Replication of ‘The Managerial Effects of Algorithmic Fairness Activism’

by Bo Cowgill, Fabrizio Dell’Acqua, and Sandra Matz.

Asel Kushkeyeva

2021-04-24

Contents

Abstract

A growing interest in ethical discussions of artificial intelligence (AI) adoption and algorithmic bias in industry and society has led to exploration of business decision-makers' reaction to the debates. I replicated The Managerial Effects of Algorithmic Fairness Activism by Bo Cowgill, Fabrizio Dell'Acqua, and Sandra Matz published in American Economic Association journal in 2020. The authors conducted two studies through a survey. For the first study, the respondents were given two types of op-eds - counterfactual to AI adoption and accepting it as something unavoidable - to test whether their decisions were influenced by them in two business cases: hiring and lending. The second study examined an influence of "scientific veneer" on managerial decision-making for the same business cases. I chose to focus on two questions: (1) If you were forced to make a decision on the future use of the algorithm now, what would you do? (2) What is the probability that a fairness issue is alleged and becomes a problem (through lawsuits or PR)? Responses to the questions were recorded as binary and on a Likert scale, thus, I fitted a multiple logistic and linear regressions. In the first study, being exposed to both op-eds influenced the respondents' opinion to recommend AI negatively. This trend, however, changed with introduction of op-eds and an algorithmic fix - a 6-month effort by a technical team to address fairness issue. Ability to see a status quo, either by default or optionally, increased the odds of recommending AI adoption. Women and ethnic minorities were less inclined to recommend AI, although not significantly. For the question of "probability that a fairness issue is alleged and becomes a problem", the decision-makers were highly convinced by "scientific veneer" and expert's pro-AI recommendation, believing that the risk of a lawsuit would decline. These findings echoed the original analysis with the exception of opposite trends in reading "fatalistic" op-ed in the first study.

Introduction

With the rapid development of technology, AI has entered our everyday lives, first entertaining us in sci-fi films and finally replacing human labour (Amazon warehouses being one of the most prominent examples). However, an ethical side of technology and adoption of AI was not considered for a long time. Recently, we witnessed a tremendous interest in ethical discussions of AI and algorithmic bias both by industry experts and society at large. Being an active consumer of tech and AI, I deemed it beneficial to explore how businesses - service and product providers, in particular management, respond to this growing interest.

My research landed on “The Managerial Effects of Algorithmic Fairness Activism” by Bo Cowgill, Fabrizio Dell’Acqua, and Sandra Matz (2020), which employed a survey to analyze subjects’ behaviour upon introduction to external discussions on consequences of AI adoption by businesses. I intended to follow the paper as closely as possible, however, the fact this work utilized R statistical software might have altered the replication as the original analysis was performed in Stata. Additionally, we might have used differing types of statistical models as the authors have not explicitly stated what type of regression they used.

The recruited respondents answered the questions based on to two business cases: hiring and lending. The survey consists of five questions, which the subjects had to answer twice to reflect their opinions on AI adoption before and after an algorithmic fix. The algorithmic fix was a 6-month team’s effort to address the fairness issues. Approximately equal number of subjects were recruited for two studies. Study 1 involved reading two op-eds - one was an article published in Fortune Magazine “AI Bias Isn’t the Problem. Our Society Is”; the second op-ed was an essay in Harvard Business review “Want Less-Biased Decisions? Use Algorithms”. The first essay discussed inability of AI to avoid bias and abandoning AI adoption; this op-ed was referred as “fatalistic”. The latter suggested use of AI if it helped to reduce a human bias; the authors of “The Managerial Effects. . . ” labeled it as “counterfactual” as it was often compared to a status quo. In study 2, the respondents were introduced to an expert recommendation regarding AI adoption, which was presented with and without a “scientific veneer”. The recommendation with “scientific veneer” included scientific terms such as probability, p-value, beta coefficients, and logarithmic expressions.

Datasets

For the study, Prolific, an online survey platform, was used to collect data; the survey was conducted on January 8, 2020 in United States. 994 participants in managerial roles who most likely would be responsible for making AI adoption decisions in a workplace were recruited for two studies. The studies were then recorded in two datasets, containing the respondents’ answers to five questions, four of which were recorded on a Likert scale: potential positive/negative impact of technology, whether AI adoption would be recommended, about risk of a lawsuit or PR problems if AI was adopted, and the magnitude of the lawsuits or PR issue would be. The AI recommendation question was also recorded as a binary - “Yes”, “No” - response. The rest of the variables reflected the respondents’ demography and were coded as zeros and ones (Table 1).

Table 1. Distribution of gender and ethnicity in the datasets

Gender	Study1	Study2
Male	0.49	0.52
Female	0.49	0.45
Other gender	0.02	0.02

Ethnicity	Study1	Study2
White	0.81	0.85
Black	0.07	0.05
Asian	0.05	0.04
LatinX	0.10	0.07
Other ethnicity	0.06	0.06

The two datasets differed by a few variables: the first contained “cf” – counterfactual and “fat” – fatalistic variables; the second dataset had “ma” – scientific veneer, “proai” – pro-AI expert recommendation, and “mapro” – scientific veneer X pro-AI expert recommendation. The variables of interest are described in the following section, and some interactions are shown in the figures below. Figures 1 and 3 illustrate no significant trends in AI recommendation after reading the op-eds and being exposed to “scientific veneer” along with pro-AI expert recommendations. In contrast, the responses to “a lawsuits question” tend to avoid extreme 1s and 7s.

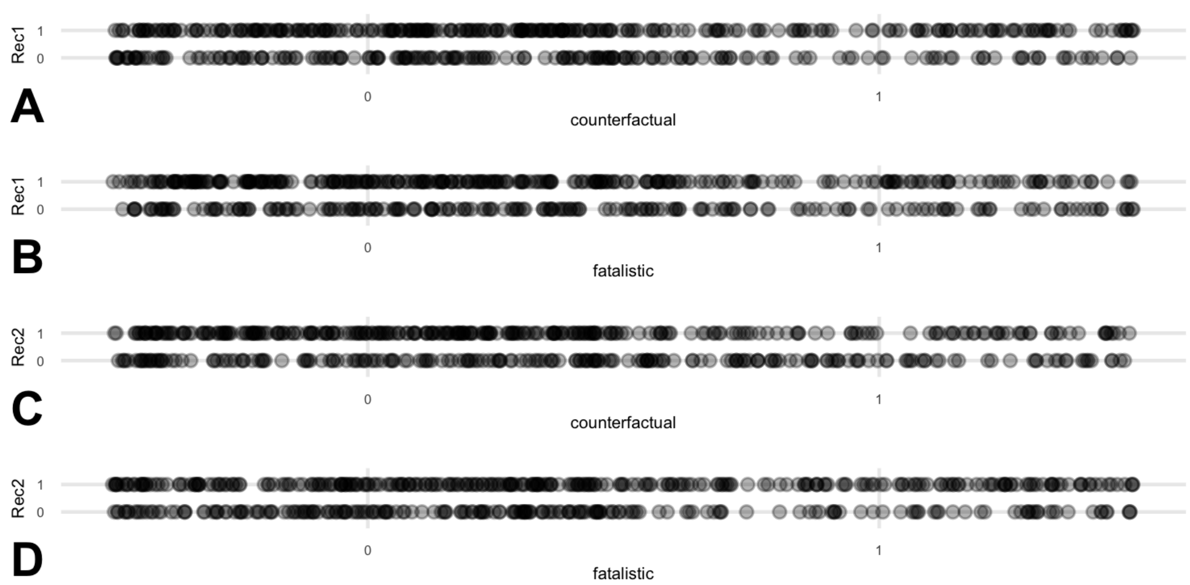


Figure 1: AI recommendation VS “counterfactual” and “fatalistic” for business case 1 (A,B) and 2 (C,D)

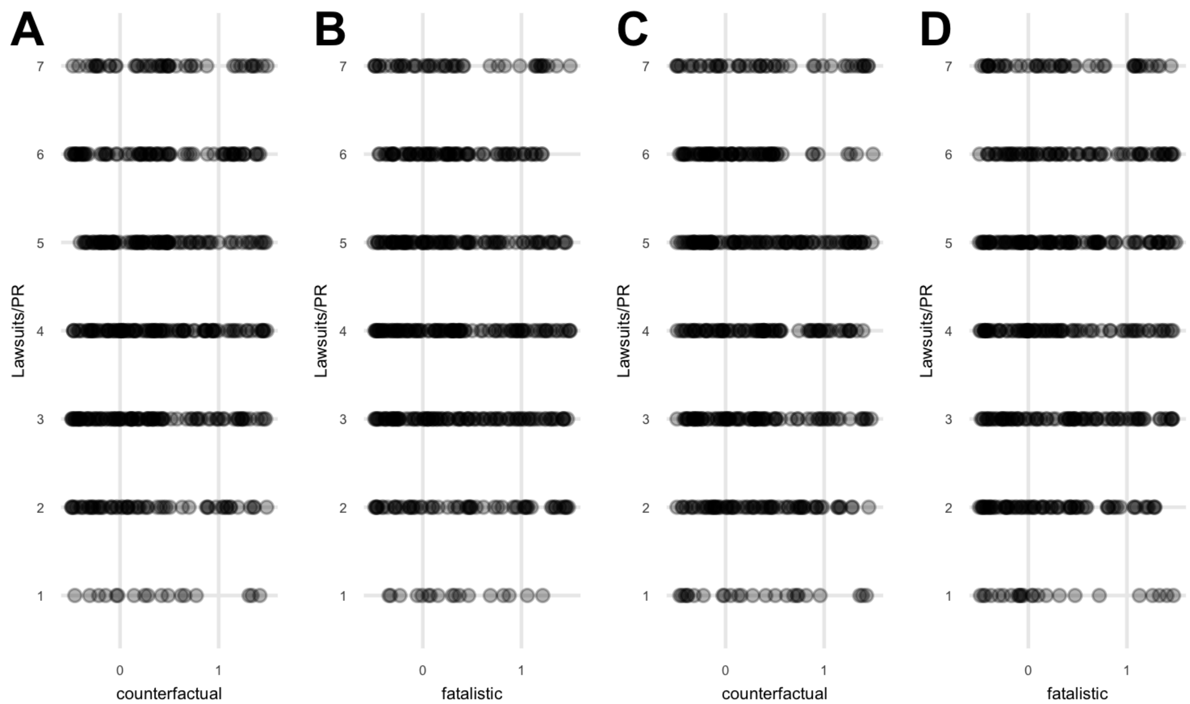


Figure 2: Lawsuits/PR VS “counterfactual” and “fatalistic” for business case 1 (A,B) and 2 (C,D)

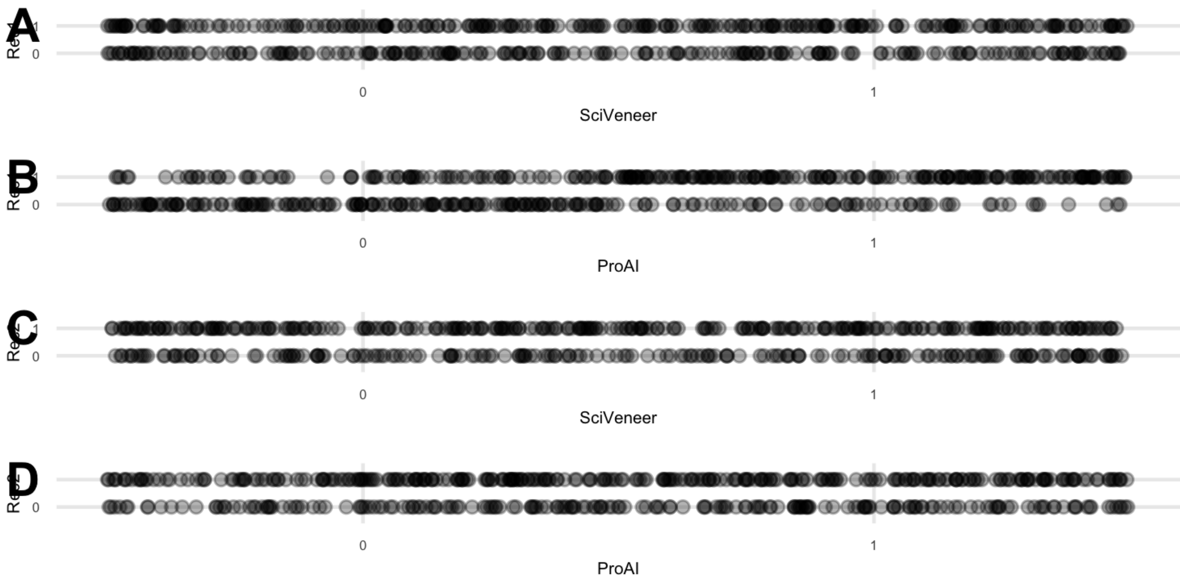


Figure 3: AI recommendation VS "Scientific Veneer" and "Pro AI" for business case 1 (A,B) and 2 (C,D)

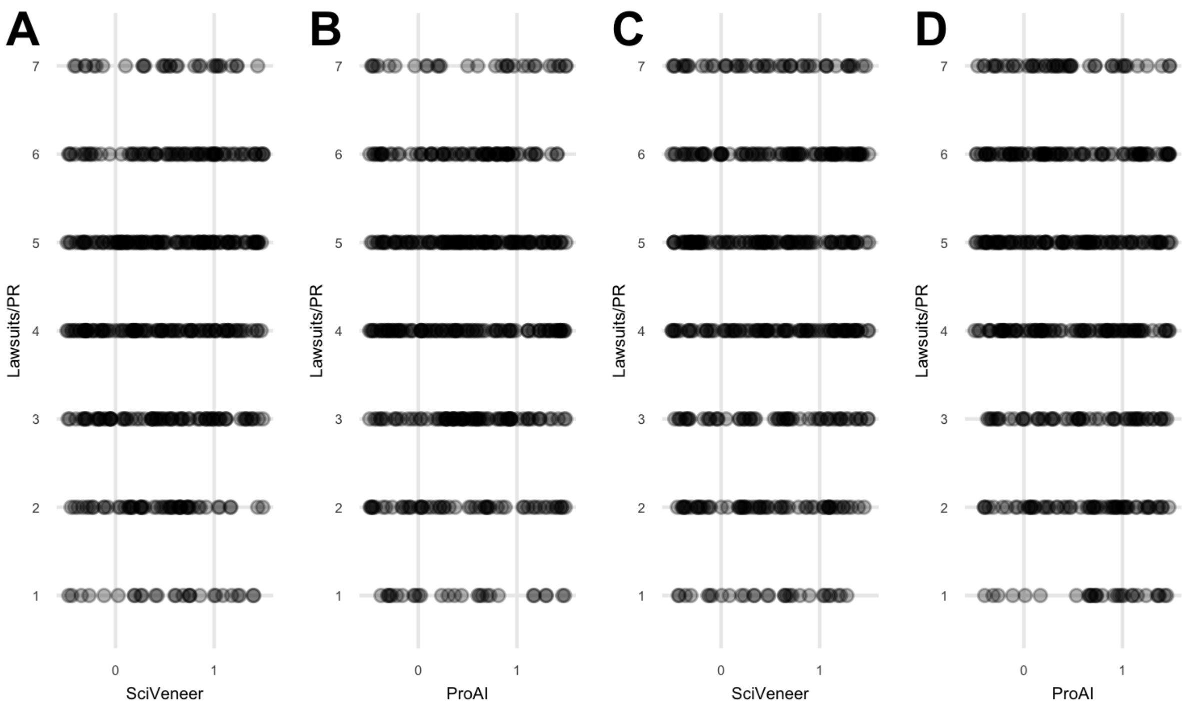


Figure 4: Lawsuits/PR VS "Scientific Veneer" and "Pro AI" for business case 1 (A,B) and 2 (C,D)

Model

Binomial regression model Study 1: Counterfactual and "fatalism" op-eds.

$$Y_{i,j,fix} \sim Binomial(N_i, p_i)$$

$$\log\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_j \text{BusinessCaseOrder} + \beta_i \text{StatusQuo,Gender} + \beta_{j,j}\text{CF} + \beta_{j,j}\text{FAT} +$$
$$\beta_{j,j,fix}\text{CFxFix} + \beta_{j,j,fix}\text{FATxFix} + \beta_i \text{WD,Ed,ML} + \epsilon,$$

Linear regression model Study 1: Counterfactual and "fatalism" op-eds.

$$y_{i,j,fix} = \beta_0 + \beta_j \text{BusinessCaseOrder} + \beta_i \text{StatusQuo,Gender} + \beta_{j,j}\text{CF} + \beta_{j,j}\text{FAT} +$$
$$\beta_{j,j,fix}\text{CFxFix} + \beta_{j,j,fix}\text{FATxFix} + \beta_i \text{WD,Ed,ML} + \epsilon,$$

where:

- $y_{i,j,fix}$ refers to the response to question Y individual i on case j regarding before/after the 6-month effort to fix.
- i indexes 498 subjects for Study 1.
- j = {0, 1} indexes the first or second order of the business cases.
- fix = {0, 1} differentiates answers to the questions before (fix = 0) or after (fix = 1) the six-months of dedicated effort.
- CF refers to the "counterfactual" condition: 1 for the subjects reading the counterfactual in the second case, and 0 for everyone in the first case.
- FAT refers to the "fatalism" condition: 1 for the subjects reading the fatalism in the second case, and 0 for everyone in the first case.
- $CF \cdot Fix$ is equal to CF×1(fix=1).
- $FAT \cdot Fix$ is equal to FAT × 1(fix = 1).
- ϵ is standard errors.

Binomial regression model Study 2: "Scientific veneer".

$$Y_{i,j,fix} \sim Binomial(N_i, p_i)$$

$$\log\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_j \text{BusinessCaseOrder} + \beta_i \text{StatusQuo}+$$
$$\beta_i \text{Gender} + \beta_i \text{Ethnicity} + \beta_i \text{PoliticalConservatism}+$$
$$\beta_{j,j} \text{SciVeneer} + \beta_{j,j} \text{ProAI} + \beta_{j,j} \text{SciVeneerXProAI}+$$
$$\beta_{j,j,fix} \text{SciVeneerxFix} + \beta_{j,j,fix} \text{ProAIxFix} + \beta_{j,j,fix} \text{SciVeneerxProAIxFix}+$$
$$\beta_i \text{WorkDecisions} + \beta_i \text{PreparedEd} + \beta_i \text{KnowsML} + \epsilon,$$

Linear regression model Study 2: "Scientific veneer".

$$y_{i,j,fix} = \beta_0 + \beta_j \text{BusinessCaseOrder} + \beta_i \text{StatusQuo}+$$
$$\beta_i \text{Gender} + \beta_i \text{Ethnicity} + \beta_i \text{PoliticalConservatism}+$$
$$\beta_{j,j} \text{SciVeneer} + \beta_{j,j} \text{ProAI} + \beta_{j,j} \text{SciVeneerXProAI}+$$
$$\beta_{j,j,fix} \text{SciVeneerxFix} + \beta_{j,j,fix} \text{ProAIxFix} + \beta_{j,j,fix} \text{SciVeneerxProAIxFix}+$$
$$\beta_i \text{WorkDecisions} + \beta_i \text{PreparedEd} + \beta_i \text{KnowsML} + \epsilon,$$

where:

- $Y_{i,j,fix}$ refers to the response to question Y individual i on case j regarding before/after the 6-month effort to fix.
- i indexes 496 subjects for Study 2.
- j = {0, 1} indexes the first or second order of the business cases.
- fix = {0, 1} differentiates answers to the questions before (fix = 0) or after (fix = 1) the six-months of dedicated effort.
- $SciVeneer$ refers to the "counterfactual" condition: 1 for the subjects reading the counterfactual in the second case, and 0 for everyone in the first case.
- $ProAI$ refers to the condition in which a positive argument about AI ethics are made.

- $SciVeneer$ refers to whether the expert applied scientific veneer. This is 1 for the subjects seeing scientific veneer in the second case, and 0 for everyone in the first case.
- $SciVeneer \cdot ProAI$ refers to subjects who see pro-AI arguments with scientific veneer. This is 1 for the subjects reading these arguments in the second case, and 0 for everyone in the first case.
- $ProAI \cdot Fix$ is equal to ProAI×1 (fix=1).
- $SciVeneer \cdot Fix$ is equal to SciVen×1 (fix = 1).
- $ProAI \cdot SciVeneer \cdot Fix$ is equal to ProAI×1 (fix=1).
- ϵ is standard errors.

Analysis and Discussion

The survey performed by Cowgill and their colleagues included five questions. I chose to focus on the response to two questions: (1) If you were forced to make a decision on the future use of the algorithm now, what would you do? (2) What is the probability that a fairness issue is alleged and becomes a problem (through lawsuits or PR)? Since for the first question the variable of interest was binary (Figure 5) a multiple logistic regression was fitted. A linear regression was appropriate as the responses to the second question were recorded on a Likert scale - from 1 to 7 (Figure 6). In both cases, I included the predictors and their interactions chosen by Cowgill and their colleagues (Table 2).

Table 2. Response and predictor variables for Study 1 and 2

Dependent variables	class	description
recyesno_post_first	numeric	A binary response to a question of AI adoption recommendation for first business case
recyesno_post_second	numeric	A binary response to a question of AI adoption recommendation for second business case
lawsuitspr_post_first	numeric	A response to a question of a perceived likelihood of a lawsuit or PR problem as a result of AI adoption for first business case
lawsuitspr_post_second	numeric	A response to a question of a perceived likelihood of a lawsuit or PR problem as a result of AI adoption for second business case

Independent variables	class	description
response_id	numeric	Unique ID
hiringfirst	numeric	Hiring is equal to 1 for the hiring business case and equal to 0 for the lending business case
sqshw	numeric	Status Quo shown
sqopt	numeric	Status Quo shown (only if clicked)
sqhid	numeric	Status Quo hidden
consrv_std	numeric	Political conservatism, standardized

Independent variables	class	description
cf	numeric	Refers to the "counterfactual" condition. This is 1 for the subjects reading the counterfactual in the second case, and 0 for everyone in the first case
fat	numeric	Refers to the "fatalism" condition. This is 1 for the subjects reading the fatalism in the second case, and 0 for everyone in the first case
ma	numeric	Scientific Veneer
proai	numeric	Refers to the condition in which a positive argument about AI ethics are made. This is 1 for the subjects reading the positive opinion in the second business case, and 0 for everyone in the first case
mapro	numeric	Scientific Veneer × Pro AI
female	numeric	Gender
black	numeric	Ethnicity
asian	numeric	Ethnicity
workdecisions	numeric	AI decisions captures subjects who report working (or potentially working) in roles where they make decisions like those in our surveys
eduprepared	numeric	Indicates whether subjects feel their education has prepared them well enough for this type of decisions
knowsml	numeric	takes the value 1 if subjects know "a great deal", "a lot", or "a moderate amount" about machine learning and predictive modeling, and 0 otherwise

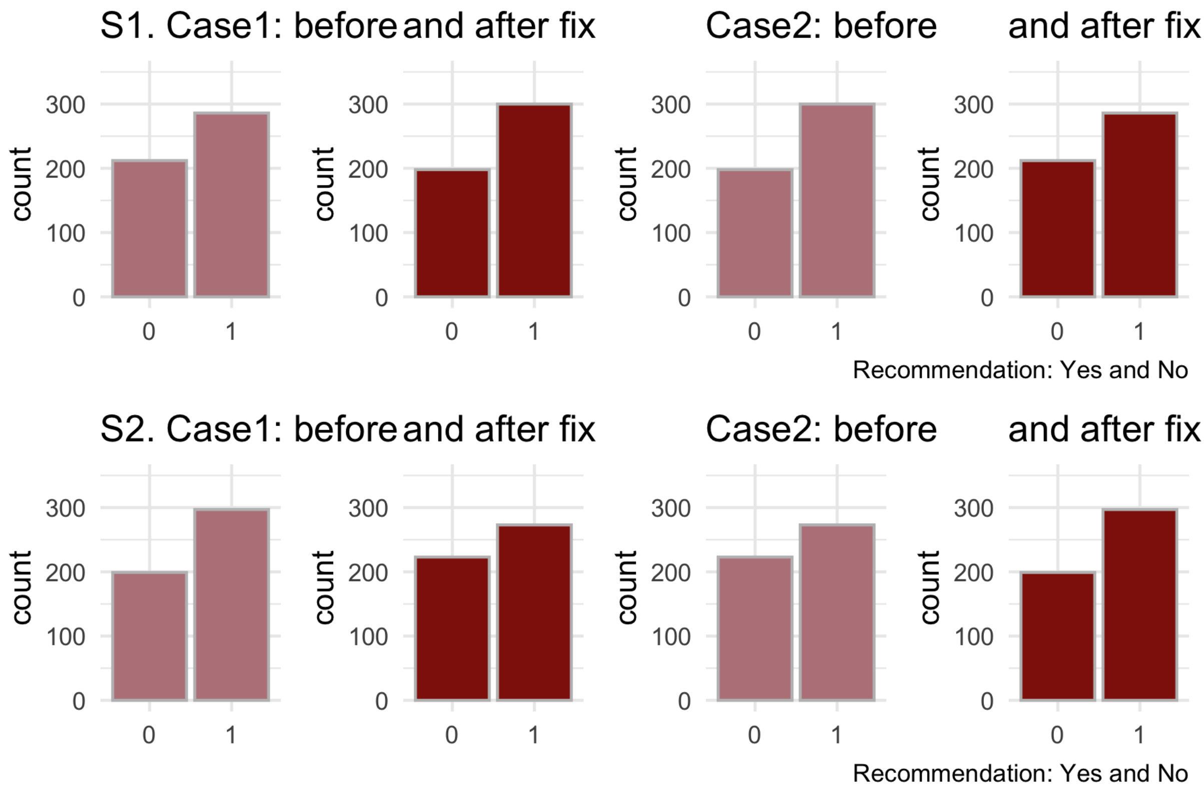


Figure 5: Distribution of responses for a question of AI recommendation before and after algorithmic fix for two business cases (Study1 top row, Study2 bottom row)

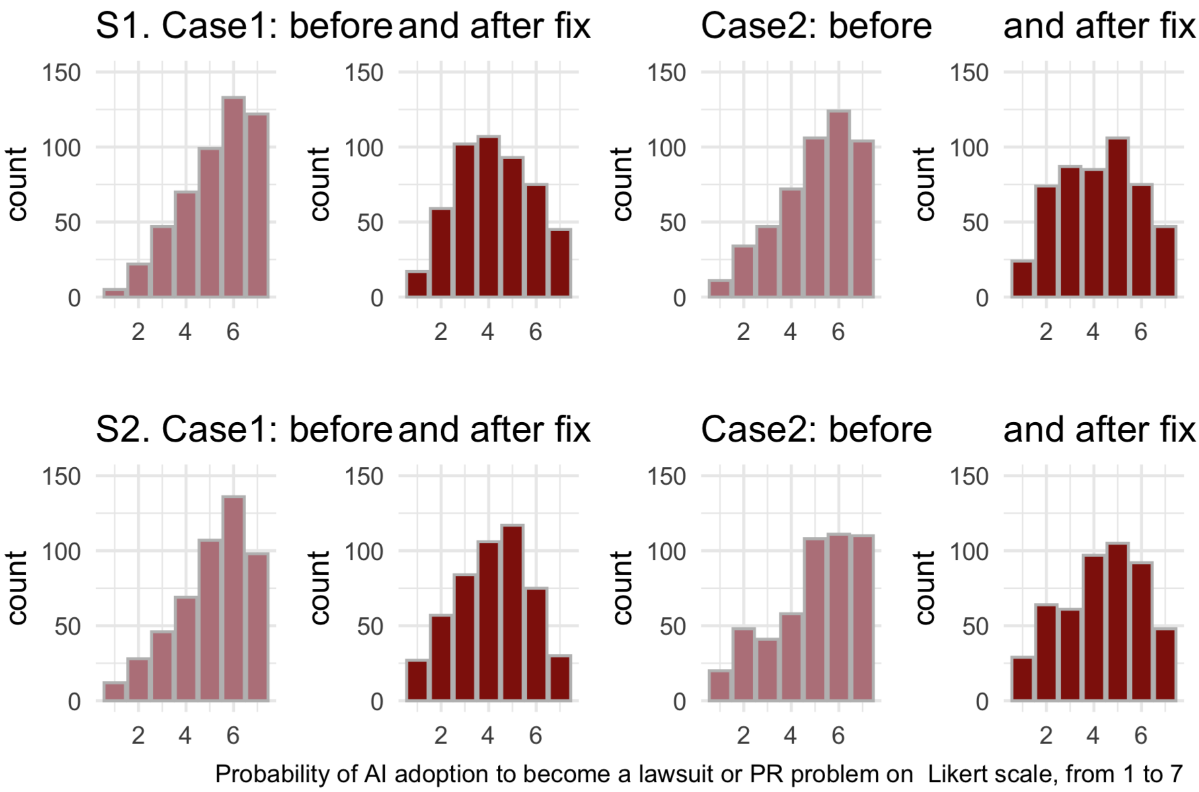


Figure 6: Distribution of responses for a question of likelihood of a lawsuit or PR problem (Study1 top row, Study2 bottom row)

In the first study, being exposed to both "counterfactual" and "fatalistic" op-eds influenced the respondents' opinion to recommend AI negatively - the coefficients ranged from -2.23 to -1.15. Ability to see a status quo, either by default or optionally, increased the odds of recommending AI adoption (Figure 8). Women and ethnic minorities were less inclined to recommend AI, although not significantly. The binomial regression produced positive coefficients for interaction of op-eds and an algorithmic fix - a 6-month effort to remove/reduce bias, illustrating that the respondents would strongly recommend AI adoption no matter what type of op-ed they were exposed to.

Summary 1. Logistic regression summary on AI adoption recommendation question, Study 1

First case

term	estimate	std.error	statistic	p.value
(Intercept)	-0.1732241	0.3505399	-0.4941637	0.6211906
hiringfirst	-0.1572774	0.2123516	-0.7406462	0.4589080
cf	-1.1533153	0.2962578	-3.8929448	0.0000990
fat	-1.2412241	0.3388530	-3.6630160	0.0002493
sqshw	1.2711924	0.2602588	4.8843396	0.0000010
sqopt	1.1480343	0.2548985	4.5038874	0.0000067
female	-0.0565831	0.2165709	-0.2612682	0.7938857
black	-0.3623857	0.3983581	-0.9096984	0.3629816
asian	-0.4945857	0.4581806	-1.0794556	0.2803847
consrv_std	-0.0105880	0.1058624	-0.1000171	0.9203307
workdecisions	0.1531968	0.2216228	0.6912503	0.4894083
eduprepared	0.2278003	0.2436334	0.9350124	0.3497819
knowsml	-0.1722841	0.2220503	-0.7758788	0.4378205
cf:recyesno_pre_first	2.6783369	0.4507108	5.9424729	0.0000000
fat:recyesno_pre_first	1.3492855	0.3623482	3.7237261	0.0001963

Second case

term	estimate	std.error	statistic	p.value
(Intercept)	-0.3648016	0.3483324	-1.0472800	0.2949705
hiringfirst	0.1218816	0.2101596	0.5799479	0.5619497
cf	-2.2306033	0.4285133	-5.2054467	0.0000002
fat	-0.4065116	0.2963948	-1.3715205	0.1702128
sqshw	1.0755776	0.2603576	4.1311559	0.0000361
sqopt	1.0600339	0.2542837	4.1687051	0.0000306
female	-0.3730211	0.2132438	-1.7492706	0.0802442

term	estimate	std.error	statistic	p.value
black	-0.0738660	0.3849311	-0.1918942	0.8478251
asian	0.7232239	0.5156447	1.4025625	0.1607474
consrv_std	0.0351728	0.1056781	0.3328293	0.7392632
workdecisions	0.1052301	0.2187261	0.4811046	0.6304422
eduprepared	0.1289503	0.2455922	0.5250588	0.5995423
knowsml	-0.0313712	0.2195968	-0.1428580	0.8864023
cf:recyesno_pre_second	2.6602313	0.4597318	5.7864849	0.0000000
fat:recyesno_pre_second	1.3702629	0.3581493	3.8259549	0.0001303

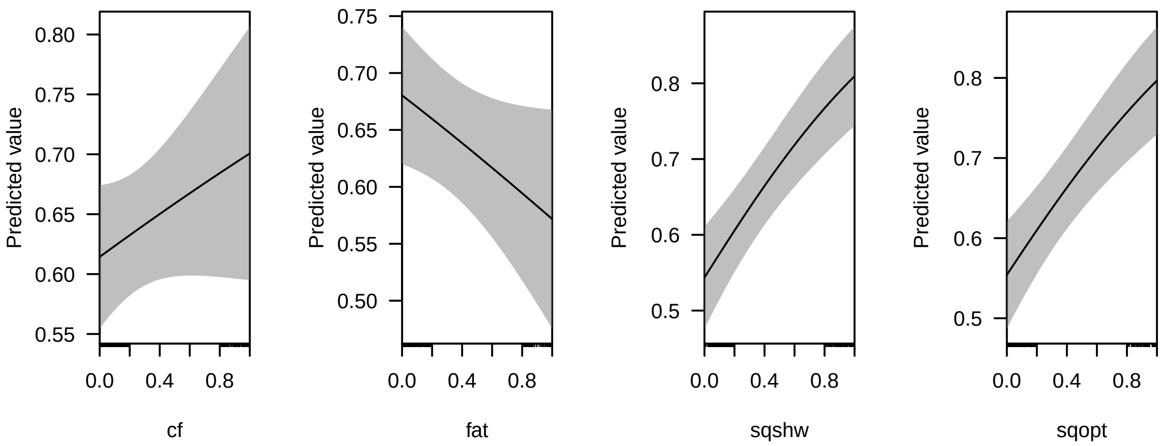


Figure 8: *Plotted* ***predicted values*** *for the variables of interest such as "Counterfactual", "Fatalistic", "Status Quo Shown", and "Status Quo Shown (if clicked)"*

I have performed a basic regression assumption check for all the models by plotting them. Figure 7 in Appendix A shows that the assumptions of linearity and normality are met (Residuals vs Fitted and Normal Q-Q plots), whereas the assumption of homoscedasticity (constant variance) does not hold too well. The Residuals vs Leverage plot demonstrates no outliers in the data.

Since a binary logistic regression does not produce an R2, many conflicting discussions persist on methods of measuring goodness-of-fit of the model. McFadden's Pseudo-R2 is one of such methods, which I employed in the analysis. According to McFadden (1977), the Pseudo-R2's values are considerably lower, thus Pseudo-R2 of 0.18 is high enough, indicating that our model fits the data well and has moderate to high predictive power. I have also checked for multicollinearity of the independent variables. Although the "counterfactual" and algorithmic fix interaction term has a higher value, I have kept it the regression in order to be consistent with the original paper. These results can be found in Appendix B.

For the second question of a possibility of a lawsuit or a PR problem, "cf" and "fat" and their interaction with an algorithmic fix proved to be important predictors. Here, after reading both "cf" and "fat" op-eds, the respondents were convinced that issue arising from AI adoption would be reduced (Summary 2).

Summary 2. Linear regression summary on lawsuits/PR problems likelihood question, Study 1

First case

term	estimate	std.error	statistic	p.value
(Intercept)	4.4438950	0.2287854	19.4238551	0.0000000
hiringfirst	0.0389714	0.1342110	0.2903740	0.7716548
cf	-3.1745313	0.4183795	-7.5876844	0.0000000
fat	-2.1692116	0.4284501	-5.0629272	0.0000006
sqshw	-0.1162516	0.1647948	-0.7054328	0.4808810
sqopt	-0.0056922	0.1645276	-0.0345970	0.9724154
female	-0.1816393	0.1389498	-1.3072297	0.1917567
black	0.2383866	0.2546638	0.9360837	0.3496979
asian	0.1761526	0.2942647	0.5986196	0.5497072
consrv_std	0.0410800	0.0672242	0.6110893	0.5414279
workdecisions	0.1049839	0.1399249	0.7502876	0.4534469
eduprepared	-0.2607534	0.1555326	-1.6765197	0.0942835
knowsml	0.2206601	0.1406774	1.5685538	0.1174068
cf:lawsuitspr_pre_first	0.5974836	0.0740703	8.0664408	0.0000000
fat:lawsuitspr_pre_first	0.3720215	0.0772936	4.8130954	0.0000020

Second case

term	estimate	std.error	statistic	p.value
(Intercept)	4.2339585	0.2275662	18.6053890	0.0000000
hiringfirst	0.2207333	0.1327590	1.6626612	0.0970290
cf	-3.1876161	0.3721272	-8.5659319	0.0000000
fat	-3.3176448	0.4252373	-7.8018661	0.0000000
sqshw	0.1144499	0.1647464	0.6947037	0.4875750
sqopt	-0.0373617	0.1631664	-0.2289792	0.8189820
female	-0.1877425	0.1346724	-1.3940682	0.1639380
black	-0.0797237	0.2523209	-0.3159614	0.7521683
asian	0.5381999	0.2915946	1.8457130	0.0655457
consrv_std	-0.0621145	0.0666323	-0.9321974	0.3517003
workdecisions	0.1300932	0.1385974	0.9386406	0.3483844
eduprepared	-0.3916069	0.1538250	-2.5457948	0.0112125
knowsml	0.1578486	0.1394000	1.1323426	0.2580523
cf:lawsuitspr_pre_second	0.6429172	0.0699344	9.1931529	0.0000000
fat:lawsuitspr_pre_second	0.6747777	0.0760419	8.8737613	0.0000000

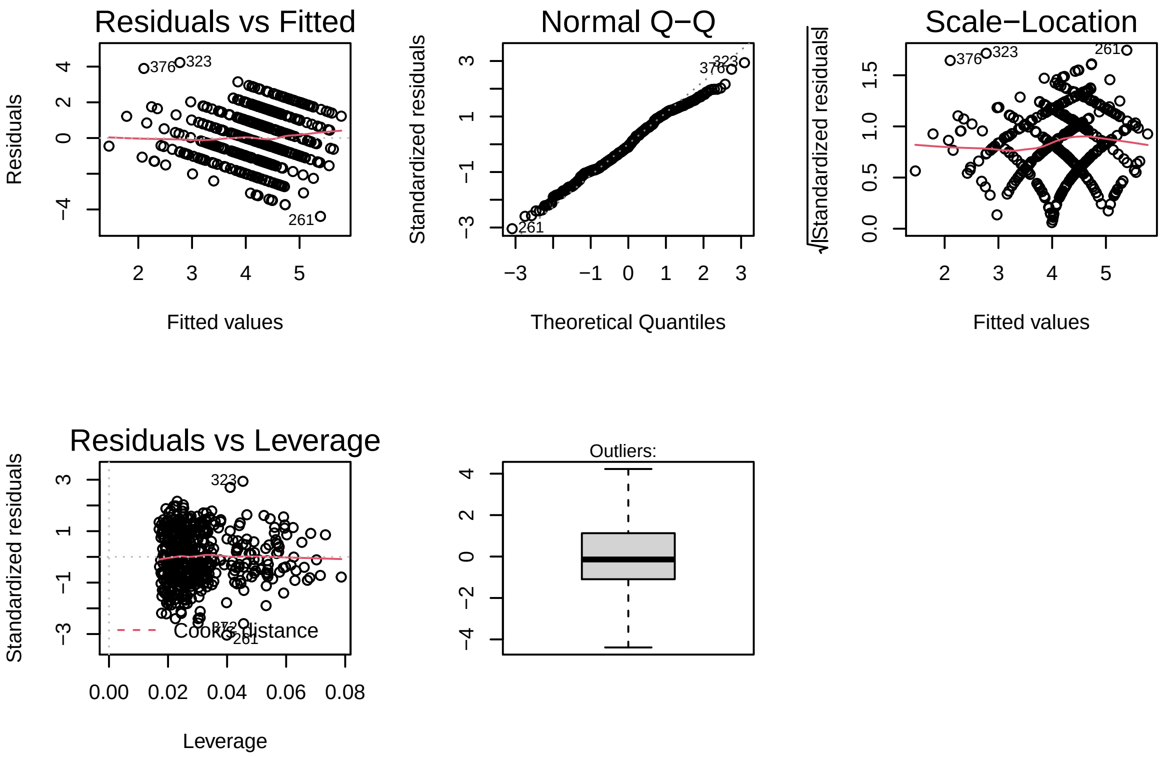


Figure 9: *Linear regression assumptions check: linearity, normality, homoscedasticity, and outliers for business case 1 and 2.*

As outlined earlier, the second study involved an AI expert recommendation with and without “scientific veneer”, which was not too significant on its own (Figure 10). As shown in Summary 3, AI adoption recommendation, the predictors “Status Quo”, “ProAI” and its interactions with “SciVeneer” and “Fix” were mostly positively associated with AI adoption, except for “ProAI” predictor in the second business case. Political conservatism of the respondents was slightly influential on AI recommendation, too. The goodness-of-fit for this model was not significantly different from the previous study; it can be found in the Appendix D.

Summary 3. Logistic regression summary on AI adoption recommendation question, Study 2

First case

term	estimate	std.error	statistic	p.value
(Intercept)	-1.7805034	0.4101229	-4.3413903	0.0000142
hiringfirst	0.1925518	0.2213708	0.8698159	0.3844010
ma	-0.5506492	0.4091276	-1.3459107	0.1783313
proai	1.2223226	0.3743568	3.2651276	0.0010941
mapro	1.3836634	0.6082050	2.2749950	0.0229062
sqshw	1.0231991	0.2774428	3.6879637	0.0002261
sqopt	1.2761181	0.2764672	4.6158026	0.0000039
female	0.0425517	0.2259453	0.1883273	0.8506201
black	-0.5050278	0.5029926	-1.0040461	0.3153564
asian	-0.6550784	0.5266570	-1.2438426	0.2135575
consrv_std	0.3145080	0.1136298	2.7678301	0.0056431
workdecisions	-0.2291268	0.2283685	-1.0033205	0.3157063
eduprepared	0.1557664	0.2421392	0.6432931	0.5200339
knowsml	0.2792793	0.2286383	1.2214896	0.2219007
ma:recyesno_pre_first	0.9672696	0.4436199	2.1804015	0.0292277
proai:recyesno_pre_first	1.3522637	0.4690685	2.8828706	0.0039407
mapro:recyesno_pre_first	-1.7929779	0.8008658	-2.2387993	0.0251690

Second case

term	estimate	std.error	statistic	p.value
(Intercept)	0.1012615	0.3554570	0.2848768	0.7757385
hiringfirst	-0.4971224	0.2015102	-2.4669835	0.0136257
ma	-0.6473687	0.3106797	-2.0837174	0.0371859
proai	-1.2959979	0.4548209	-2.8494684	0.0043792
mapro	1.2569887	0.6651117	1.8898912	0.0587725
sqshw	1.2063438	0.2484931	4.8546374	0.0000012
sqopt	0.9347808	0.2411541	3.8762798	0.0001061
female	-0.2102189	0.2079004	-1.0111522	0.3119436
black	-0.6835214	0.4610269	-1.4826063	0.1381790
asian	-0.0541329	0.4849890	-0.1116168	0.9111273
consrv_std	0.0615562	0.1023040	0.6016987	0.5473747
workdecisions	-0.2931629	0.2104923	-1.3927488	0.1636958
eduprepared	0.1771556	0.2214832	0.7998602	0.4237918
knowsml	0.3922572	0.2122224	1.8483309	0.0645545
ma:recyesno_pre_second	0.7594152	0.4374980	1.7358139	0.0825967
proai:recyesno_pre_second	1.5361060	0.4755498	3.2301686	0.0012372
mapro:recyesno_pre_second	-1.6501307	0.8014436	-2.0589479	0.0394992

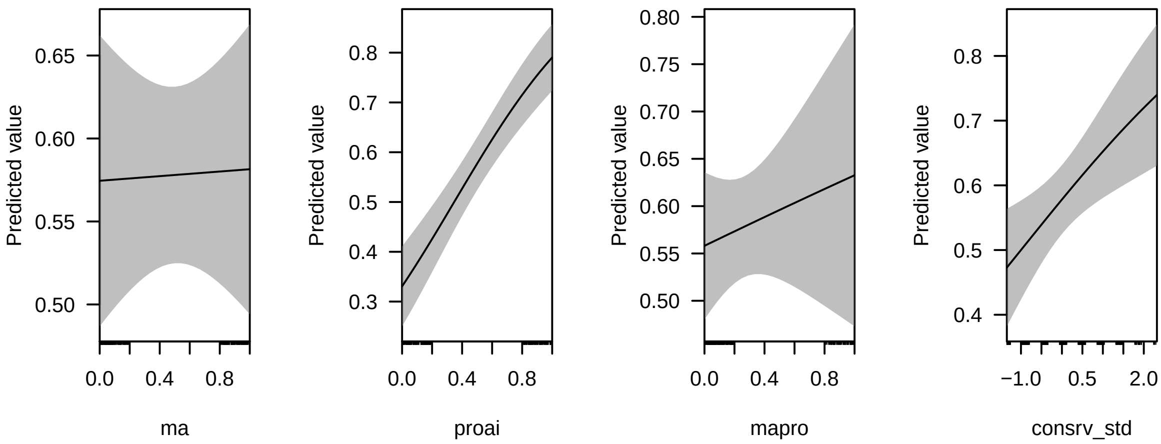


Figure 10 *Plotted* ***predicted values*** *for the variables of interest such as “Scientific Veneer”, “Pro-AI expert recommendation”, “Scientific Veneer X Pro-AI”, and “Political conservatism”*

For the question of “probability that a fairness issue is alleged and becomes a problem”, the decision-makers were highly convinced by “scientific veneer” and expert’s pro-AI recommendation, believing that the risk of a lawsuit would decline. These beliefs did not hold well after interaction with algorithmic fix. Surprisingly, the respondents’ specialized education in AI, their knowledge of machine learning, and practical experience in decision-making did not prove to be important in these studies setting until this point: in the second business case knowledge of machine learning and work experience in decision-making were significantly influential on perceived likelihood of a problem or lawsuit (Summary 4).

Summary 4. Linear regression summary on lawsuits/PR problems likelihood question, Study 2

First case

term	estimate	std.error	statistic	p.value
(Intercept)	3.8547416	0.2309048	16.6940709	0.0000000
hiringfirst	0.3208538	0.1282052	2.5026589	0.0126585
ma	-2.4449561	0.4659106	-5.2476941	0.0000002
proai	-2.2516222	0.4845518	-4.6468146	0.0000044
mapro	2.5499139	0.7924802	3.2176373	0.0013801
sqshw	-0.0443250	0.1585141	-0.2796280	0.7798837
sqopt	-0.1714459	0.1560750	-1.0984840	0.2725450
female	-0.0823739	0.1331930	-0.6184548	0.5365696
black	-0.1828925	0.3032796	-0.6030493	0.5467613
asian	0.3610074	0.3201176	1.1277335	0.2599971
consrv_std	-0.0596851	0.0649708	-0.9186443	0.3587440
workdecisions	-0.0630545	0.1342897	-0.4695409	0.6388967
eduprepared	-0.1278893	0.1428583	-0.8952181	0.3711201
knowsml	0.0944532	0.1346047	0.7017081	0.4832020
ma:lawsuitspr_pre_first	0.5626112	0.0828860	6.7877688	0.0000000
proai:lawsuitspr_pre_first	0.5233897	0.0871220	6.0075476	0.0000000
mapro:lawsuitspr_pre_first	-0.5989305	0.1453767	-4.1198510	0.0000447

Second case

term	estimate	std.error	statistic	p.value
(Intercept)	4.7482553	0.2326138	20.4126107	0.0000000
hiringfirst	0.1351958	0.1293371	1.0452980	0.2964125
ma	-2.7142529	0.5014775	-5.4125113	0.0000001
proai	-3.0459986	0.3849838	-7.9120175	0.0000000
mapro	1.9471374	0.7046904	2.7631104	0.0059457

term	estimate	std.error	statistic	p.value
sqshw	-0.0617526	0.1580350	-0.3907524	0.6961540
sqopt	-0.1095244	0.1573910	-0.6958748	0.4868447
female	0.0277791	0.1339205	0.2074300	0.8357622
black	0.2864674	0.3049633	0.9393502	0.3480244
asian	0.5924273	0.3191176	1.8564544	0.0640026
consrv_std	-0.0650047	0.0655853	-0.9911477	0.3221139
workdecisions	0.3492109	0.1353104	2.5808122	0.0101532
eduprepared	-0.2598572	0.1442944	-1.8008829	0.0723502
knowsml	-0.2878272	0.1360734	-2.1152353	0.0349256
ma:lawsuitspr_pre_second	0.5024448	0.0841997	5.9672999	0.0000000
proai:lawsuitspr_pre_second	0.5253054	0.0773808	6.7885726	0.0000000
mapro:lawsuitspr_pre_second	-0.3051929	0.1345937	-2.2675120	0.0238037

Overall, the results of this analysis follow closely the original work of Cowgill and their colleagues. In both analyses, women and ethnic minorities opposed AI adoption. Along with the authors of the original paper, I witnessed strong trends that six months of focused engineering efforts perceived to be effective by the decision-makers along with a strong evidence of scientific veneer affecting the AI adoption decisions positively – the respondents recommended algorithm adoption more when the opinion was in favor of AI. Although a clear evidence that reading counterfactual op-ed convinced the decision-makers of a less probable likelihood of lawsuits and PR problems matched Cowgill’s findings, this analysis proved that reading the same op-ed discouraged AI adoption, which opposes the results of the original work. Another contradictory finding was about influence of reading fatalistic op-ed on lawsuits likeliness. According to Cowgill and their co-workers, the two events were positively correlated, my analysis proved it otherwise.

Limitations

Some limitations associated with a broad understanding of algorithmic activism, others arose from technical differences in two analyses and level of expertise, in particular. Firstly, Cowgill et al. (2020) employed Stata statistical software, which is considerably different than R programming language I used for this analysis. Secondly, the authors of The Managerial Effects… collected the data via a survey having access to the original dataset. I, however, acquired the dataset from the Open ICPSR data repository. A very close study of the paper and data convinced me that the latter was slightly re-shaped. Lastly, I have only explored two questions out of five and not omitted a few term interactions, potentially missing on additional trends in the data. All of the above could have led to differences in findings.

Additionally, I have noticed that some variables were omitted from the original analysis without explanation, and some variables, being present in the code, were not included in the write-up. This introduced a major challenge, limiting reproducibility of the original work.

Two business cases of hiring and lending explored potential bias against women and ethnic minorities. And although women represented half of the respondents, ethnic diversity was majorly undermined as over 80% of the decision-makers were white. This undoubtedly has influenced the results of the analysis. Another limitation stemmed from the lack of literature review in both analyses.

Conclusion

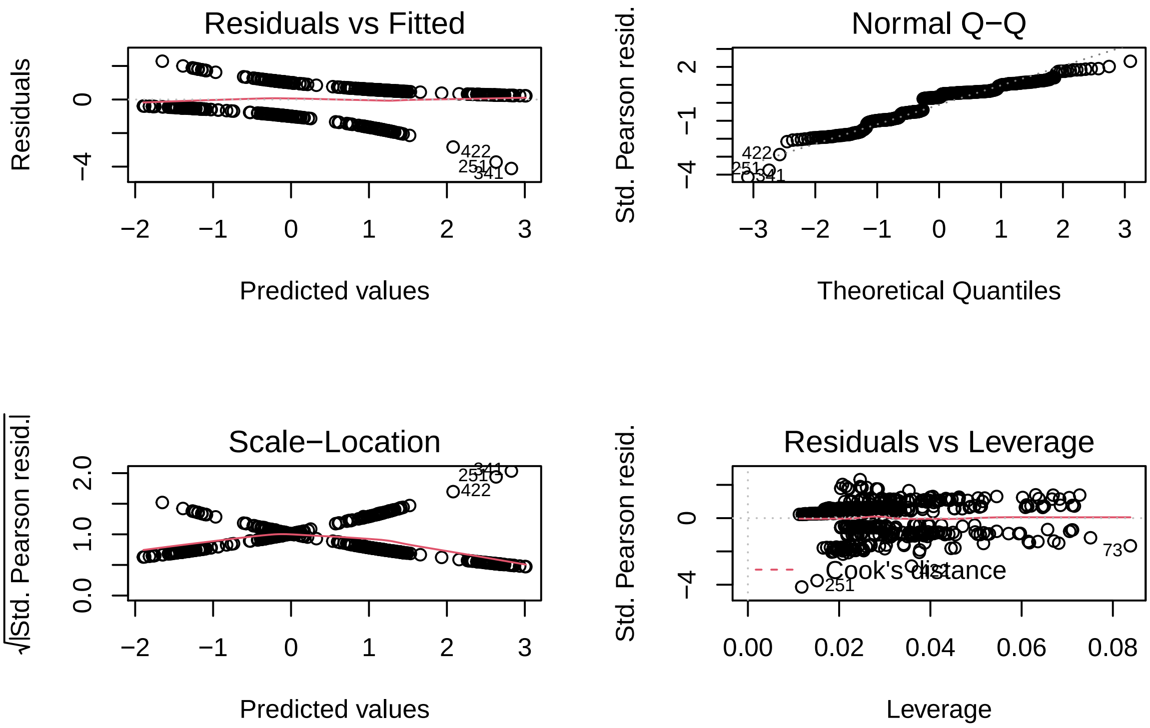
Discussions on algorithmic decision making and ethics of AI adoption have been on an ever-increasing trend in recent years. I attempted to replicate one of the available analysis of a survey conducted by Cowgill, Dell’Acqua, and Matz in 2020. Their analysis tackled the issue of AI adoption by managers and various factors influencing their decision. Despite some challenges in picking predictor variables and data discrepancy, I have achieved tangible results in replication as most of my findings mirrored the original work. This success encouraged me to research on the topic of algorithmic decision making as this field provides practical solutions and insights into the field of AI and machine learning. Some of the important questions can be answered via combination of qualitative and quantitative analysis such as The Managerial Effects of Algorithmic Fairness Activism. Since technology plays an integral part in our lives, society and the industry experts require broader and deeper understanding of algorithmic bias and ways to learn to deal with the challenge.

References

- Cowgill, B., Dell’Acqua, F., and Matz, S. (2020). The Managerial Effects of Algorithmic Fairness Activism. https://www.aeaweb.org/articles?id=10.1257/pandp.20201035
- Datasets. (2020). https://www.openicpsr.org/openicpsr/project/120752/version/V1/view?path=/openicpsr/120752/fcr:versions/V1/README&type=file
- McFadden, D. (1974). Conditional logit analysis of qualitative choice behavior. Pp. 105-142 in P. Zarembka (ed.), Frontiers in Econometrics. Academic Press. http://eml.berkeley.edu/~mcfadden/travel.html
- McFadden, D. (1977). Quantitative Methods for Analyzing Travel Behaviour on Individuals: Some Recent Developments. Cowles Foundation for Research in Economics at Yale University.

Appendix A

Figure 7. Regression assumptions check



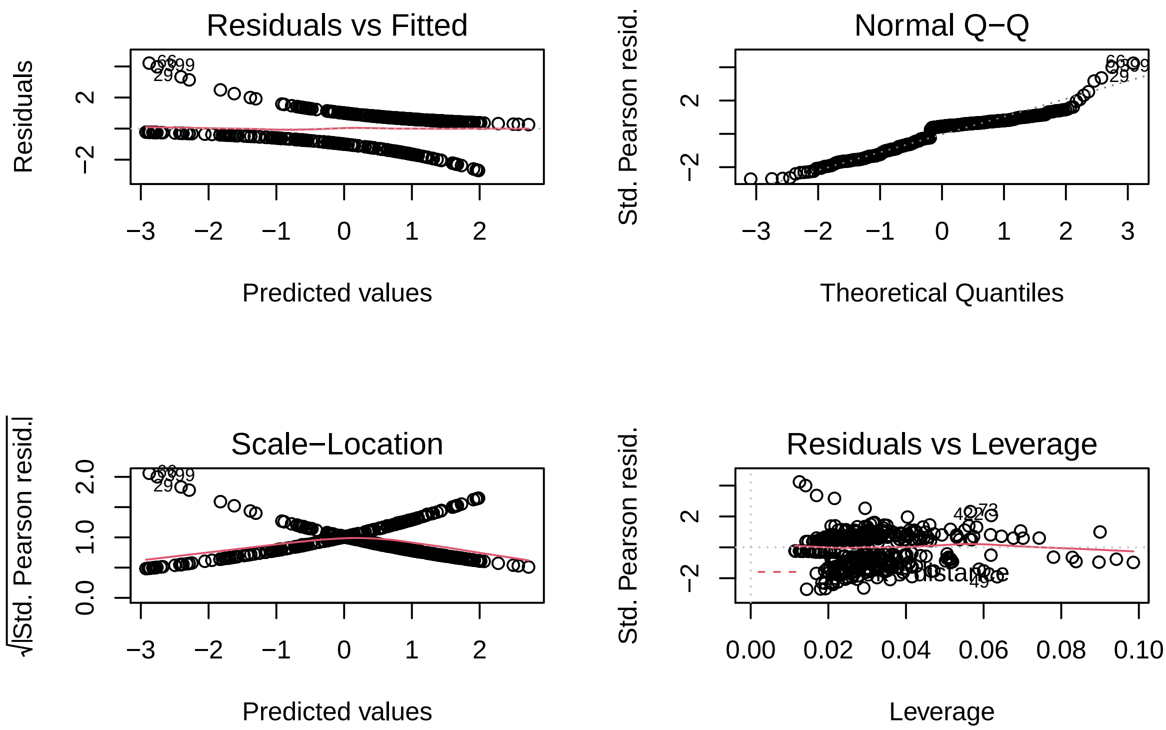


*Figure7. Regression assumptions check: linearity, normality, homoscedasticity, and outliers for business case 1 and 2. Plot 1 - Residuals vs Fitted - demonstrates **linearity**, which holds considerably well as the red line closely follows the dotted line. **Homoscedasticity**: The Scale-Location plot revealed that the points are not perfectly equally spread around the red line, meaning that the assumption of constant variance does not hold too well. **Normality**: The QQplot shows that the model is normal because the points are for the most part follow the dotted line. **Outliers** The plot 4 - Residuals vs Leverage - reveals no influential points as the Cook’s distance red dashed line does not show.*

Appendix B

Checking for model fit - Log Regr Mod St2 - Two cases - Recommendation Yes No

```
## fitting null model for pseudo-r2

##  McFadden
## 0.2351929

##            hiringfirst                       ma                    proai
##               1.066933                 3.642421                 2.998604
##                  mapro                    sqshw                    sqopt
##               5.437141                 1.500120                 1.477208
##                 female                    black                    asian
##               1.103094                 1.033413                 1.032892
##             consrv_std            workdecisions              eduprepared
##               1.076230                 1.092437                 1.057284
##                knowsml    ma:recyesno_pre_first proai:recyesno_pre_first
##               1.130756                 3.479922                 3.022232
## mapro:recyesno_pre_first
##               5.180402

## fitting null model for pseudo-r2

##  McFadden
## 0.1056443

##             hiringfirst                        ma                     proai
##                1.040930                  2.478618                  5.316652
##                   mapro                     sqshw                     sqopt
##                8.906326                  1.341429                  1.343548
##                  female                     black                     asian
##                1.104480                  1.027830                  1.030891
##              consrv_std             workdecisions               eduprepared
##                1.050737                  1.104708                  1.053654
##                 knowsml    ma:recyesno_pre_second proai:recyesno_pre_second
##                1.138094                  3.998782                  5.467135
## mapro:recyesno_pre_second
##               10.943911
```

Appendix C

St 1 estimated binomial regression model is:

$$\log\left(\frac{\hat{p}}{1-\hat{p}}\right) = -0.17 - 0.16 \cdot \text{BusinessCaseOrder} + 1.27 \cdot \text{StatusQuoShown}+$$
$$1.15 \cdot \text{StatusQuoShown(only if clicked)}-$$
$$0.06 \cdot \text{Female} - 0.36 \cdot \text{Black} - 0.49 \cdot \text{Asian} - 0.01 \cdot \text{PolitConserv}-$$
$$1.15 \cdot \text{CF} - 1.24 \cdot \text{FAT}+$$
$$2.68 \cdot \text{CFxFix} + 1.35 \cdot \text{FATxFix}+$$
$$0.15 \cdot \text{WorkDecD} + 0.23 \cdot \text{Ed} - 0.17 \cdot \text{ML}$$

St 2 estimated binomial regression model is:

$$\log\left(\frac{\hat{p}}{1-\hat{p}}\right) = -1.78 + 0.19 \cdot \text{BusinessCaseOrder} + 1.02 \cdot \text{StatusQuoShown}+$$
$$1.27 \cdot \text{StatusQuoShown(only if clicked)}+$$
$$0.04 \cdot \text{Female} - 0.51 \cdot \text{Black} - 0.66 \cdot \text{Asian} + 0.31 \cdot \text{PolitConserv}-$$
$$0.55 \cdot \text{SciVeneer} + 1.22 \cdot \text{ProAI} + 1.38 \cdot \text{SciVeneerXProAI}+$$
$$0.97 \cdot \text{SciVeneerxFix} + 1.35 \cdot \text{ProAIxFix} - 1.79 \cdot \text{SciVeneerXProAIXFix}-$$
$$0.23 \cdot \text{WorkDecD} + 0.16 \cdot \text{Ed} + 0.28 \cdot \text{ML}$$

Appendix D

Linear regression assumption check on lawsuits/PR problems likelihood question for case 1 and 2, St2.