter 3

B.1 Proof of Lemma 3.1

Employee's utility maximization under the general ability level (a_t, a_i) and salary rate γ is:

$$\max_{w_t,w_i} u = \max_{w_t,w_i} \gamma\left(\beta_t w_t + \beta_i w_i\right) - \left(\frac{w_t^2}{a_t} + \frac{w_i^2}{a_i}\right).$$

Take first-order conditions with respect to the labor decision $w_t \& w_i$:

$$\frac{\partial u}{\partial w_t} = -\frac{2w_t}{a_t} + \beta_t\gamma = 0$$
$$\frac{\partial u}{\partial w_i} = -\frac{2w_i}{a_i} + \beta_i\gamma = 0$$

We also check the second order condition:

$$\frac{\partial^2 u}{\partial w_t^2} = -\frac{2}{a_t},$$
$$\frac{\partial^2 u}{\partial w_i^2} = -\frac{2}{a_i}.$$

One can easily see that the Hessians are negative definite as the second-order conditions are negative.

The equilibrium employee decisions, the output level, and net utility are:

$$w_t^* = \frac{1}{2}\beta_t a_t\gamma,\ w_i^* = \frac{1}{2}\beta_i a_i\gamma$$
$$o^* = \beta_t w_t^* + \beta_i w_i^* = \frac{1}{2}\left(\beta_t^2 a_t + \beta_i^2 a_i\right)\gamma$$
$$u^* = \gamma o^* - \left[\frac{(w_t^*)^2}{a_t} + \frac{(w_i^*)^2}{a_i}\right] = \frac{1}{4}\left(\beta_t^2 a_t + \beta_i^2 a_i\right)\gamma^2$$

■

B.2 Proof of Lemma 3.2

Under the competition for the bonus salary, employees make the choices of labor by maximizing the final output under the bonus rate as long as the net utility is higher than that under the base salary. The optimization under the general ability level (a_t, a_i) is,

$$\begin{aligned} \max_{w_t, w_s} \quad & \beta_t w_t + \beta_i w_s \\ s.t. \quad & \gamma_{B} \left(\beta_t w_t + \beta_i w_s\right) - \left(\frac{w_t^2}{a_t} + \frac{w_s^2}{a_i}\right) \geqslant \frac{1}{4}\left(\beta_t^2 a_t + \beta_i^2 a_i\right) \underline{\gamma}^2 \end{aligned}$$

Construct the Lagrangian as follows:

$$\mathcal{L} = \beta_t w_t + \beta_i w_s + \lambda \left[\gamma_{B} \left(\beta_t w_t + \beta_i w_s\right) - \left(\frac{w_t^2}{a_t} + \frac{w_s^2}{a_i}\right) - \frac{1}{4}\left(\beta_t^2 a_t + \beta_i^2 a_i\right) \underline{\gamma}^2\right]$$

With the following FOCs:

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial w_t} &= \beta_t + \left(-\frac{2w_t}{a_t} + \beta_t \gamma_{B}\right) \lambda = 0 \\ \frac{\partial \mathcal{L}}{\partial w_s} &= \beta_i + \left(-\frac{2w_s}{a_i} + \beta_i \gamma_{B}\right) \lambda = 0 \\ \frac{\partial \mathcal{L}}{\partial \lambda} &= \left(\frac{w_t^2}{a_t} + \frac{w_s^2}{a_i}\right) + \frac{1}{4}\left(\beta_t^2 a_t + \beta_i^2 a_i\right) \underline{\gamma}^2 - \gamma_{B} \left(\beta_t w_t + \beta_i w_s\right) = 0 \end{aligned}$$

We can obtain two sets of possible optimal value of $(w_t^*, w_s^*, \lambda^*)$ to be

$$\left(\frac{\beta_t a_t \left(\gamma_{B} + \sqrt{\gamma_{B}^2 - \underline{\gamma}^2}\right)}{2}, \frac{\beta_i a_i \left(\gamma_{B} + \sqrt{\gamma_{B}^2 - \underline{\gamma}^2}\right)}{2}, \frac{1}{\sqrt{\gamma_{B}^2 - \underline{\gamma}^2}}\right) \textit{ and}$$

$$\left(\frac{\beta_t a_t \left(\gamma_{B} - \sqrt{\gamma_{B}^2 - \underline{\gamma}^2}\right)}{2}, \frac{\beta_i a_i \left(\gamma_{B} - \sqrt{\gamma_{B}^2 - \underline{\gamma}^2}\right)}{2}, -\frac{1}{\sqrt{\gamma_{B}^2 - \underline{\gamma}^2}}\right)$$

By checking the condition of bordered Hessian, we find that solution set 1 meets the criteria of maximum point. Therefore, the labor decisions, the highest possible output, and the utility of an employee are:

$$w_t^* = \frac{\gamma_{_B} + \sqrt{\gamma_{_B}^2 - \underline{\gamma}^2}}{2}\beta_t a_t,\ w_i^* = \frac{\gamma_{_B} + \sqrt{\gamma_{_B}^2 - \underline{\gamma}^2}}{2}\beta_i a_i,$$

$$o^* = \beta_t w_t^* + \beta_i w_s^* = \frac{\gamma_{_B} + \sqrt{\gamma_{_B}^2 - \underline{\gamma}^2}}{2}\left(\beta_t^2 a_t + \beta_i^2 a_i\right),$$

$$u^* = \gamma_{_B} o^* - \left[\frac{(w_t^*)^2}{a_t} + \frac{(w_i^*)^2}{a_i}\right] = \frac{1}{4}\left(\beta_t^2 a_t + \beta_i^2 a_i\right)\underline{\gamma}^2.$$

■

B.3 Proof of Proposition 3.1

From the breakdown of productivity change by employee type, we have:

$$\begin{aligned}
\Delta o^{HH} =& 0,\\
\Delta o^{HL} =& \underbrace{o_{\underline{\gamma}}^{HL} - o_{max}^{LH}}_{\textit{Competition Effect}},\\
\Delta o^{LH} =& o_{max}^{HL} - o_{\underline{\gamma}}^{LH},\\
=& \underbrace{\left(o_{max}^{HL} - o_{\underline{\gamma}}^{LH_{AI}}\right)}_{\textit{Competition Effect}} + \underbrace{\left(o_{\underline{\gamma}}^{LH_{AI}} - o_{\underline{\gamma}}^{LH}\right)}_{\textit{Learning Effect}},\\
\Delta o^{LL} =& \underbrace{o_{\underline{\gamma}}^{LL_{AI}} - o_{\underline{\gamma}}^{LL}}_{\textit{Learning Effect}}.
\end{aligned}$$

In the environment we are interested in, (H, L) naturally ranks over (L, H) before AI, ranking swaps after AI, and there is intensified competition between them. We have $o_{\underline{\gamma}}^{HL} < o_{\gamma_B}^{HL} < o_{max}^{LH}$ and $o_{\underline{\gamma}}^{LH_{AI}} < o_{\gamma_B}^{LH_{AI}} < o_{max}^{HL}$. Therefore, the *competition effect* is negative on (H, L) and positive on (L, H). Because *learning effect* is always positive, we have $\Delta o^{HL} < 0$ and $\Delta o^{LH} > 0$.

B.4 Proof of Theorem 3.1

We first compare the firm's productivity before and after AI ($O_{AI} - O_{noAI}$):

$$(o_{\underline{\gamma}}^{HL} - o_{max}^{LH})p_{HL} + (o_{max}^{HL} - o_{\underline{\gamma}}^{LH})p_{LH} + (o_{\underline{\gamma}}^{LL_{AI}} - o_{\underline{\gamma}}^{LL}),$$

or it can be simplified as

$$\Delta o^{HL} p_{HL} + \Delta o^{LH} p_{LH} + \Delta o^{LL} p_{LL}.$$

For this expression to be negative, we derive the condition in terms of $a_t^H - a_t^L$ and p_{HL} as $\hat{a}^{(o)} = 2\frac{(o_{max}^{HH} - o_{\underline{\gamma}}^{HL})p_{HL} - (o_{max}^{HL} - o_{\underline{\gamma}}^{HH})p_{LH}}{\beta_t^2(Kp_{HL} + \underline{\gamma}p_{LH} + f\underline{\gamma}p_{LL})}$ and $\hat{p}_{HL}^{(o)} = \frac{\Delta o^{LH} p_{LH} + \Delta o^{LL} p_{LL}}{\Delta o^{HL}}$.

Similarly, we derive the parameter condition for $\pi_{AI} - \pi_{noAI} < 0$ with respect to $a_t^H - a_t^L$ and p_{HL} as $\hat{a}^{(\pi)} = 2\frac{(M_2 o_{max}^{HH} - M_1 o_{\underline{\gamma}}^{HL})p_{HL} - (M_2 o_{max}^{HL} - M_1 o_{\underline{\gamma}}^{HH})p_{LH}}{\beta_t^2(M_2 K p_{HL} + M_1 \underline{\gamma} p_{LH} + M_1 f \underline{\gamma} p_{LL})}$ and $\hat{p}_{HL}^{(\pi)} = \frac{\Delta \pi^{LH} p_{LH} + \Delta \pi^{LL} p_{LL}}{\Delta \pi^{HL}}$. ■

B.5 Proof of Proposition 3.2

Employee's utility for those who rank at first, third, and last follows the natural utility identified before. That is, before AI:

$$u^{HH} = \frac{1}{4}\left(\beta_t^2 a_t^H + \beta_i^2 a_i^H\right)\gamma_B^2, u^{LH} = \frac{1}{4}\left(\beta_t^2 a_t^L + \beta_i^2 a_i^H\right)\underline{\gamma}^2, u^{LL} = \frac{1}{4}\left(\beta_t^2 a_t^L + \beta_i^2 a_i^L\right)\underline{\gamma}^2$$

After AI:

$$u^{HH} = \frac{1}{4}\left(\beta_t^2 a_t^H + \beta_i^2 a_i^H\right)\gamma_B^2, u^{HL} = \frac{1}{4}\left(\beta_t^2 a_t^H + \beta_i^2 a_i^L\right)\underline{\gamma}^2, u^{LL} = \frac{1}{4}\left(\beta_t^2 a_{te}^L + \beta_i^2 a_i^L\right)\underline{\gamma}^2$$

For the utility for the ones at rank 2, we need to solve their decision by minimizing the cost with an output constraint. For example, for (H, L) type's decision before AI:

$$\begin{aligned} \min_{w_t, w_i} \quad & \left(\frac{w_t^2}{a_t^H} + \frac{w_i^2}{a_i^L}\right) \\ s.t. \quad & \beta_t w_t + \beta_i w_i = \frac{\gamma_B + \sqrt{\gamma_B^2 - \underline{\gamma}^2}}{2}\left(\beta_t^2 a_t^L + \beta_i^2 a_i^H\right) \end{aligned}$$

Solving for the employees' decision, we derive for the utility of (H, L) before AI to be

$$\frac{(\gamma_B + \sqrt{\gamma_B^2 - \underline{\gamma}^2})(\beta_t^2 a_t^L + \beta_i^2 a_i^H)\gamma_B}{4}\left(2\gamma_B - \frac{(\gamma_B + \sqrt{\gamma_B^2 - \underline{\gamma}^2})(\beta_t^2 a_t^L + \beta_i^2 a_i^H)}{\beta_t^2 a_t^H + \beta_i^2 a_i^L}\right)$$

It is straightforward to see that this utility is higher than the after-AI value. ■

B.6 Proof of Proposition 3.3

Compare the profit under the guaranteed salary to the base value:

$$\pi^{g}_{AI} - \pi^{ng}_{AI} = M_2(o^{LH_{AI}}_{\gamma_B} - o^{HL}_{max})p_{LH} + (M_2 o^{HL}_{\gamma_B} - M_1 o^{HL}_{\underline{\gamma}})p_{HL}.$$

Because the first term $(o^{LH_{AI}}_{\gamma_B} - o^{HL}_{max}) < 0$, the overall value can only be positive when $(1-\gamma_B)\gamma_B > (1-\underline{\gamma})\underline{\gamma}$. Under this condition on the salary rate, we can easily derive the thresholds for $\pi^{g}_{AI} - \pi^{ng}_{AI} > 0$ in terms of tangible skill gap $a^H_t - a^L_t$ and p_{HL} as $\hat{a}^{(g)} = 2\frac{M_2(o^{HH}_{\gamma_B} - o^{HL}_{max})p_{LH} - (M_1 o^{HL}_{\underline{\gamma}} - M_2 o^{HL}_{\gamma_B})p_{HL}}{\beta_t^2 M_2 \gamma_B (1-f) p_{LH}}$ and $\hat{p}^{(g)}_{HL} = \frac{M_2(o^{HL}_{max} - o^{LH_{AI}}_{\gamma_B})p_{LH}}{M_2 o^{HL}_{\gamma_B} - M_1 o^{HL}_{\underline{\gamma}}}$ $\Delta a_t =\leqslant \hat{a}^{(g)}$, $p_{HL} \geqslant \hat{p}^{(g)}_{HL}$. ■

B.7 Proof of Proposition 3.4

From Equation (23), we have two candidates for the maximum value as

$$\pi_{AI}(\hat{f}^{(l)}) = M_2(o^{HH}_{\gamma_B} p_{HH} + o^{HL}_{max} p_{HL}) + M_1 \left[o^{HL}_{\underline{\gamma}} p_{LH} + \frac{\gamma}{2}\left(\beta_t^2 a^H_t - \beta_i^2 (a^H_i - 2a^L_i)\right) p_{LL} \right]$$

$$\pi_{AI}(1) = M_2 o^{HH}_{\gamma_B}(p_{HH} + p_{LH}) + M_1 o^{HL}_{\underline{\gamma}}(p_{HL} + p_{LL})$$

Compare these two values and derive the condition in terms of p_{HL}, we get the threshold $\hat{p}^{(f)}_{HL} = \frac{(M_2 o^{HH}_{\gamma_B} - M_1 o^{HL}_{\underline{\gamma}})p_{LH} + M_1 \frac{\gamma}{2}\beta_i^2 (a^H_i - a^L_i) p_{LL}}{M_2 o^{HL}_{max} - M_1 o^{HL}_{\underline{\gamma}}}$. ■

B.8 Proof of Proposition 3.5

When the firm chooses the base AI level at $\hat{f}^{(l)}_0$, the profit is more likely to be the maximum as shown in Figure 24. With $\hat{f}^{(l)}_0 == \frac{\hat{f}\bar{\gamma} - [\gamma_B p_i + K(1-p_i)]\theta}{\bar{\gamma} - [\gamma_B p_i + K(1-p_i)]\theta}$, we can easily see that this value is increasing in $\hat{f}$.

For $\hat{f} = 1 - \frac{\beta_i^2 (a^H_i - a^L_i)}{\beta_t^2 (a^H_t - a^L_t)}$, it increase when a^L_t decreases or a^L_i increases. ■

Bibliography

Abernethy, M. A., C.-Y. Hung, and L. van Lent. 2020, January. Expertise and Discretionary Bonus Decisions. *Management Science* 66 (1): 415–432.

Acemoglu, D., and P. Restrepo. 2019, May. Automation and New Tasks: How Technology Displaces and Reinstates Labor. *Journal of Economic Perspectives* 33 (2): 3–30.

Agrawal, A., J. S. Gans, and A. Goldfarb. 2019, May. Artificial Intelligence: The Ambiguous Labor Market Impact of Automating Prediction. *Journal of Economic Perspectives* 33 (2): 31–50.

Ali, M., P. Sapiezynski, M. Bogen, A. Korolova, A. Mislove, and A. Rieke. 2019, November. Discrimination Through Optimization: How Facebook's Ad Delivery Can Lead to Skewed Outcomes. In *Proceedings of the ACM on Human-Computer Interaction*, Volume 3, 1–30.

Allen, R., and P. R. Choudhury. 2022, January. Algorithm-Augmented Work and Domain Experience: The Countervailing Forces of Ability and Aversion. *Organization Science* 33 (1): 149–169.

Angwin, J., J. Larson, S. Mattu, L. Kirchner, and ProPublica. 2016, May. Machine Bias. https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing.

Arinez, J. F., Q. Chang, R. X. Gao, C. Xu, and J. Zhang. 2020, August. Artificial Intelligence in Advanced Manufacturing: Current Status and Future Outlook. *Journal of Manufacturing Science and Engineering* 142 (11).

Aseri, M., M. Dawande, G. Janakiraman, and V. Mookerjee. 2018, October. Procurement Policies for Mobile-Promotion Platforms. *Management Science* 64 (10): 4590–4607.

Aseri, M., A. Mehra, V. Mookerjee, and H. Xu. 2021, March. Should an Ad-Agency Offer Geoconquesting or Protection From It? *HKUST Business School Research Paper No. 2021-018 (also appears in Targeted and Social Network Marketing eJournal)*.

Athey, S. C., K. A. Bryan, and J. S. Gans. 2020, May. The Allocation of Decision Authority to Human and Artificial Intelligence. *AEA Papers and Proceedings* 110:80–84.

Autor, D. H. 2015, August. Why Are There Still So Many Jobs? The History and Future of Workplace Automation. *Journal of Economic Perspectives* 29 (3): 3–30.

Bagwell, K. 2007. The Economic Analysis of Advertising. Handbook of Industrial Organization, Elsevier.

Balseiro, S. R., O. Besbes, and G. Y. Weintraub. 2015, January. Repeated Auctions with Budgets in Ad Exchanges: Approximations and Design. *Management Science* 61 (4): 864–884.

Bansal, G., B. Nushi, E. Kamar, W. S. Lasecki, D. S. Weld, and E. Horvitz. 2019, October. Beyond Accuracy: The Role of Mental Models in Human-AI Team Performance. *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing* 7:2–11.

Barocas, S., and A. D. Selbst. 2016. Big Data's Disparate Impact. *California Law Review* 104 (3): 671–732.

Bartlett, R., A. Morse, R. Stanton, and N. Wallace. 2022, January. Consumer-Lending Discrimination in the FinTech Era. *Journal of Financial Economics* 143 (1): 30–56.

Bergemann, D., and A. Bonatti. 2019. Markets for Information: An Introduction. *Annual Review of Economics* 11 (1): 85–107.

Brynjolfsson, E., and T. Mitchell. 2017, December. What Can Machine Learning Do? Workforce Implications. *Science* 358 (6370): 1530–1534.

Brynjolfsson, E., T. Mitchell, and D. Rock. 2018, May. What Can Machines Learn and What Does It Mean for Occupations and the Economy? *AEA Papers and Proceedings* 108:43–47.

Brynjolfsson, E., D. Rock, and C. Syverson. 2018, January. Artificial Intelligence and the Modern Productivity Paradox: A Clash of Expectations and Statistics. In *The Economics of Artificial Intelligence: An Agenda*, 23–57. University of Chicago Press.

Burström, T., V. Parida, T. Lahti, and J. Wincent. 2021, April. AI-Enabled Business-Model Innovation and Transformation in Industrial Ecosystems: A Framework, Model and Outline for Further Research. *Journal of Business Research* 127:85–95.

Cadsby, C. B., F. Song, and F. Tapon. 2007, April. Sorting and Incentive Effects of Pay for Performance: An Experimental Investigation. *Academy of Management Journal* 50 (2): 387–405.

Calvano, E., G. Calzolari, V. Denicolò, and S. Pastorello. 2020, October. Artificial Intelligence, Algorithmic Pricing, and Collusion. *American Economic Review* 110 (10): 3267–3297.

Castelo, N., M. W. Bos, and D. R. Lehmann. 2019, October. Task-Dependent Algorithm Aversion. *Journal of Marketing Research* 56 (5): 809–825.

Celis, L. E., A. Mehrotra, and N. K. Vishnoi. 2019, May. Toward Controlling Discrimination in Online Ad Auctions.

Chan, T. Y., J. Li, and L. Pierce. 2014, August. Compensation and Peer Effects in Competing Sales Teams. *Management Science* 60 (8): 1965–1984.

Chawla, S., and M. Jagadeesan. 2020, July. Fairness in Ad Auctions Through Inverse Proportionality. *arXiv:2003.13966 [cs]*.

Chouldechova, A., and A. Roth. 2018, October. The Frontiers of Fairness in Machine Learning. Technical report.

Collins, C., D. Dennehy, K. Conboy, and P. Mikalef. 2021, October. Artificial Intelligence in Information Systems Research: A Systematic Literature Review and Research Agenda. *International Journal of Information Management* 60:102383.

Cowgill, B., and C. E. Tucker. 2020, February. Algorithmic Fairness and Economics.

Cui, R., J. Li, and D. J. Zhang. 2019, August. Reducing Discrimination with Reviews in the Sharing Economy: Evidence from Field Experiments on Airbnb. *Management Science* 66 (3): 1071–1094.

Cui, R., M. Li, and S. Zhang. 2022, March. AI and Procurement. *Manufacturing & Service Operations Management* 24 (2): 691–706.

Datta, A., M. C. Tschantz, and A. Datta. 2015, April. Automated Experiments on Ad Privacy Settings. *Proceedings on Privacy Enhancing Technologies* 2015 (1): 92–112.

Dietvorst, B. J., J. P. Simmons, and C. Massey. 2015, February. Algorithm Aversion: People Erroneously Avoid Algorithms After Seeing Them Err. *Journal of Experimental Psychology: General* 144 (1): 114–126.

(US)

Dietvorst, B. J., J. P. Simmons, and C. Massey. 2018, March. Overcoming Algorithm Aversion: People Will Use Imperfect Algorithms If They Can (Even Slightly) Modify Them. *Management Science* 64 (3): 1155–1170.

Dwork, C., and C. Ilvento. 2018. Fairness under Composition. 20 pages.

Edelman, B., M. Luca, and D. Svirsky. 2017, April. Racial Discrimination in the Sharing Economy: Evidence from a Field Experiment. *American Economic Journal: Applied Economics* 9 (2): 1–22.

Erickson, G. M. 1985. A Model of Advertising Competition. *Journal of Marketing Research* 22 (3): 297–304.

Facebook 2019, August. Updates To Housing, Employment and Credit Ads in Ads Manager. https://www.facebook.com/business/news/updates-to-housing-employment-and-credit-ads-in-ads-manager.

Felten, E., M. Raj, and R. Seamans. 2021, December. Occupational, Industry, and Geographic Exposure to Artificial Intelligence: A Novel Dataset and Its Potential Uses. *Strategic Management Journal* 42 (12): 2195–2217.

Felten, E. W., M. Raj, and R. Seamans. 2018, May. A Method to Link Advances in Artificial Intelligence to Occupational Abilities. *AEA Papers and Proceedings* 108:54–57.

Fjelland, R. 2020, June. Why General Artificial Intelligence Will Not Be Realized. *Humanities and Social Sciences Communications* 7 (1): 1–9.

Frank, M. R., D. Autor, J. E. Bessen, E. Brynjolfsson, M. Cebrian, D. J. Deming, M. Feldman, M. Groh, J. Lobo, E. Moro, D. Wang, H. Youn, and I. Rahwan. 2019, April. Toward Understanding the Impact of Artificial Intelligence on Labor. *Proceedings of the National Academy of Sciences* 116 (14): 6531–6539.

Fu, R., M. Aseri, P. Singh, and K. Srinivasan. 2021, October. "Un"Fair Machine Learning Algorithms. *Management Science*.

Fügener, A., J. Grahl, A. Gupta, and W. Ketter. 2021, September. Will Humans-in-the-Loop Become Borgs? Merits and Pitfalls of Working with AI. *MIS Quarterly* 45 (3): 1527–1556.

Fügener, A., J. Grahl, A. Gupta, and W. Ketter. 2022, June. Cognitive Challenges in Human–Artificial Intelligence Collaboration: Investigating the Path Toward Productive Delegation. *Information Systems Research* 33 (2): 678–696.

Furman, J., and R. Seamans. 2019. AI and the Economy. *Innovation Policy and the Economy* 19:161–191.

Gallego, A., and T. Kurer. 2022, May. Automation, Digitalization, and Artificial Intelligence in the Workplace: Implications for Political Behavior. *Annual Review of Political Science* 25 (1): 463–484.

Garnelo, M., and M. Shanahan. 2019, October. Reconciling Deep Learning with Symbolic Artificial Intelligence: Representing Objects and Relations. *Current Opinion in Behavioral Sciences* 29:17–23.

Gelauff, L., A. Goel, K. Munagala, and S. Yandamuri. 2020, June. Advertising for Demographically Fair Outcomes.

Gerhart, B., and M. Fang. 2014, March. Pay for (Individual) Performance: Issues, Claims, Evidence and the Role of Sorting Effects. *Human Resource Management Review* 24 (1): 41–52.

Google 2020, June. Upcoming Update to Housing, Employment, and Credit Advertising Policies.

Groshen, and Holzer. 2019. Improving Employment and Earnings in Twenty-First Century Labor Markets: An Introduction. *RSF: The Russell Sage Foundation Journal of the Social Sciences* 5 (5): 1.

Groysberg, B., S. Abbott, M. R. Marino, and M. Aksoy. 2021, January. Compensation Packages That Actually Drive Performance. *Harvard Business Review*.

He, W., S.-L. Li, J. Feng, G. Zhang, and M. C. Sturman. 2021, February. When Does Pay for Performance Motivate Employee Helping Behavior? The Contextual Influence of Performance Subjectivity. *Academy of Management Journal* 64 (1): 293–326.

Heaven, W. D. 2020, October. Artificial General Intelligence: Are We Close, and Does It Even Make Sense to Try? https://tinyurl.com/4r3r77y8.

Ilvento, C., M. Jagadeesan, and S. Chawla. 2020, January. Multi-Category Fairness in Sponsored Search Auctions. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, 348–358. Barcelona Spain: ACM.

Internet Advertising Bureau, and PwC. 2020, May. IAB Internet Advertising Revenue Report 2019. Technical report.

Iyer, K., R. Johari, and M. Sundararajan. 2014, December. Mean Field Equilibria of Dynamic Auctions with Learning. *Management Science* 60 (12): 2949–2970.

Jacobides, M. G., S. Brusoni, and F. Candelon. 2021, December. The Evolutionary Dynamics of the Artificial Intelligence Ecosystem. *Strategy Science* 6 (4): 412–435.

Jain, H., B. Padmanabhan, P. A. Pavlou, and T. S. Raghu. 2021, September. Editorial for the Special Section on Humans, Algorithms, and Augmented Intelligence: The Future of Work, Organizations, and Society. *Information Systems Research* 32 (3): 675–687.

Jarrahi, M. H. 2018, July. Artificial Intelligence and the Future of Work: Human-AI Symbiosis in Organizational Decision Making. *Business Horizons* 61 (4): 577–586.

Jingyu Li, Mengxiang Li, Xincheng Wang, and J. B. Thatcher. 2021, September. Strategic Directions for AI: The Role of CIOs and Boards of Directors. *MIS Quarterly* 45 (3): 1603–1643.

Jung, C., S. Kannan, C. Lee, M. Pai, A. Roth, and R. Vohra. 2020, July. Fair Prediction with Endogenous Behavior. In *Proceedings of the 21st ACM Conference on Economics and Computation*, 677–678. Virtual Event Hungary: ACM.

Kawaguchi, K. 2021, March. When Will Workers Follow an Algorithm? A Field Experiment with a Retail Business. *Management Science* 67 (3): 1670–1695.

Kellogg, K. C., M. A. Valentine, and A. Christin. 2020, January. Algorithms at Work: The New Contested Terrain of Control. *Academy of Management Annals* 14 (1): 366–410.

Kingsley, S., C. Wang, A. Mikhalenko, P. Sinha, and C. Kulkarni. 2020. Auditing Digital Platforms for Discrimination in Economic Opportunity Advertising. 29.

Kleinberg, J., J. Ludwig, S. Mullainathan, and A. Rambachan. 2018, May. Algorithmic Fairness. *AEA Papers and Proceedings* 108:22–27.

Kumar, A., and Y. R. Tan. 2015, August. The Demand Effects of Joint Product Advertising in Online Videos. *Management Science* 61 (8): 1921–1937.

Kumar, V., B. Rajan, R. Venkatesan, and J. Lecinski. 2019, August. Understanding the Role of Artificial Intelligence in Personalized Engagement Marketing. *California Management Review* 61 (4): 135–155.

Lambrecht, A., and C. Tucker. 2019, July. Algorithmic Bias? An Empirical Study of Apparent Gender-Based Discrimination in the Display of STEM Career Ads. *Management Science* 65 (7): 2966–2981.

Lazear, E. P. 2000. Performance Pay and Productivity. *The American Economic Review* 90 (5): 75.

Lazear, E. P. 2018, August. Compensation and Incentives in the Workplace. *Journal of Economic Perspectives* 32 (3): 195–214.

Lebovitz, S., N. Levina, and H. Lifshitz-Assaf. 2021, September. Is AI Ground Truth Really True? The Dangers of Training and Evaluating AI Tools Based on Experts' Know-What. *MIS Quarterly* 45 (3): 1501–1525.

Lee, D., A. Gopal, and S.-H. Park. 2020, September. Different but Equal? A Field Experiment on the Impact of Recommendation Systems on Mobile and Personal Computer Channels in Retail. *Information Systems Research* 31 (3): 892–912.

Levin, J., and P. Milgrom. 2010, May. Online Advertising: Heterogeneity and Conflation in Market Design. *American Economic Review* 100 (2): 603–607.

Levy, F. 2018, July. Computers and Populism: Artificial Intelligence, Jobs, and Politics in the Near Term. *Oxford Review of Economic Policy* 34 (3): 393–417.

Li, X., J. Grahl, and O. Hinz. 2022, June. How Do Recommender Systems Lead to Consumer Purchases? A Causal Mediation Analysis of a Field Experiment. *Information Systems Research* 33 (2): 620–637.

Logg, J. M., J. A. Minson, and D. A. Moore. 2019, March. Algorithm Appreciation: People Prefer Algorithmic to Human Judgment. *Organizational Behavior and Human Decision Processes* 151:90–103.

Longoni, C., and L. Cian. 2022, January. Artificial Intelligence in Utilitarian vs. Hedonic Contexts: The "Word-of-Machine" Effect. *Journal of Marketing* 86 (1): 91–108.

Lou, B., and L. Wu. 2021, September. AI on Drugs: Can Artificial Intelligence Accelerate Drug Development? Evidence from a Large-Scale Examination of Bio-Pharma Firms. *MIS Quarterly* 45 (3): 1451–1482.

Luo, X., S. Tong, Z. Fang, and Z. Qu. 2019, November. Frontiers: Machines vs. Humans: The Impact of Artificial Intelligence Chatbot Disclosure on Customer Purchases. *Marketing Science* 38 (6): 937–947.

Mahroof, K. 2019, April. A Human-Centric Perspective Exploring the Readiness Towards Smart Warehousing: The Case of a Large Retail Distribution Warehouse. *International Journal of Information Management* 45:176–190.

Makridakis, S. 2017, June. The Forthcoming Artificial Intelligence (AI) Revolution: Its Impact on Society and Firms. *Futures* 90:46–60.

Marotta, V., Y. Wu, K. Zhang, and A. Acquisti. 2023. The Welfare Impact of Targeted Advertising Technologies. *Information Systems Research*:60.

Mele, C., M. Marzullo, S. Morande, and T. R. Spena. 2022. How Artificial Intelligence Enhances Human Learning Abilities: Opportunities in the Fight Against COVID-19. *Service Science*:14.

Mitchener, L., D. Tuckey, M. Crosby, and A. Russo. 2022, April. Detect, Understand, Act: A Neuro-symbolic Hierarchical Reinforcement Learning Framework. *Machine Learning* 111 (4): 1523–1549.

Nasr, M., and M. C. Tschantz. 2020, January. Bidding Strategies with Gender Nondiscrimination Constraints for Online Ad Auctions. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, 337–347. Barcelona Spain: ACM.

Obermeyer, Z., B. Powers, C. Vogeli, and S. Mullainathan. 2019, October. Dissecting Racial Bias in an Algorithm Used to Manage the Health of Populations. *Science* 366 (6464): 447–453.

Oestreicher-Singer, G., and A. Sundararajan. 2012. Recommendation Networks and the Long Tail of Electronic Commerce. *MIS Quarterly* 36 (1): 65–83.

Rai, A., P. Constantinides, and S. Sarker. 2019, March. Next Generation Digital Platforms: Toward Human-AI Hybrids. *MIS Quarterly* 43 (1): iii–ix.

Raisch, S., and S. Krakowski. 2021, January. Artificial Intelligence and Management: The Automation–Augmentation Paradox. *Academy of Management Review* 46 (1): 192–210.

Ransbotham, S., S. Khodabandeh, D. Kiron, F. Candelon, M. Chu, and B. LaFountain. 2020, October. Expanding AI's Impact With Organizational Learning. Technical report, MIT Sloan Management Review and Boston Consulting Group.

Reisenbichler, M., T. Reutterer, D. A. Schweidel, and D. Dan. 2022, May. Frontiers: Supporting Content Marketing with Natural Language Generation. *Marketing Science* 41 (3): 441–452.

Roemer, J. E., and A. Trannoy. 2016, December. Equality of Opportunity: Theory and Measurement. *Journal of Economic Literature* 54 (4): 1288–1332.

Rubel, O., and A. Prasad. 2016, July. Dynamic Incentives in Sales Force Compensation. *Marketing Science* 35 (4): 676–689.

Ryman-Tubb, N. F., P. Krause, and W. Garn. 2018, November. How Artificial Intelligence and Machine Learning Research Impacts Payment Card Fraud Detection: A Survey and Industry Benchmark. *Engineering Applications of Artificial Intelligence* 76:130–157.

Schanke, S., G. Burtch, and G. Ray. 2021, September. Estimating the Impact of "Humanizing" Customer Service Chatbots. *Information Systems Research* 32 (3): 736–751.

Senoner, J., T. Netland, and S. Feuerriegel. 2021, December. Using Explainable Artificial Intelligence to Improve Process Quality: Evidence from Semiconductor Manufacturing. *Management Science*.

Shimao, H., W. Khern-am-nuai, and K. N. Kannan. 2021, May. Addressing Fairness in Machine Learning Predictions: Strategic Best-Response Fair Discriminant Removed Algorithm. SSRN Scholarly Paper ID 3389631, Social Science Research Network, Rochester, NY.

Sjödin, D., V. Parida, M. Palmié, and J. Wincent. 2021, September. How AI Capabilities Enable Business Model Innovation: Scaling AI Through Co-Evolutionary Processes and Feedback Loops. *Journal of Business Research* 134:574–587.

Speicher, T., M. Ali, G. Venkatadri, F. N. Ribeiro, G. Arvanitakis, F. Benevenuto, K. P. Gummadi, P. Loiseau, and A. Mislove. 2018, January. Potential for Discrimination in Online Targeted Advertising. In *Conference on Fairness, Accountability and Transparency*, 5–19: PMLR.

Statista 2021. Advertising Market in the U.S. Technical report, Statista.

Stroh, L. K., J. M. Brett, J. P. Baumann, and A. H. Reilly. 1996, June. Agency Theory and Variable Pay Compensation Strategies. *Academy of Management Journal* 39 (3): 751–767.

Sturm, T., J. P. Gerlach, L. Pumplun, N. Mesbah, F. Peters, C. Tauchert, N. Nan, and P. Buxmann. 2021, September. Coordinating Human and Machine Learning for Effective Organizational Learning. *MIS Quarterly* 45 (3): 1581–1602.

Sweeney, L. 2013, May. Discrimination in Online Ad Delivery. *Communications of the ACM* 56 (5): 44–54.

Tong, S., N. Jia, X. Luo, and Z. Fang. 2021, September. The Janus Face of Artificial Intelligence Feedback: Deployment Versus Disclosure Effects on Employee Performance. *Strategic Management Journal* 42 (9): 1600–1631.

Tong, S., X. Luo, and B. Xu. 2020, January. Personalized Mobile Marketing Strategies. *Journal of the Academy of Marketing Science* 48 (1): 64–78.

Tucker, C. 2018, January. Privacy, Algorithms, and Artificial Intelligence. In *The Economics of Artificial Intelligence: An Agenda*, 423–437. University of Chicago Press.

van den Broek, E., A. Sergeeva, and M. Huysman. 2021, September. When the Machine Meets the Expert: An Ethnography of Developing AI for Hiring. *MIS Quarterly* 45 (3): 1557–1580.

Webb, M. 2020. The Impact of Artificial Intelligence on the Labor Market. 61.

Yu, K.-H., A. L. Beam, and I. S. Kohane. 2018, October. Artificial Intelligence in Healthcare. *Nature Biomedical Engineering* 2 (10): 719–731.

Zhang, Y., Q. V. Liao, and R. K. E. Bellamy. 2020, January. Effect of Confidence and Explanation on Accuracy and Trust Calibration in AI-Assisted Decision Making. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, FAT* '20, 295–305. New York, NY, USA: Association for Computing Machinery.