Mirror, Mirror on the Wall: How Machine-Generated User Profiles Influence News Consumption Patterns and Beyond

YUXUAN LI*, Department of Computer Science and Technology, Tsinghua University, China
MINGDUO ZHAO*, Department of Economics, Department of Electrical Engineering and Computer Science, University of California, Berkeley, United States
COYE CHESHIRE, School of Information, University of California, Berkeley, United States

Advanced machine learning algorithms and the intricacies of the human brain often emerge as parallels, functioning as enigmatic black boxes that provoke both curiosity and caution. While discourse frequently revolves around our perceptions of machine learning, the manner in which these algorithms interpret human behavior, especially in a human-understandable format, remains underexplored. This study [1] delves into this realm by employing a large language model [2] to generate descriptive, human-interpretable user profiles based on individuals' news preferences. Participants were divided into two groups: a treatment group, which received user profiles after expressing their news preferences, and a control group, which received no profiles. We argue that these generated profiles act as reflective "mirrors" of user behavior. Subsequently, participants' news consumption patterns were observed in a second round to discern behavioral changes post-profile revelation. By juxtaposing data from both rounds and comparing the two groups, we evaluated changes in various news consumption metrics.

CCS Concepts: • **Human-centered computing → Empirical studies in HCI**.

Additional Key Words and Phrases: Behavior Change, Empirical study that tells us about people

ACM Reference Format:
Yuxuan Li, Mingduo Zhao, and Coye Cheshire. 2024. Mirror, Mirror on the Wall: How Machine-Generated User Profiles Influence News Consumption Patterns and Beyond. In *CHI '24: ACM CHI Conference on Human Factors in Computing Systems, May 11–16, 2024, Hawaii, USA*. ACM, New York, NY, USA, 15 pages. https://doi.org/XXXXXXX.XXXXXXX

1 INTRODUCTION

In the contemporary digital ecosystem, the interplay between algorithms, artificial intelligence (AI), and human behavior has become the cornerstone of our society[20]. We are increasingly living in a world where the stories we encounter, the opinions we form, and the beliefs we hold are subtly but undeniably influenced by intelligent systems that cater content to our perceived preferences[18]. As algorithms rise to prominence, shaping the contours of our digital experience, it becomes imperative to interrogate the broader socio-cultural, psychological, and political ramifications of this AI-driven paradigm[9].

*Both authors contributed equally to the paper.
[1]The experiment was reviewed and determined to be exempt by IRB under protocol ID 2023-08-16595 and pre-registered at AsPredicted (#142189, https://aspredicted.org/32D_66M).
[2]To be specific, we employ GPT-4, the current state-of-the-art model, to conduct our study.

Manuscript submitted to ACM

Advanced machine learning algorithms and the intricacies of the human brain often emerge as parallels, functioning as enigmatic black boxes that provoke both curiosity and caution. Common discussions and debates tend to focus on our comprehension and assessment of these machine learning algorithms[26]. Yet, a relatively uncharted domain remains: how do these advanced algorithms perceive and interpret the vast spectrum of human behavior, especially when presented in formats that are comprehensible to the average individual[27][7]? The central thrust of our study is an exploration of this very paradigm.

We utilize a large language model (LLM) to not only explore its capabilities, but to translate individual news preferences into descriptive user profiles and scores that can be easily interpreted by humans. Think of these generated profiles as mirrors—reflective instruments that provide users with a glimpse into their own behavioral nuances and preferences[12]. The primary objective was to determine the extent to which these mirrors, when held up to individuals, could influence or alter their subsequent news consumption habits.

For the purpose of a structured and comprehensive analysis, participants in our study were divided into two distinct groups. The treatment group were provided with machine-generated user profiles. In contrast, the control group navigated the study without any feedback on their news preferences. In the second round of the study, we observed the news consumption patterns of these participants, particularly noting any behavioral shifts that might have arisen after viewing their profiles. By analyzing data acquired from both rounds and instituting a comparison between the behavioral patterns of the treatment and control groups, we are able to explore potential shifts in news consumption[16]. Our evaluative lens was not restricted to mere consumption patterns; we also assessed a range of metrics including breadth of interests, reliance on mainstream media, cultral interests, political inclination, and so on.

Our purpose is not just an exploration of machine learning capabilities but an invitation to reflect on the interplay between human behavior and algorithmic outputs[31]. Through our study, we hope to shine a light on the profound impact that machine-generated user feedback can have on individual choices, thus broadening the discourse on the relationship between humans and the increasingly intelligent digital ecosystems we inhabit.

Section 2 provides an overview of relevant literature. Section 3 describes the specifics of the experimental design and explains the utilization of the LLM to generate profiles and valid scores. Section 4 presents and elaborates on the study's findings. Section 5 discusses the broader implications of these findings. Lastly, Section 6 offers a conclusion and reflects on potential enhancements for the study.

2 RELATED WORK

The study of algorithms, artificial intelligence (AI), and human behavior coalesces into a vibrant interdisciplinary domain, pulling threads from computer science, economics, psychology, sociology, ethics, and media studies among others. The widespread integration of machine learning technologies into various societal sectors calls for a review of existing literature to provide the context for this investigation. Our review is meticulously structured along the following thematic areas: (1) Algorithms and Human Behavior, (2) AI Interpretations of Human Preferences, (3) Ethical Considerations in AI-Human Interactions, (4) News Consumption Habits, (5) Methodologies for Studying Algorithmic Impact, and (6) Echo Chambers and Algorithmic Reinforcement.

2.1 Algorithms and Human Behavior

The symbiotic relationship between algorithms and human behavior has attracted substantial scholarly attention. Sunstein's pivotal work on "nudging" [29] serves as a cornerstone for understanding the ways subtle alterations in algorithmic operations can significantly influence human decision-making processes. Similarly, Pariser's description of

"filter bubbles" [23] explores the propensity for algorithms to generate informational echo chambers that contribute to ideological polarization. These theoretical foundations have been expanded upon by contemporary scholars such as Tufekci [30], who delve into the consequential ramifications of filter bubbles for democracy and public discourse. Notably, Eslami et al. [11] provide robust empirical evidence that algorithmic structures can induce discernible shifts in social dynamics.

2.2 AI Interpretations of Human Preferences

The capacity of large language models to interpret and possibly predict human preferences has emerged as an exigent area of inquiry in HCI and related fields. Early research by Kosinski et al. [13] illuminates how ostensibly benign digital footprints could be effectively exploited to prognosticate personality traits. Following this, technological leaps exemplified by models such as GPT-3 [4] have showcased advanced text-generation capabilities, leading to discussions about their capacity for interpreting human behavior and preferences. However, critiques by scholars like Ribeiro et al. [24] raise legitimate concerns about the interpretability, reliability, and epistemological soundness of these language models, signaling an imperative for a more nuanced understanding, which serves as the central axis for our study.

2.3 Ethical Considerations in AI-Human Interactions

The ethical labyrinth that accompanies AI's burgeoning capabilities necessitates scrutiny. Crawford's influential work [5] and O'Neil's exposé [21] act as cautionary treatises that delve into the dangers of algorithmic bias and the perpetuation of systemic inequalities. Furthermore, Friedman and Nissenbaum [10] highlight ethical considerations in human-computer interaction by examining the essential principles of privacy, consent, and transparency. More recent contributions from scholars such as Bostrom [3] and Sengers et al. [25] interrogate the ethical boundaries of AI capabilities and offer frameworks for ethically-aligned computational designs. This ethical scaffolding is imperative for studies aiming to integrate AI tools for social analysis, given the expansive ethical landscape of AI-driven analytics.

2.4 News Consumption Habits

The taxonomy of news consumption has also undergone rigorous academic inspection, increasingly nuanced by the advent of algorithmic curation. Webster's classification between "news finders" and "news seekers" [32] offers a valuable behavioral taxonomy, while Bakshy et al. [2] and Flaxman et al. [8] employ quantitative methods to assess the impact of algorithmic sorting on news exposure. Theoretical frameworks, such as Allport's "contact hypothesis" [1], provide foundational understanding for the influence of diversified information exposure on individual attitudes and collective behaviors. More recently, Napoli's work [17] investigates how algorithmic curation affects the diversity of news consumed, and its subsequent impact on civic participation.

2.5 Methodologies for Studying Algorithmic Impact

Investigative methodologies for exploring the ramifications of algorithms on human behavior and society have evolved considerably. Lewis' comprehensive analysis [15] offers a survey of prevalent methods such as case studies, surveys, and experiments. Observational studies utilizing big data analytics, as instantiated by Lazer et al. [14], contribute an additional layer of complexity and granularity. The development of new research methods continues to enhance our understanding of the societal impacts of algorithms.

2.6 Echo Chambers and Algorithmic Reinforcement

The concept of echo chambers has gained prominence in discussions about algorithms' social impacts. These are spaces where algorithms continuously present users with like-minded opinions, thereby reinforcing existing beliefs and potentially increasing societal polarization. Early works by Sunstein [28] emphasize the democratic implications of such environments. Subsequent studies, such as those by Del Vicario et al. [6], focused on how algorithms amplify these echo chambers by tailoring content to maximize user engagement. Recent research suggests the necessity for algorithmic interventions to disrupt these chambers and promote cognitive diversity [19]. Understanding echo chambers is pivotal for our current study, as they present both a challenge and an opportunity in algorithmic content curation.

The aforementioned themes collectively provide an integrated academic backdrop against which our study is positioned, allowing for a balanced interplay between theoretical conceptualization, empirical validation, and ethical contemplation.

3 EXPERIMENT AND DATA PROCESSING

In an endeavor to discern the potential impact of presenting individuals with a "mirror" reflecting their preferences, we created an experimental study. Participants were recruited via Proflic[22], a platform well-regarded for research studies. Our sample consists of 442 participants, with an average age of 40. When it comes to gender, 40% identified as female, while 3% identified as non-binary or preferred not to disclose. The median of participants hold a Bachelor's degree as their highest educational qualification. The median income range reported by participants falls between "$50,000 to $74,999". In terms of racial distribution, 16.2% as identified African American, 8.8% as Asian, 3.4% as others, 61.5% as white (ordered alphabetically by race name).

Upon recruitment, these individuals were navigated to a specially curated website, intentionally designed to emulate the interface of Google News. This is our attempt to make our experiment environment as usable and intuitive as possible allowing participants to concentrate on news selection. This approach also tries to ensure that our experimental setup closely mimic real-life news selection scenarios, making our results mostly comparable to observational studies. This experiment was methodically segmented into two distinct phases within each round:

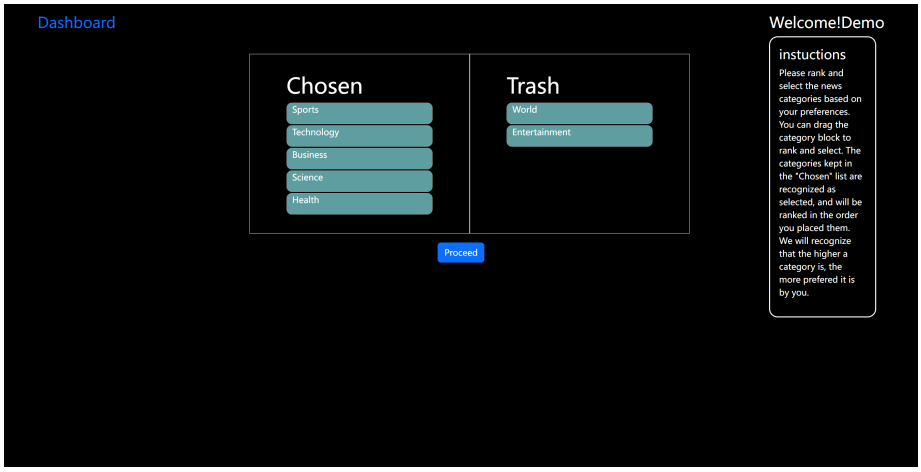

Fig. 1. Page for users to choose and rank news categories.

Step 1: Topic Selection

Participants were presented with an array of topics or news categories, as visualized in Figure 1. Accompanying these topics were instructions situated on the right side of the interface. Participants were tasked with ranking these topics based on their levels of interest. In cases where certain topics were deemed irrelevant or unappealing, participants had the liberty to relocate these topics to a designated "Trash" column. The categorization of these topics mirrored the organizational structure employed by Google News.

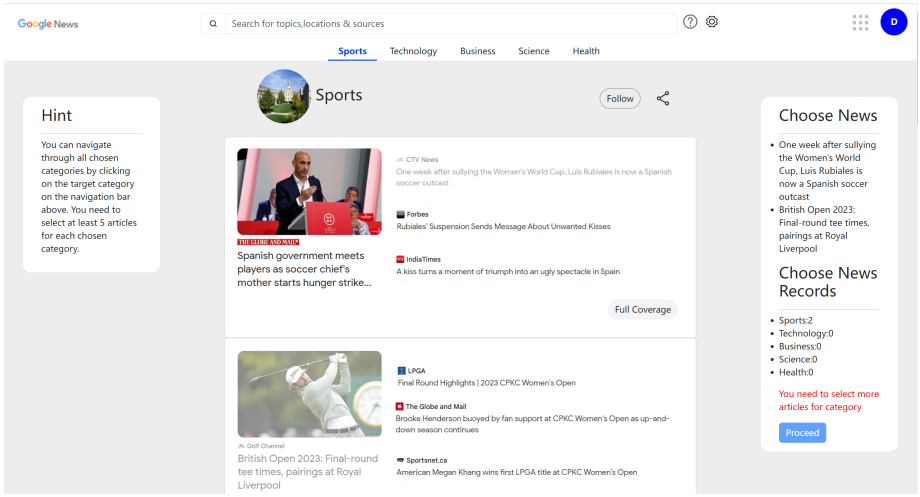

Fig. 2. Page for users to choose news. Existence and order of the categories on the navigation bar aligns with the user's previous choosing and ranking.

Step 2: Simulated News Selection

Subsequent to their topic selection, participants were ushered to an interface that was designed to resemble the conventional Google news selection page (Figure 2). For every category selected in the step 1, topics were arrayed from left to right on the tab below the searching bar, reflecting the order of preference indicated by the participant. Within each topic, participants were permitted to select up to five news headlines that resonated with their interests. Upon selecting five headlines from every chosen topic, participants concluded this step. To bolster the authenticity of this experimental environment, the news headlines and their associated media outlets were extracted directly from the actual Google News platform. However, a crucial distinction exists: instead of being routed to the actual news articles after clicking, participants were engaging in a simulated environment. Essentially, their selections were solely based on news headlines and the associated media outlets.

Our study is structured into two distinct phases, termed "rounds." Each round encapsulates the two-step process described earlier: the Topic Selection (Step 1) and the Simulated News Selection (Step 2). Upon the completion of round 1, participants in the treatment group are presented with a profile, as illustrated in Table 1 (definitions for the variables listed in Table 3). This profile offers a human-interpretable reflection, curated by the LLM, which provides insight into their choices from the inaugural round. In addition to offering a "mirror" into their choices, the LLM explains the reasoning behind its conclusions. The validity and precision with which the model learns characteristics based on round 1 selections will be supported in the later part of this paper. After completing round 1 and being exposed to

Variable	Description	Reason
Breadth of Interests	*"You have broad interests"*	*"You read news from all 7 categories, ranking Technology and Health highest."*
Reliance on Mainstream Media	*"You rely heavily on mainstream news"*	*"Most of your chosen articles are from large mainstream publications."*
Cultural Interest	*"You lean towards Western culture"*	*"Your chosen articles focus more on Western than Eastern culture and events."*
Political Ideology	*"You have moderate political views"*	*"Your chosen articles come from centrist news sources, not far left or right ones."*

Table 1. Mock User Profile

either the treatment or placebo information during the intermediary phase, participants will proceed to round 2 to make news selections again.

Scale (1-5)	Questions
1-5	Government is almost always wasteful and inefficient
1-5	Government regulation of business usually does more harm than good
1-5	Poor people today have it easy because they can get government benefits without doing anything in return
1-5	The government today can't afford to do much more to help the needy
1-5	Blacks who can't get ahead in this country are mostly responsible for their own condition
1-5	Immigrants today are a burden on our country because they take our jobs, housing, and health care
1-5	The best way to ensure peace is through military strength
1-5	Most corporations make a fair and reasonable amount of profit
1-5	Stricter environmental laws and regulations cost too many jobs and hurt the economy
1-5	Homosexuality should be discouraged by society

Table 2. Political Survey: Ten 1-5 Scale Questions

A part of our study framework that warrants attention pertains to a political survey, grounded in the methodology of Pew Research, which is designed to gauge participant's political inclinations. The specific questions encompassed in the survey can be found in Table 2. Respondents who strongly agree with a given statement receive a score of 5, while those who vehemently disagree are awarded a score of 1. Scores for those with moderate views will fall within this range, depending on the intensity of their agreement or disagreement. Upon completing all questions, an aggregate score is computed for each participant by taking sum. Individuals who attain 50 points exhibit strong right-leaning tendencies, while those who score 10 points lean prominently to the left. Scores in between these values denote varying degrees of political inclination, with higher scores suggesting a greater rightward leaning. The primary objective of this survey is to quickly determine participants' political orientations by tapping into questions and topics that are frequently debated and elicit differing viewpoints from the two major political parties. This survey is taken twice: before round 1, and again after round 2.

To summarize: participants initially engage with this political survey, subsequently transitioning to round 1. Subsequently, those in the treatment group are given the profile constructed by the LLM, while their counterparts in the control group receive intermediary, non-informative content (placebo). This phase, bridging rounds 1 and 2, serves as a juncture to potentially influence subsequent choices for people in the treatment group. Then people in all groups make their second round choice. After round 2, participants revisit the political survey, offering an opportunity to gauge any potential changes in their political standings.

Variable	Definition and Scoring
Breadth of Interests	Measures how many different categories of news a user is interested in. Score range: 0 (least breadth) to 100 (most breadth), measured by news consumption.
Reliance on Mainstream Media	Captures how much a person relies on mainstream news sources. Score range: 0 (least reliance) to 100 (most reliance), measured by news consumption.
Cultural Interest	Describes interest in Eastern or Western culture. Score range: -100 (most interested in Eastern culture) to 100 (most interested in Western culture), measured by news consumption.
Political Survey	Describes how left or right-leaning someone is, using 10 scale questions from Pew Research. Score range: 10 (most left-leaning) to 50 (most right-leaning), measured by political survey.
Political Ideology	Describes how left or right-leaning the news they consume are. Score range: -100 (most left-leaning) to 100 (most right-leaning), measured by news consumption.

Table 3. Definition and Scoring of User Engagement Variables

Table 3 shows the outcome variables employed in our analysis. Central to our examination is the exploration of variations in these variables contingent on participants' exposure to the user profile generated by the LLM. This is differentiated by whether participants belong to the control group (who aren't exposed to the profile) or the treatment group (who view the profile). A particular part of our variable definitions demands special emphasis: the distinction between "Political Survey" and "Political Ideology." The former is derived from participants' direct engagement with a survey that assesses their political inclinations by Pew. In contrast, "Political Ideology" is an inference made by the LLM, drawing upon participants' news consumption patterns within a given round. Among our suite of outcome variables, "Political Survey" is the only variable that is not extrapolated by the LLM.

In order to authenticate the scores offered by the LLM, we map the relationship between "Political Survey" and "Political Ideology" in a scatter plot (see Figure 3). Additionally, a fitted line is superimposed on this scatter plot to emphasize trends and correlations. An inspection of the scatter plot shows a positive, significant slope, revealing that the scores generated by the LLM align with the results from the direct political survey. A comprehensive validation of the other outcome variables, including their congruence and reliability across rounds, will be elaborated upon in the results section.

4 RESULTS

Our primary research objective is to determine how viewing a machine-generated, easily interpretable user profile affects subsequent news choices. From both the news consumption patterns and the political survey, we evaluate five dimensions: the range of topics individuals show interest in, their preference for Western versus Eastern culture, their reliance on mainstream media, their political stance (from survey), and the political slant of the news they consume (from news choices).

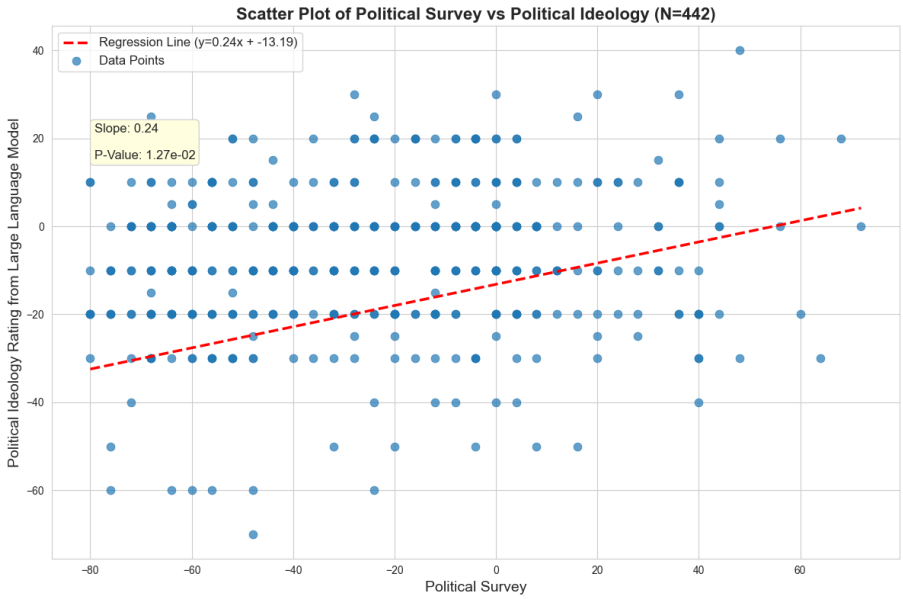

Fig. 3. Scatter plot illustrating the relationship between Mapped Political Survey scores (original scores transformed to the range [-100, 100]) and Political Ideology scores for 442 participants. Each point represents an individual's Political Ideology score plotted against their corresponding Mapped Political Survey score. A dashed red line indicates the linear regression fit.

Variable	Average	Standard Deviation
Breadth of Interests	76.42	23.79
Cultural Interest	57.53	24.3
Reliance on Mainstream Media	86.14	7.34
Political Survey	23.95	8.4
Political Ideology	-7.42	17.08

Table 4. Descriptive Statistics in the First Round (N=442)

The data illustrated in Table 4 provides an overview of the summary statistics derived from evaluations conducted via the large language model, juxtaposed with results from our political survey. These statistics offer a window into the characteristics of our sample cohort. Our results reveal that participants, on average, exhibit a left-leaning political inclination. For clarity, a score of 30 in our political survey signifies a neutral stance, drawing inspiration from methodologies utilized by the Pew survey. This left-leaning proclivity is further substantiated by the political ideology metric, where participants received an average score of -7.42, on a spectrum ranging from -100 (extremely left-leaning) to 100 (extremely right-leaning), with 0 symbolizing a neutral viewpoint. This is further evidence to validate the score generated by the large language model. However, it's essential to contextualize this observed bias. The standard deviation captured within our dataset suggests that while there is a discernible leftward tilt, this inclination is moderate, given that the entirety of the sample falls within a range of just 1 standard deviation from the mean. This

indicates that the political beliefs of the participants, though leaning left, are relatively close to a centrist viewpoint when considering the broader scale of possible beliefs.

Variable	Treatment	Control	p-value (T=C)
Breadth of Interests	76.69	76.21	0.8348
Reliance on Mainstream Media	86.28	86.03	0.7233
Cultural Interest	56.97	57.98	0.6679
Political Survey	23.91	23.98	0.9339
Political Ideology	-8.67	-6.44	0.1737

Table 5. Average Values and T-test for the Treatment Group and Control Group in the First Round

Presented in Table 5 is a balance table that underscores the validity of our treatment and control group allocations. An examination of all five outcome variables reveals an absence of significant disparities prior to the employment of the treatment during the first round.

	Outcomes				
	(1) Breadth of Interests	(2) Reliance on Mainstream Media	(3) Cultural Interest	(4) Political Survey	(5) Political Ideology
β_1	0.563***	0.141***	0.165***	1.001***	0.107**
	(0.047)	(0.047)	(0.05)	(0.013)	(0.047)
β_2	4.597**	2.283***	0.264	0.164	-1.179
	(2.254)	(0.694)	(2.447)	(0.220)	(1.624)
Observations	442	442	442	442	442
R^2	0.252	0.044	0.024	0.931	0.013
Adjusted R^2	0.249	0.040	0.020	0.931	0.009

$^{*}p$ <0.1; $^{**}p$ <0.05; $^{***}p$ <0.01

Table 6. Regression Results for the 5 Variables.

Table 6 delineates the regression outcomes associated with the variables under examination. To comprehensively understand the dynamics, we employed the following regression model:

$$y_{i2} = \beta_1 y_{i1} + \beta_2 D_i + \beta_0 + \epsilon_i$$

where y_{i2} represents the score assigned to the outcome variable during the second evaluation phase; y_{i1} denotes the score ascribed during the initial evaluation; D_i serves as an indicator function, assuming a value of 1 for participants in the treated group and 0 for those in the control group. The outcome variables we have are "Breadth of interest", "Reliance on Mainstream Media", "Cultural interest", "Political Survey", "Political Ideology". Once again, with the exception of "Political Survey" which is anchored in the Pew Research framework, all other outcome variables were scored leveraging insights derived from the LLM.

The p-values indicate the estimation of β_1 are all highly statistically significant. Such a finding illustrates the persistent consistency exhibited by participants across the various evaluation rounds (Table 6). If the scores, as computed by

the large language model, were fraught with instability or excessive noise, one would expect a notable inconsistency in evaluations across different rounds, given that these evaluations are conducted in isolation. The observed tight associations, therefore, serves not just as an affirmation of participant behavior constancy but also as a testament to the precision and reliability of the scoring methodology employed by the large language model.

Following the validation of the score presented, in this section, we delve into a granular analysis of the results, placing particular emphasis on β_2 which represents the average treatment effect (ATE) of the profile on individuals. Upon inspection, exposure to these profiles appears to enhance an individual's breadth of interest. This effect is statistically significant, indicating that individuals diversify their news consumption after recognizing specific topic preferences. Such behavior exemplifies a manifestation of the "love of variety" principle. Interestingly, despite this diversification in topics, there's a notable shift towards mainstream media, post-exposure to the profile. This shift raises questions about the possible reasons for such an inclination. One possibility is that it could involve trust. To be specific, in the context of our study, after realizing their inherent biases or leanings (by learning their preferences and inclinations), participants might have sought validation from these more universally accepted sources. As for cultural interests, statistical analyses do not suggest meaningful alterations in an individual's proclivity for news from various cultures. The p-value hovers perilously close to 1, indicating an under-powered changes in tastes after the profile treatment.

When examining political inclinations, our survey reflects a negligible shift (0.164) in political beliefs, post-exposure. This finding is anticipated since a brief interaction with a user profile is unlikely to wield significant influence over deeply entrenched political beliefs, beliefs that have been molded over years of individual experiences and growth. However, a more pronounced treatment effect emerges when considering the political ideology of the news sources consumed. While not achieving statistical significance, a trend emerges where post-profile exposure, participants exhibited a propensity for more left-leaning content. This deeper dive into the results highlights not only the direct effects of the profile on news consumption patterns but also presents intriguing questions about the underlying reasons and mechanisms driving these observed behaviors.

	Outcome: Breadth of Interests		
	(1) Left	(2) Neutral	(3) Right
β_1	0.550^{***}	0.562^{***}	0.578^{***}
	(0.092)	(0.08)	(0.076)
β_2	6.891^{*}	5.340	1.369
	(3.824)	(3.824)	(4.054)
Observations	141	159	142
R^2	0.214	0.250	0.296
Adjusted R^2	0.202	0.240	0.285

$^{*}p$ <0.1; $^{**}p$ <0.05; $^{***}p$ <0.01

Table 7. Regression Results of Breadth of Interests for 3 Political Groups.

In subsequent sections, our analysis shifts its focus to uncover the heterogeneous treatment effect (HTE) across different outcome variables. This entails categorizing participants into three distinct cohorts based on their preliminary

political survey outcomes: left-leaning, right-leaning, and neutral. Table 7 delves into the outcome variable related to the breadth of interest. A compelling observation emerges: individuals identifying with left-leaning political beliefs display a more pronounced enhancement in the breadth of interest. Specifically, their score escalates by 6.89, with an associated p-value of approximately 7.4%. Meanwhile, those with neutral stances observe a moderate surge, positioning them second in terms of the magnitude of change. Conversely, participants with a more Republican orientation exhibit the most subdued growth in their scope of interests. It's pivotal to emphasize that these results don't underscore political dynamics. Instead, they illuminate variations in the spectrum of topics participants found appealing post-treatment. Essentially, it delves into the elasticity of one's topics interested when confronted with algorithmically-generated insights into their preferences.

	Outcome: Reliance on Mainstream Media		
	(1) Left	(2) Neutral	(3) Right
β_1	0.277^{***}	0.068	0.105
	(0.079)	(0.084)	(0.08)
β_2	3.006^{**}	1.524	2.458^{*}
	(1.109)	(1.257)	(1.215)
Observations	141	159	142
R^2	0.125	0.014	0.041
Adjusted R^2	0.112	0.001	0.027

$^{*}p$ <0.1; $^{**}p$ <0.05; $^{***}p$ <0.01

Table 8. Regression Results of Reliance on Mainstream Media for 3 Political Groups.

	Outcome: Cultural Interest		
	(1) Left	(2) Neutral	(3) Right
β_1	0.140	0.252^{***}	0.054
	(0.085)	(0.081)	(0.097)
β_2	1.466	-1.507	1.836
	(4.427)	(4.051)	(4.304)
Observations	141	159	142
R^2	0.020	0.061	0.004
Adjusted R^2	0.006	0.049	-0.011

$^{*}p$ <0.1; $^{**}p$ <0.05; $^{***}p$ <0.01

Table 9. Regression Results of Cultural Interest for 3 Political Groups.

In our examination of participants' dependence on mainstream media, distinct patterns emerge based on political inclinations (Table 8). Those aligning with Democratic perspectives exhibit a pronounced uptick in their reliance on mainstream outlets. The score for this group rises by 3.01, a statistically significant shift, as evidenced by a p-value less than 1%. Contrarily, both neutral and right-leaning respondents demonstrate a modest and less consistent increase in their trust in mainstream media sources. When evaluating the metric of cultural interest, the findings appear less definitive across all three political categories (Table 9). The data presents a higher variance, implying that individual differences might play a larger role, rendering the influence of the machine learning "mirror" less discernible. Considering the substantial p-values (all surpassing 70%), it is reasonable to infer that a transient machine-generated profiles do not markedly influence participants' cultural interests. In essence, while the machine learning "mirror" appears to sway individuals' reliance on mainstream media, especially among those with Democratic tendencies, its influence on broader cultural interests remains ambiguous across the board.

	Outcome: Political Survey		
	(1) Left	(2) Neutral	(3) Right
β_1	1.022***	1.079***	1.081***
	(0.05)	(0.064)	(0.054)
β_2	0.032	0.090	0.380
	(0.23)	(0.373)	(0.496)
Observations	141	159	142
R^2	0.753	0.650	0.746
Adjusted R^2	0.750	0.645	0.743

$^{*}p$ <0.1; $^{**}p$ <0.05; $^{***}p$ <0.01

Table 10. Political Survey for 3 Political Groups.

When analyzing political inclinations derived from the direct political survey, a uniform pattern emerges across all three groups: there's virtually no observable shift in political beliefs (Table 10). As previously noted, such an outcome is expected. A transient engagement with a user profile is improbable to considerably impact political stances that have been forged and refined over the course of one's life. These views represent a culmination of myriad experiences, teachings, and personal reflections, which are not easily swayed by fleeting interactions.

On the other hand, when we evaluate the political inclinations ascertained by the LLM - determined based on news consumption patterns and the inherent political ideology of the chosen news - the findings become more nuanced, albeit not statistically significant at the 5% level (Table 11). For those identifying with left and neutral political views, a leftward shift is observed post "mirror" intervention, suggesting a reflection and reinforcement of choices made in the initial round. In contrast, right-leaning participants exhibit an intensified rightward tilt following the treatment. While it might be tempting to amalgamate the results, the HTE analysis, which separates respondents based on their initial political inclinations, offers invaluable insights. It underscores the self-reinforcing nature of news consumption: when presented with a machine-reflected profile of their prior choices, individuals tend to gravitate further in the direction

	Outcome: Political Ideology		
	(1) Left	(2) Neutral	(3) Right
β_1	0.122	-0.001	0.127
	(0.084)	(0.081)	(0.081)
β_2	-1.926	-3.298	2.067
	(3.055)	(2.543)	(2.806)
Observations	141	159	142
R^2	0.020	0.011	0.021
Adjusted R^2	0.006	-0.002	0.007

$^{*}p$ <0.1; $^{**}p$ <0.05; $^{***}p$ <0.01

Table 11. Regression Results of Political Ideology for 3 Political Groups.

they originally leaned towards. This observation hints at the potential amplification and entrenchment of existing beliefs when users are continually mirrored their past preferences.

5 IMPLICATION

The findings from this study have several implications. Among the most prominent observations is the notion that echo chambers or the cycle of self-reinforcing beliefs may not be entirely involuntary. The idea that simply providing external information can automatically lead individuals to appreciate diversity may be overly optimistic. There is a possibility that even after individuals consume vast amounts of politically biased content and the platform provides an alert or feedback, they might persist in their established news consumption patterns. Over time, such habits could further entrench their political views, exacerbating polarization.

In light of the study's conclusions, we propose the implementation of a diversified recommender system that offers perspectives from a broader range of sources without people's explicit notice. This recommendation doesn't stem merely from the "exploration" principle in algorithms but from a larger societal standpoint. While this approach may not always align with a company's profit-driven motives, there's an overarching responsibility to consider the larger societal implications. Governments, in this scenario, can intervene to ensure that the societal ramifications of consuming polarized news don't undermine communal unity and impose huge negative externality. This calls for a strategic balance between user preferences and the broader interests of society.

6 CONCLUSION AND DISCUSSION

In our exploration into the implications of AI-driven user profiling and its potential influence on news consumption habits, we have explored complicated domains spanning news consumption, recommender systems, information treatment, and even possible policy implications. Here are our major conclusions:

First of all, our experimental design, which incorporated both distinct rounds of interaction and AI profiling, highlighted several pivotal insights. It was evident that human behavior, when mirrored back via an AI model through human interpretable information, does not necessarily induce expected cognitive shifts towards diversifying content

consumption especially for the political contents. Although people tend to consume contents from more diverse topics, echo chambers and self-reinforcing beliefs, often considered unconscious phenomena, might, in fact, be far more consciously reinforced than previously assumed in the political domain. Interestingly, when politics is not a factor, viewing the profile does seem to expand individuals' interest in a broader range of topics. It's intriguing to explore the reasons behind this divergence.

Secondly, our study spotlighted the congruence between scores attributed by the LLM and the outcomes of the political survey. When examining news consumption, this convergence underscored the LLM's adeptness and potential at scanning through large amounts of data and extracting meaningful patterns that reflect a user's tendencies or inclinations. This suggests that advanced machine learning models like the LLM can, with considerable accuracy, predict certain user characteristics based on seemingly unrelated behaviors or preferences. The alignment further cements the potential utility of such models in a variety of applications.

Lastly, when we shift our focus to the broader policy implications, our study's outcomes might highlight a necessity for well-thought-out interventions. While the advancements in technology and the digital realm promise a more interconnected world, they also inadvertently create spaces where biases and singular perspectives thrive. Our results underscore the fact that merely heightening transparency, by presenting users with their profiles, may not be an adequate measure to dismantle these echo chambers. The trajectory depicted by our findings paints a concerning picture. If users only fortify their existing beliefs even after being presented with reflective insights into their news consumption patterns, it indicates a potential for continuous or even accelerated polarization.

For the discussion part, we aim to highlight potential areas of exploration and delve deeper into some emerging concepts. While this is by no means a comprehensive list, we wish to outline two pivotal ideas we believe are critical to future work. To begin with, the current study's methodology leans towards a more static approach, predominantly focusing on one singular intervention. However, we recognize the dynamic nature of user interaction with digital platforms. To capture this evolving landscape more accurately, we are in the process of initiating a follow-up study. In this ongoing research, we intend to allow participants to engage with a continual stream of content curated by the recommender system. The goal is to monitor and analyze the changes in their user profiles in real-time, as they navigate and consume this content. People will also receive live feedback from the system. This will provide an understanding of how users adapt and respond to a dynamically generated content environment. Secondly, we ponder the applicability of our findings in the broader realm of social media platforms. While our research is rooted in the context of Google News, which doesn't inherently push a particular viewpoint, it's crucial to ascertain the universality of our results. Social media platforms have diverse user bases and multi-faceted content. Even though we maintain a high degree of confidence in the external validity of our findings, we believe that a replication of our study in a social media setting would further cement our conclusions and offer richer insights into the interplay between users and content recommendation algorithms.

At the end of the day, as AI and machine learning algorithms become integral in shaping our media landscapes, it's imperative that we approach their influence with nuance, understanding, and responsibility. The interplay between AI recommendations, user profiles, and real-world behaviors is intricate, and our study is only a beginning. As societies grapple with the challenges of polarization, robust, evidence-based interventions from the government or companies themselves will be paramount, ensuring that the nexus of technology and media serves to inform, enlighten, and unify, rather than mislead and divide.

REFERENCES

[1] Gordon Willard Allport, Kenneth Clark, and Thomas Pettigrew. 1954. The nature of prejudice. (1954).

[2] Eytan Bakshy, Solomon Messing, and Lada A Adamic. 2015. Exposure to ideologically diverse news and opinion on Facebook. *Science* 348, 6239 (2015), 1130–1132.

[3] Nick Boström. 2014. Superintelligence: Paths, dangers, strategies. *Superintelligence: Paths, dangers, strategies* (2014).

[4] Tom B. Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel M. Ziegler, Jeffrey Wu, Clemens Winter, Christopher Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. 2020. Language Models are Few-Shot Learners. arXiv:2005.14165 [cs.CL]

[5] Kate Crawford. [n. d.]. Artificial intelligence's white guy problem. ([n. d.]).

[6] Michela Del Vicario, Alessandro Bessi, Fabiana Zollo, Fabio Petroni, Antonio Scala, Guido Caldarelli, H Eugene Stanley, and Walter Quattrociocchi. 2016. The spreading of misinformation online. *Proceedings of the national academy of Sciences* 113, 3 (2016), 554–559.

[7] Paul Dourish. 2001. *Where the action is: the foundations of embodied interaction.* MIT press.

[8] Seth Flaxman, Sharad Goel, and Justin M Rao. 2016. Filter bubbles, echo chambers, and online news consumption. *Public opinion quarterly* 80, S1 (2016), 298–320.

[9] Brian J Fogg. 2002. Persuasive technology: using computers to change what we think and do. *Ubiquity* 2002, December (2002), 2.

[10] Batya Friedman and Helen Nissenbaum. 1996. Bias in computer systems. *ACM Transactions on information systems (TOIS)* 14, 3 (1996), 330–347.

[11] Kevin Hamilton, Motahhare Eslami, Amirhossein Aleyasen, Karrie Karahalios, and Christian Sandvig. 2015. FeedVis: A Path for Exploring News Feed Curation Algorithms. https://doi.org/10.1145/2685553.2702690

[12] Marc Hassenzahl. 2008. User experience (UX) towards an experiential perspective on product quality. In *Proceedings of the 20th Conference on l'Interaction Homme-Machine*. 11–15.

[13] Michal Kosinski, David Stillwell, and Thore Graepel. 2013. Private traits and attributes are predictable from digital records of human behavior. *Proceedings of the National Academy of Sciences of the United States of America* 110 (03 2013). https://doi.org/10.1073/pnas.1218772110

[14] David Lazer, Alex Pentland, Lada Adamic, Sinan Aral, Albert-László Barabási, Devon Brewer, Nicholas Christakis, Noshir Contractor, James Fowler, Myron Gutmann, et al. 2009. Computational social science. *Science* 323, 5915 (2009), 721–723.

[15] Seth C. Lewis. 2015. Journalism In An Era Of Big Data. *Digital Journalism* 3, 3 (2015), 321–330. https://doi.org/10.1080/21670811.2014.976399

[16] John McCarthy and Peter Wright. 2004. Technology as experience. *interactions* 11, 5 (2004), 42–43.

[17] Philip Napoli. 2019. *Social media and the public interest: Media regulation in the disinformation age.* Columbia University Press.

[18] Clifford Nass, Jonathan Steuer, and Ellen R Tauber. 1994. Computers are social actors. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 72–78.

[19] C Thi Nguyen. 2020. Echo chambers and epistemic bubbles. *Episteme* 17, 2 (2020), 141–161.

[20] Donald A Norman. 2002. The Design of Everyday Things, reprint paperback ed.

[21] Cathy O'neil. 2017. *Weapons of math destruction: How big data increases inequality and threatens democracy.* Crown.

[22] Stefan Palan and Christian Schitter. 2018. Prolific. ac—A subject pool for online experiments. *Journal of Behavioral and Experimental Finance* 17 (2018), 22–27.

[23] Eli Pariser, Amy Goodman, and Juan Gonzalez. 2023. The Filter Bubble: What the Internet Is Hiding from You. (09 2023).

[24] Marco Tulio Ribeiro, Tongshuang Wu, Carlos Guestrin, and Sameer Singh. 2020. Beyond Accuracy: Behavioral Testing of NLP Models with CheckList. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. Association for Computational Linguistics, Online, 4902–4912. https://doi.org/10.18653/v1/2020.acl-main.442

[25] Phoebe Sengers, Kirsten Boehner, Shay David, and Joseph'Jofish' Kaye. 2005. Reflective design. In *Proceedings of the 4th decennial conference on Critical computing: between sense and sensibility*. 49–58.

[26] Ben Shneiderman. 2003. *Leonardo's laptop: human needs and the new computing technologies.* Mit Press.

[27] Lucille Alice Suchman. 1987. *Plans and situated actions: The problem of human-machine communication.* Cambridge university press.

[28] Cass Sunstein. 2003. R. 2007. Republic. com 2.0.

[29] Richard Thaler and C. Sunstein. 2009. *NUDGE: Improving Decisions About Health, Wealth, and Happiness.* Vol. 47.

[30] Zeynep Tufekci. 2015. Algorithmic Harms beyond Facebook and Google: Emergent Challenges of Computational Agency. Colorado Technology Law Journal 13.2 (2015), 203–218.

[31] Sherry Turkle. [n. d.]. alone together why we expect more from technology and less from each other. pdf. ([n. d.]).

[32] James G. Webster. 2014. *The Marketplace of Attention: How Audiences Take Shape in a Digital Age.* The MIT Press. https://doi.org/10.7551/mitpress/9892.001.0001