Journal of
Theoretical and Applied Electronic Commerce Research

MDPI

Article

Cutting Edge or Crystal Clear? How Does the Supervisibility of Algorithm Throughput Influence the Coping Behaviors of Targeted Advertising Audiences?

Lijun Chen [1,*], Rui Sun [2,*], Lele Fan [1] and Lei Zhuang [2]

[1] School of Business Administration, Jimei University, Xiamen 361021, China; skylarkfan@163.com
[2] School of Business Administration, Huaqiao University, Quanzhou 362021, China; hquzl@hqu.edu.cn
* Correspondence: lijunchen@jmu.edu.cn (L.C.); sunrui@hqu.edu.cn (R.S.)

Abstract

The increasing complexity of artificial intelligence systems and algorithms can negatively affect individual attitudes and prompt them to adopt coping behaviors. Based on the technology threat avoidance theory, this study explores the mechanisms through which the supervisibility of recommendation algorithm throughput influences the coping behaviors of targeted advertising audiences using scenario-based online experiments conducted between August 2022 and January 2023. In particular, we examine the mediating effect of perceived threat and the moderating effects of privacy concerns and self-efficacy. The results reveal that perceived threat arising from the characteristics of recommendation algorithm throughput determines the possible coping behaviors of targeted advertising audiences. Moreover, perceived threat mediates the impact of the supervisibility of algorithm throughput on audiences' coping behaviors. Furthermore, users' privacy concerns and self-efficacy positively moderate the influence of the supervisibility of algorithm throughput on perceived threat and the relationship between perceived threat and audiences' coping behaviors, respectively. The results help to explain the black box of the relationship between the characteristics of recommendation algorithms and audience responses from an algorithm throughput perspective. The findings offer implications for improving the effectiveness of personalized recommendations and provide a new perspective for studying the transparency of algorithmic recommendations in consumer behavior.

Keywords: recommendation algorithm; throughput type; algorithm supervisibility; targeted advertising; coping behaviors

Check for updates

Academic Editors: Joonheui Bae and Min Jung Cho

Received: 7 April 2026
Revised: 28 May 2026
Accepted: 28 May 2026
Published: 30 May 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

The advancement and extensive application of big data technologies and artificial intelligence (AI) across various domains have consistently been accompanied by critical discourse on the potential emergence of autonomous consciousness. This persistent concern revolves around the possibility that these systems may evade oversight and become uncontrollable. Recent developments in autonomous AI systems have enabled machines to learn patterns from data without explicit human labeling [1]. Public apprehension about algorithmic controllability has been amplified by such systems, as research shows that people increasingly question whether these systems remain under meaningful human supervision [2]. This concern reflects the public's ongoing unease about the controllability of AI and algorithmic technologies.

In marketing, targeted advertising, a type of online advertising that focuses on consumers' characteristics, interests, and preferences, relies on large amounts of user data and algorithmic models to deliver specific online advertisements based on consumer behaviors [3]. Although offering such precise recommendations reduces search costs and improves recommendation effectiveness, it also creates certain problems, including information cocoons, price discrimination, and algorithmic manipulation [4]. Excessively precise recommendations not only make consumers feel that the algorithm knows them better than they know themselves but also raise concerns that the algorithm may manipulate users or even pose a threat [5,6]. When faced with recommendations, people question the basis and reasons behind them [7], leading to negative behaviors such as resisting and avoiding recommendations which have attracted the attention of scholars [8].

Studies that adopt privacy computing theory [9] state that consumers' coping behaviors are mainly based on privacy protection considerations such as privacy concerns and risks [5,6,10–13]. With the continuous development and iteration of algorithmic technologies and their increasingly ubiquitous use in recommendations, audiences are concerned not only about privacy leakage but also about the broader manipulation of human life by algorithms [14–17]. Such concerns have recently intensified with the rapid development of AI and the proliferation of large language models such as ChatGPT. Audiences' coping behaviors resulting from such concerns ultimately draw their attention to the controllability of new technology. However, most research on consumers' coping behaviors focuses on endogenous factors that influence their attitudes and behaviors. Few studies have examined exogenous factors such as the characteristics of the technology itself. Furthermore, discussions on algorithm models have focused on the technical level, such as optimizing algorithm models, and few studies have jointly examined algorithm features and audience responses. Given that an algorithm is a data-processing procedure based on massive data, the "supervisibility" of its processing not only reflects its complexity and high-level nature but also determines whether its output is interpretable and transparent, which ultimately affects audience attitudes substantially.

To bridge the aforementioned gaps, this study explores the mechanisms underlying the relationship between the characteristics of recommendation algorithms and the coping behaviors of targeted advertising audiences (hereafter referred to as audiences' coping behaviors) by referring to the technology threat avoidance theory (TTAT). To collect data, we adopt two experimental scenarios based on the characteristics of different recommendation algorithms (supervised vs. unsupervised). By examining audiences' coping behaviors in these two real-world scenarios, as well as the mediating effect of perceived threat and the moderating effects of privacy concerns and self-efficacy, this study clarifies the relationship between the supervisibility of recommendation algorithms and audiences' coping behaviors, thereby adding to the literature on technology threat and consumer responses.

2. Literature Review and Hypotheses Development

2.1. Supervisibility: Conceptualization and Differentiation

To understand how algorithm design properties affect user coping responses, we introduce the construct of *supervisibility*. We define *supervisibility* as:

Definition 1. *The user's perceived extent to which the internal data processing (throughput) of a recommendation algorithm is designed to be auditable and adjustable by human agents, specifically reflecting whether the algorithm operates under pre-defined human rules (supervised) or autonomously discovers patterns without explicit human labeling (unsupervised).*

This definition contains three essential features. First, supervisibility is a perceptual construct—it captures how users interpret algorithmic design, not objective technical features. Second, it focuses on the throughput phase (how input data are transformed into recommendations), not on data collection or output presentation. Third, its core distinction is between supervised vs. unsupervised learning architectures as communicated to users, which signals the degree of human oversight over rule-setting.

2.1.1. Distinguishing Supervisibility from Transparency, Explainability, Interpretability, and Controllability

Supervisibility is related to but conceptually distinct from several existing constructs. Transparency refers to the degree to which an algorithm's logic is disclosed or made visible to users [18]. An algorithm can be transparent (e.g., providing a detailed technical white paper) yet still be perceived as low in supervisibility if it uses unsupervised deep learning that no human can fully supervise. Explainability concerns the ability to provide post hoc justifications for a specific algorithmic output [19], whereas supervisibility concerns the ex-ante architecture of rule determination. Interpretability refers to the degree to which a human can understand a model's decision process [20]; supervisibility does not require full understanding but rather the perception that human supervision is possible. Controllability refers to a user's direct ability to modify algorithmic outputs (e.g., through settings), while supervisibility refers to perceived human oversight over the learning process itself. Table 1 summarizes these distinctions.

Table 1. Distinguishing supervisibility from adjacent constructs.

Construct	Core Question	Focus	Key References
Transparency	Is the algorithm's logic disclosed to the user?	Information disclosure	[18,21]
Explainability	Can the system justify a specific output?	Post hoc reasoning	[19]
Interpretability	To what extent can a human understand the model's decision process?	Model comprehensibility	[20]
Controllability	Can the user directly modify the output or settings?	User agency	[2]
Supervisibility	Does the user perceive that human-defined rules (vs. autonomous detection) govern the throughput?	Human oversight of learning architecture	This study

2.1.2. Theoretical Rationale for Supervisibility

Drawing on the Stimulus–Organism–Response (SOR) framework [22], we conceptualize algorithm design features as stimuli that influence users' cognitive and emotional states (the organism), such as perceived threat, which in turn shape responses (coping behaviors). Among various algorithm design properties, supervisibility is particularly relevant because it directly signals whether the system remains under human control. When users perceive low supervisibility (e.g., an unsupervised algorithm that learns autonomously), they perceive the system as less controllable, heightening perceived threat. According to Technology Threat Avoidance Theory (TTAT) [23], perceived controllability is a central driver of threat appraisal and subsequent avoidance-oriented coping. Supervisibility directly shapes this perceived controllability, thereby providing a theoretically grounded explanation for how algorithmic design properties influence user coping responses.

2.1.3. Rationale for the Supervised vs. Unsupervised Manipulation

We posit that non-expert users naturally infer human involvement from descriptions of supervised learning (e.g., "rules set by developers") and machine autonomy from descriptions of unsupervised learning (e.g., "machine learns rules on its own"). Therefore, our supervised condition is designed to operationalize high supervisibility and our unsupervised condition to operationalize low supervisibility. We validate this operationalization through a manipulation check, which confirms that participants indeed perceive the two conditions as significantly different in terms of supervisibility.

2.2. The Technology Threat Avoidance Theory as an Explanatory Framework: Construct Mapping and Boundary Conditions

To explain how the supervisibility of recommendation algorithms influences users' coping responses, we draw on the Technology Threat Avoidance Theory (TTAT). TTAT was originally developed to understand how individual computer users respond to security threats such as malware and viruses [23]. Its core proposition is that when individuals perceive a technology-related threat, they engage in two cognitive appraisals: threat appraisal (assessing the severity of the threat and their susceptibility to it) and coping appraisal (evaluating the effectiveness, cost, and their ability to perform protective measures). These appraisals collectively shape avoidance motivation, which in turn drives coping behaviors.

While TTAT originated in the domain of information security, its underlying cognitive appraisal mechanism is generalizable to other technology-induced threats that involve perceived loss of control. For example, prior research has successfully extended TTAT to privacy concerns in online behavioral advertising, where perceived threat mediates the relationship between privacy concerns and advertising avoidance [11]. In our context, the threat is not a security violation per se, but rather a perceived loss of control and autonomy caused by an opaque, autonomous recommendation algorithm. Specifically, when an algorithm operates with low supervisibility (e.g., unsupervised learning), users may feel that the system's decision rules are beyond human oversight, triggering a threat appraisal: "This algorithm might manipulate me without accountability." This appraisal then motivates coping responses—either problem-focused (taking actions to regain control) or emotion-focused (venting or avoiding).

Given that TTAT's appraisal mechanism extends to threats of lost control, we now map its core constructs onto the specific context of algorithmic supervisibility. In the original theory, threat refers to the likelihood of a security breach (e.g., a virus infection); in our extension, it becomes the perceived loss of control due to an autonomous algorithm. Threat appraisal originally involves assessment of severity and susceptibility; here, it captures the perception that a low-supervisibility algorithm poses a controllability threat. Coping appraisal originally concerns the perceived effectiveness of antivirus actions; in our context, it focuses on users' perceived ability to avoid or manipulate recommendations (e.g., deleting browsing history, switching platforms). Finally, avoidance behavior originally meant quitting the system or changing usage patterns; in our study, it manifests as problem-focused coping (e.g., adjusting privacy settings) and emotion-focused coping (e.g., expressing dissatisfaction). This mapping demonstrates that TTAT offers a coherent theoretical lens for understanding how supervisibility affects user responses.

Our extension preserves TTAT's core assumption that threat appraisal precedes coping, but we also acknowledge two important boundary conditions that distinguish the algorithmic context from the original security domain. First, TTAT typically assumes that users possess some degree of control over the threat object (e.g., they can install antivirus software). In our context, however, low-supervisibility algorithms may be perceived as uncontrollable by design and this very lack of control constitutes the threat rather than mitigating it. Second, TTAT primarily addresses episodic threats, such as detecting a specific virus or malware

attack. Algorithmic transparency concerns, by contrast, are continuous and ambient: users experience them repeatedly over time during routine interactions with recommendation systems. These differences do not invalidate TTAT's applicability; rather, they specify the scope of our theoretical extension and suggest promising directions for future research, such as longitudinal investigations of coping dynamics. By anchoring our investigation in TTAT, we provide a theoretically grounded explanation of how perceived controllability—signaled through algorithm supervisibility—evokes threat appraisals and subsequent coping behaviors. This framework directly supports our hypotheses concerning the mediating role of perceived threat and the moderating effects of privacy concerns and self-efficacy.

2.3. Supervisibility of Algorithm Throughput and Audiences' Coping Behaviors

People are curious about why and how algorithmic decisions are made [24]. This decision process involves data throughput within an algorithmic system. Data throughput refers to the collation of the input data into a specific structure through supervised or unsupervised machine learning techniques. Supervisibility refers to the extent to which and the ease with which a system can be monitored, reviewed, intervened with, and guided by humans. Supervised machine learning is carried out under preset rules, whereas in unsupervised machine learning, the machine automatically learns to form a set of rules that are difficult for people to clearly understand and applies them to produce future recommendations. The complexity of the technology involved in this stage makes it difficult for users to understand or explain; therefore, it is also known as the algorithmic black box [25]. The technical details of supervised and unsupervised algorithms have been extensively discussed in the literature [26–28]; importantly, supervised algorithms can embed human biases [26,27], while unsupervised algorithms operate autonomously and are even harder to control, as their rules are not fully known even to developers.

The high-order recommendation algorithms used in targeted advertising are personalized recommendation services based on data analysis [28]. Their self-learning and decision-making abilities allow these systems to analyze data and complete the push, respectively. The computing power of autonomous algorithms can process large amounts of data and far exceeds that of the human brain. To ensure the controllability of algorithms and reduce the threat to users' privacy, developers often ensure that algorithms are supervised using a series of technical settings. However, this does not align with developers' pursuit of value [28]. Moreover, as network data become increasingly large (with unstructured and unlabeled data increasing disproportionately relative to labeled data), there is an increasing interest in using unlabeled data through unsupervised algorithms to optimize network performance [29]. This interest stems from the need to free supervised algorithms from their restrictive requirements [1]. Further, it is expensive to manage large amounts of labeled network data, manual annotation is inconvenient, and the highly dynamic nature of computer networks means that such data can rapidly become obsolete [30].

As the complexity of AI systems and algorithms rises, they are increasingly regarded as incomprehensible black boxes because understanding their decisions and performance require increasing levels of professional knowledge. This can affect audience attitudes, prompting them to engage in certain coping behaviors. Coping, a psychological concept, describes the cognitive strategies and behaviors adopted by individuals in the face of threats to reduce the resulting pressure. In consumer behavior, the coping style defined by Lazarus and Folkman [31] has been widely cited as the adaptive cognitive or behavioral effort made by individuals to reduce the pressure of internal and external demands.

In this field, TTAT describes how users form avoidance motives and behaviors when encountering technological threats. Its basic premise is that when faced with technological threats, if users think they can avoid the threat by taking protective measures, they may take

active or passive measures by adopting problem- and emotion-focused coping behaviors, respectively. As unsupervised algorithms are more difficult to understand and control than supervised algorithms, targeted advertising audiences become more passive, making it more likely for them to show emotion-focused behaviors. Therefore, by combining the characteristics of recommendation algorithm throughput with TTAT, this study proposes the following hypotheses:

H1: *The supervisibility of the recommendation algorithm throughput significantly impacts audiences' coping behaviors.*

H1a: *Supervised recommendation algorithms lead targeted advertising audiences to adopt problem-focused coping behaviors.*

H1b: *Unsupervised recommendation algorithms lead targeted advertising audiences to adopt emotion-focused coping behaviors.*

2.4. Mediating Effect of Perceived Threat

The increasing integration of algorithm technologies in daily life makes it crucial to recognize the role of algorithms [32]. Transparency requires that recommendation decisions made through algorithm throughput are clear and understandable to users [32]. However, this is often difficult to achieve. For example, Netflix uses an algorithm to recommend content based on users' viewing and search histories. However, it still faces criticism because it is difficult to assess the transparency of the recommendation decisions; in other words, visibility is not always the same as transparency. Regardless of whether they are supervised or unsupervised, the recommendation algorithm throughput process adopted by advertisers contains certain technical barriers for the audience. Even if the principles behind algorithm recommendation and the data throughput process are published to the audience to improve transparency, it may not yield the desired results because of differences in individuals' algorithm literacy. Indeed, as algorithms and models become more advanced and complex, audience attitudes become more ambiguous. People are increasingly concerned about the control of these algorithms over human interactions [32], their reinforcement of bias [33], and the distortion of existing biases and prejudices [34]. Studies show that highly transparent information strengthens users' perceptions of danger, leading them to perceive greater threats [35], and triggers corresponding coping behaviors. The higher the threat perceived by consumers, the more likely they are to adopt negative behaviors and attitudes [36]. Similarly, the more difficult the process is to understand and control, the stronger the negative attitudes and behaviors. Based on the above, the following hypotheses are proposed:

H2: *Compared with a supervised algorithm, an unsupervised algorithm is more likely to cause targeted advertising audiences to perceive threats.*

H3: *Perceived threat positively impacts audiences' coping behaviors.*

H4: *Perceived threat mediates the relationship between the supervisibility of recommendation algorithm throughput and audiences' coping behaviors.*

2.5. Moderating Effects of Privacy Concerns and Self-Efficacy

Privacy concerns are common issues for consumers and businesses. In the Internet era, with the continuous evolution of data mining and recommendation technologies, personal information faces a greater risk of leakage. Privacy concerns are Internet users' apprehen-

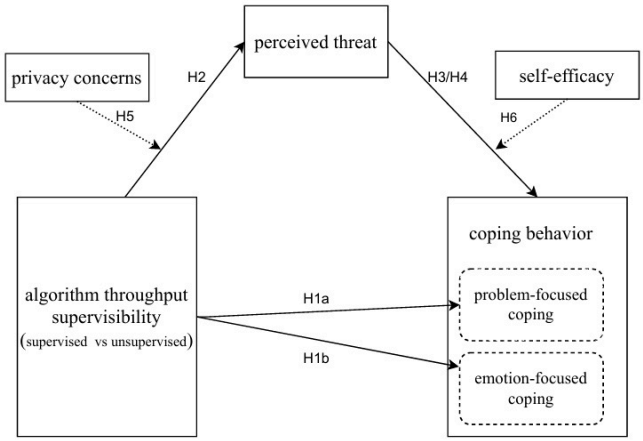
sions about the collection and use of their personal information by websites [37]. They are affected by a variety of factors, while simultaneously impacting consumers' attitudes and behaviors in different ways. For example, privacy concerns weaken consumers' perceptions of fairness and affect their risk beliefs [38], leading to more privacy-protective behaviors. With the continuous development of algorithm technologies, consumer data always face the threat of being leaked. Advanced algorithms can quickly and accurately restore an individual's complete digital portrait through in-depth mining, data aggregation, and cross-comparison across different types of data. The White Paper on the Protection of User Rights of Telecom and Internet Users, issued by the China Academy of Information and Communications Technology (CAICT), points out that "high-precision algorithms restore the identity attribute of a large number of anonymous data that have been removed, and then associate it with specific individuals, weakening the effect of anonymity technology" [39]. However, as algorithm models become more advanced, deeper data mining occurs and individuals pay greater attention to privacy issues, resulting in stronger perceived threats. Based on the above, the following hypothesis is proposed:

H5: *Privacy concerns moderate the impact of the supervisibility of the recommendation algorithm throughput on perceived threat.*

The concept of self-efficacy was first proposed by Bandura [40]. According to Bandura's self-cognitive theory, the personal factors of cognition, emotion, and biology as well as behavior and the environment are determinants of a person's behavior, and self-efficacy is a key factor in this process [41]. An individual's self-efficacy significantly impacts their behavior; people with higher self-efficacy are often more willing to take positive actions. In the current context, self-efficacy is reflected in an individual's belief in their ability to deal with privacy risks. Self-efficacy has a significant impact on individuals' intention to keep their information safe [42–44]; further, users with high self-efficacy take countermeasures faster and more adaptively when dealing with technical threats [45]. Based on the above, the following hypothesis is proposed:

H6: *Self-efficacy moderates the impact of perceived threat on audiences' coping behaviors.*

Figure 1 illustrates the study's conceptual model based on the aforementioned hypotheses.



Notes: Solid arrows represent hypothesized direct paths. Dashed arrows represent moderating effects. H1 (overall effect of algorithm supervisibility on coping behaviors) is decomposed into H1a (supervised → problem-focused coping) and H1b (unsupervised → emotion-focused coping). H4 represents the indirect effect of algorithm type on coping behaviors via perceived threat, sharing the same arrow with H3 (direct effect of perceived threat on coping). H2, H5, and H6 are as shown.

Figure 1. Conceptual model.

3. Methods

3.1. Sample

Data were collected from August 2022 to January 2023 using the Credamo platform. A total of 298 valid responses were obtained. The sample consisted of 202 female (67.8%) and 96 male (32.2%) participants. Ages ranged from 18 to 58 years (M = 30.98, SD = 8.11). Regarding education, 209 participants (70.1%) held a bachelor's degree, 42 (14.1%) held a postgraduate degree or higher, and 47 (15.8%) had a vocational school degree or other qualifications. Geographically, participants came from 31 provincial-level regions across China, with the largest proportions from Guangdong (19.5%), Shandong (9.4%), Heilongjiang (9.4%), Jiangsu (8.4%), and Anhui (5.4%). Table 2 provides a detailed summary of the demographic characteristics.

Table 2. Demographic characteristics of the sample (N = 298).

Characteristic	Category	*n*	%
Gender	Female	202	67.8
	Male	96	32.2
Age (years)	18–30	162	54.4
	31–40	88	29.5
	41–58	48	16.1
Education	Junior high school	1	0.3
	Vocational/high school	15	5.0
	Associate degree	33	11.1
	Bachelor's degree	209	70.1
	Master's degree	37	12.4
	Doctorate degree	2	0.7
	Missing/unclear	1	0.3
Geographic region	Shandong	28	9.4
	Guangdong	58	19.5
	Jiangsu	25	8.4
	Heilongjiang	28	9.4
	Hebei	15	5.0
	Anhui	16	5.4
	Zhejiang	12	4.0
	Fujian	9	3.0
	Beijing	8	2.7
	Shanghai	7	2.3
	Other provinces/regions	92	30.9

Note: In the education variable, vocational/technical high school was combined as "Vocational/high school"; associate degree is listed separately. For geographic region, only provinces/regions with frequencies ≥ 7 are shown individually; "Others" includes the remaining provinces/regions (covering all 31 provincial-level administrative divisions in China).

Although the public discourse on generative AI has intensified since late 2022, our theoretical focus is on the fundamental distinction between supervised and unsupervised algorithmic throughput—a design characteristic that remains conceptually stable regardless of recent AI advancements. The data collection period coincided with early public concerns about algorithmic controllability (e.g., debates around algorithm transparency and user manipulation), meaning the findings are still relevant for understanding consumer responses to algorithm design properties.

3.2. Experimental Design, Manipulation, and Procedure

This study adopted a single-factor between-subjects design, with the **supervisibility** of the targeted advertising recommendation algorithm as the independent variable. Data were collected through a combination of scenario-based experiments and questionnaires.

3.2.1. Pre-Test of the Manipulation

Prior to the main experiment, the effectiveness of the scenario manipulation was first validated. The materials for the two experimental scenarios were as follows:

> *"One day, you open a shopping app, search for 'plus-size clothing', click on several search results, and browse casually for a few minutes. You then exit the app and switch to other platforms. When you reopen the original shopping app, you see the recommended content as shown in Figure 2 in a prominent position"*



Figure 2. Recommended content. **Note:** The Chinese text in this figure includes common e-commerce phrases such as "buy one get three", "returnable if ineffective", "designed for weight loss", "repeat customers", and product names (e.g., "Fat flow residual").

In Scenario 1, the recommended content was accompanied by the following explanation: "This recommendation is generated based on your relevant information using a supervised, basic-level algorithm. 'Supervised' means that the algorithm's rules are pre-defined by human developers, and the process is controllable." That is, in Scenario 1, the processing of the targeted advertising recommendation algorithm is supervised.

In Scenario 2, the recommended content was accompanied by the following explanation: "This recommendation is generated based on your relevant information using an unsupervised, high-level autonomous algorithm. 'Unsupervised' means that the algorithm's rules are autonomously learned by the machine and are not fully controlled by the developers." That is, in Scenario 2, the processing of the targeted advertising recommendation algorithm is unsupervised.

After presenting the scenario, participants were asked: "*Based on the scenario you just read, is the operation of this recommendation algorithm controllable and supervised by humans?*" (1 = completely uncontrollable and unsupervised; 7 = completely controllable and supervised by humans).

An independent-samples *t*-test showed that the ratings in the supervised condition ($M = 5.48$, $SD = 1.16$) were significantly higher than those in the unsupervised condition

(M = 1.96, SD = 1.21), t (48) = 10.52, $p < 0.001$, Cohen's d = 2.98. This confirms that our manipulation was successful and that participants clearly distinguished between the two algorithmic conditions.

3.2.2. Main Experimental Procedure

In the main experiment, participants first provided demographic information and were informed of the purpose of the survey and the precautions to be taken. Next, a brief explanation of the supervised and unsupervised algorithms behind targeted advertising was given. Participants were then randomly assigned to one of the two scenarios. After fully understanding the scenario, participants were asked to answer questions about their feelings and potential coping behaviors in that situation.

3.2.3. Design Rationale for the Recommended Products

In both scenarios, participants were shown exactly the same set of recommended products, which included plus-size clothing and weight-loss related items. This product combination was generated based on participants' search histories and served two purposes.

First, holding product content constant across the two scenarios ensured that any differences in participants' responses could be attributed to the algorithm type manipulation (supervised vs. unsupervised) rather than to variation in the products themselves.

Second, the inclusion of weight-loss products was intended to create a cross-category association—that is, the algorithm's ability to link a user's search for plus-size clothing to a potential weight-loss need. Critically, this cross-category association can be interpreted in two different ways depending on the scenario. In the supervised condition, participants were told that the association was based on rules pre-defined by developers. In the unsupervised condition, participants were told that the association was autonomously learned by the machine through pattern detection. Thus, while the recommended products remained identical across conditions, the perceived generative mechanism behind the recommendation differed. This design lies at the core of our manipulation of supervisibility: the algorithm's throughput is perceived as human-supervised (rule-based) in one condition and as autonomous (self-learning) in the other.

We acknowledge that the wording "not fully controlled by the developers" inherently conveys both algorithmic autonomy and reduced human oversight. This is not a confound but rather an integral part of our manipulation, as low supervisibility is defined precisely by such reduced human control. Also, this phrasing may also prime threat perceptions beyond the mere understanding of unsupervised learning.

3.3. Variable Measurement

To ensure content validity, the items used to measure the constructs were derived from established scales from domestic and international research and adapted according to the research objectives.

Following Malhotra et al. [46], privacy concerns were measured using the three dimensions of collection, control, and awareness. The first dimension had four items such as "When the platform asks me for personal information, I often feel troubled." The second dimension had three items such as "Online privacy is actually a matter of consumers exercising control and autonomy over how their information is collected, used, and shared." Finally, the third dimension had three items such as "Companies should disclose the way they collect, process, and use data online."

Self-efficacy was measured using the General Self-Efficacy Scale translated and revised by Wang et al. [47]. This scale, which was initially developed by Schwarzer and Jerusalem [48], has been widely used globally. Wang et al. [47] verified its reliability and validity as well as its application in China. The General Self-Efficacy Scale consists of

10 items such as "If I try my best, I can always solve the problem," "Even if others oppose me, I still have a way to obtain what I want," and "For me, adhering to ideals and achieving goals is very easy." We used the General Self-Efficacy Scale because our study focuses on overall coping behaviors (both problem-focused and emotion-focused) in response to algorithmic threats, which involve not only privacy concerns but also technical and controllability dimensions. General self-efficacy captures individuals' global confidence in handling difficult situations, making it theoretically appropriate for predicting a broad range of coping responses [40].

The measurement of perceived threat was based on Liang and Xue [23], Shi [49], and Chen [50]. The degree of perceived threat of the participants in the scenario was measured using the following four items: "The information flow of this targeted advertisement poses a threat to me," "The consequences of the information flow of this targeted advertisement threatens me," "The information flow of this targeted advertisement harms my interests," and "The information flow of this targeted advertisement is dangerous to my interests."

The measurement of audiences' coping behaviors was based on Zhang et al. [5,6] and was divided into problem-focused and emotion-focused coping behaviors. The measurement items of problem-focused coping behaviors were as follows: "I take practical actions such as changing passwords, setting permissions, and logging out of personal information," "I take practical actions such as filing complaints for actions on online platforms," "I have switched to using other online shopping platforms and stopped using this one," and "I have reduced the use of this online shopping platform." The measurement items of emotion-focused coping behaviors were as follows: "I feel very depressed," "I will express my dissatisfaction," "I feel uneasy about the inability to ensure the security of my personal information," and "I feel uncomfortable." For the main analyses (mediation and moderation), we created a composite coping score by averaging all eight items, as this reflects the overall avoidance-oriented response to technological threats conceptualized by TTAT. For exploratory purposes, we also examined problem-focused and emotion-focused coping separately in H1a and H1b to provide dimension-specific insights. This operationalization is applied consistently throughout the analyses. In the following results, unless otherwise specified (e.g., when testing H1a and H1b separately), the term "coping behaviors" refers to this composite score.

All of the aforementioned items were rated on a 7-point Likert scale, representing "1 = completely disagree," "2 = disagree," "3 = relatively disagree," "4 = uncertain," "5 = relatively agree," "6 = agree," and "7 = completely agree."

4. Results

4.1. Preliminary Analysis

Using the Statistical Package for the Social Sciences (SPSS, version 26), a reliability test was conducted on the scales for privacy concerns, self-efficacy, perceived threat, and audiences' coping behaviors (problem-focused coping and emotion-focused coping). The Cronbach's α coefficients of these scales were 0.880, 0.905, 0.953, 0.949, 0.912, and 0.920, respectively. All of the values were above 0.8, indicating that the research data met the acceptable threshold and that reliability was high.

To further assess data quality, we examined potential common method bias using Harman's single-factor test [51]. An unrotated exploratory factor analysis on all measurement items showed that the first factor accounted for 36.528% of the total variance, which is below the recommended threshold of 50%. This suggests that common method bias is not a serious concern in this study.

Next, a confirmatory factor analysis was conducted to test the convergent validity and discriminant validity of the variables. As the average variance extracted (AVE)

values of privacy concerns and self-efficacy were below 0.5, four items with low factor loadings were excluded: three from the privacy concerns scale ("Companies should disclose the way they collect, process, and use data online," "A good consumer online privacy policy should have a clear and conspicuous disclosure," and "It is very important for me to know how my personal information will be used") and one from the self-efficacy scale ("If I put in the necessary effort, I can always solve most difficult problems"). These privacy concern items were deleted because they primarily measured general awareness of data practices, which, while part of the original conceptualization [46], were less central to the specific context of algorithmic supervisibility and cross-category inference in targeted advertising. The self-efficacy item was removed due to its absolute phrasing ("always solve"), which may have introduced ambiguity. After re-running the confirmatory factor analysis, the AVE values (0.536/0.509/0.804/0.700/0.813) and composite reliability values (0.887/0.902/0.943/0.903/0.946) of the five factors (privacy concerns/self-efficacy/perceived threat/problem-focused coping behaviors/emotion-focused coping behaviors) exceeded 0.5 and 0.7, respectively, indicating good convergent validity, as shown in Table 3. Further, the square roots of the AVE values of each factor (0.732/0.714/0.897/0.837/0.902) were generally above the absolute values of the correlation coefficients between the factors, indicating good discriminant validity, as shown in Table 4.

Table 3. Result of model AVE and CR indexes.

Factor	AVE	CR
PRC	0.536	0.887
SES	0.509	0.902
PTH	0.804	0.943
PFC	0.700	0.903
EFC	0.813	0.946

Table 4. Discriminant validity: Pearson correlation and AVE square root values.

	PRC	SES	PTH	PFC	
PRC	0.732				
SES	0.012	0.714			
PTH	0.469	0.118	0.897		
PFC	0.426	0.152	0.753	0.837	
EFC	0.358	0.050	0.868	0.869	0.902

Note: Diagonal numbers are AVE square root values.

Table 5 presents the overall model fit. The CFI (0.954), NFI (0.907), NNFI (0.948), and IFI (0.954) met the recommended thresholds (>0.9), indicating acceptable fit. However, several indices fell short of conventional criteria: GFI (0.858) and AGFI (0.830) were below the recommended 0.90, and RMR (0.090) exceeded the preferred value of 0.05. The χ^2 test was significant (χ^2 (340) = 643.251, $p < 0.001$), which is common in large samples (N = 298) and thus not conclusive. Taken together, the model fit is acceptable but not ideal. The suboptimal GFI and AGFI values may be attributed to model complexity and the high intercorrelations among coping behavior dimensions.

To further justify the adequacy of the measurement model, we note that the incremental fit indices (CFI, NFI, NNFI, and IFI)—which are less sensitive to sample size and model complexity—all exceed 0.90, indicating a good relative fit. The standardized RMR (SRMR = 0.052) falls below the conservative cutoff of 0.08, further supporting the acceptable fit. Moreover, lower GFI and AGFI values are commonly observed in complex models with multiple latent factors (five factors) and a large number of indicators (28 items), as

well as in samples of moderate size (N = 298). We acknowledge that these limitations do not invalidate our findings; rather, they suggest that future research could refine the measurement of coping behaviors or test alternative factor structures.

Table 5. Model fit indices.

Common Index	χ^2	df	*p*	χ^2/df	GFI	RMSEA	RMR	CFI	NFI	NNFI
Criterion	-	-	>0.05	<3	>0.9	<0.10	<0.05	>0.9	>0.9	>0.9
Value	643.251	340	0.000	1.892	0.858	0.055	0.090	0.954	0.907	0.948
Other indices	TLI	AGFI	IFI	PGFI	PNFI	SRMR	RMSEA 90% CI			
Criterion	>0.9	>0.9	>0.9	>0.9	>0.9	<0.1	-			
Value	0.948	0.830	0.954	0.718	0.816	0.052	0.048~0.061			

Notes: GFI = goodness-of-it index, RMSEA = root mean square error of approximation, RMR = root mean square residual, CFI = comparative fit index, NFI = normed fit index, NNFI = non-normed fit index, TLI = Tucker–Lewis index, AGFI = adjusted goodness-of-fit index, IFI = incremental fit index, PGFI = parsimonious goodness-of-fit index, PNFI = parsimonious normed fit index, SRMR = standardized root mean square residual. All reported fit indices are based on the measurement model.

4.2. Descriptive Statistics

Pearson's correlation coefficients were used to analyze the correlations between the items and total scores to verify the homogeneity between the items and scales. Table 6 shows that the correlation coefficients of all the variables were significant, providing preliminary support for subsequent hypotheses testing. In particular, the strong positive correlations between perceived threat and the two coping dimensions (r = 0.753 and 0.868) are consistent with the core premise of TTAT. However, these very high correlations also raise concerns about potential conceptual overlap between perceived threat and coping behaviors. The shared variance between perceived threat and problem-focused coping is approximately 57% (r^2 = 0.567), and with emotion-focused coping it is approximately 75% (r^2 = 0.753). Such high overlap suggests that the constructs may not be fully distinct, which could affect discriminant validity. While these correlations are theoretically expected under TTAT (where threat appraisal is a strong driver of coping), the magnitude invites caution. Future research could refine the measurement of these constructs, for example by using behavioral indicators or temporal separation, to better establish their distinctiveness while preserving their theoretical relationship.

Table 6. Correlation coefficients among the variables (Pearson's correlations).

	MEAN	SD	Perceived Threat	Privacy Concerns	Self-Efficacy	Problem-Focused Coping Behaviors	Emotion-Focused Coping Behaviors
Perceived threat	4.057	1.407	1				
Privacy concerns	5.558	0.983	0.469 ***	1			
Self-efficacy	5.313	0.893	0.118 *	0.012	1		
Problem-focused coping behaviors	4.107	1.303	0.753 ***	0.426 ***	0.152 **	1	
Emotion-focused coping behaviors	4.165	1.534	0.868 ***	0.358 ***	0.050	0.869 ***	1

Note: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

4.3. Hypotheses Testing

4.3.1. Impact of the Supervisibility of the Recommendation Algorithm Throughput on Audiences' Coping Behavior

The hypotheses were tested by treating problem- and emotion-focused coping behaviors as the dependent variables. To test H1, we conducted a regression analysis of the

supervisibility of the recommendation algorithm throughput (0 = unsupervised; 1 = supervised) on audiences' coping behaviors. The results showed that the R^2 value of the model was 0.409, indicating that the supervisibility of the recommendation algorithm throughput explained 40.9% of the variation in audiences' coping behaviors. Additionally, the regression coefficient of the supervisibility of the recommendation algorithm throughput was −1.752 (t = −14.326, $p < 0.01$), indicating that it had a significant negative impact on audiences' coping behaviors. Thus, H1 was supported.

The aforementioned impact was further tested by conducting paired-samples t-tests separately within each algorithm condition (comparing problem-focused and emotion-focused coping in the same group). In the supervised condition, problem-focused coping (M = 3.48, SD = 1.01) was significantly higher than emotion-focused coping (M = 3.03, SD = 1.02), t (147) = 10.109, $p < 0.001$, Cohen's d = 0.835. In the unsupervised condition, emotion-focused coping (M = 5.29, SD = 1.06) was significantly higher than problem-focused coping (M = 4.73, SD = 1.26), t (149) = −11.416, $p < 0.001$, Cohen's d = 0.932 (absolute t = 11.416). As shown in Figure 3, in the supervised algorithm scenario, the participants were more inclined to adopt problem-focused coping behaviors, whereas in the unsupervised algorithm scenario, participants were more inclined to adopt emotion-focused coping behaviors. Thus, H1a and H1b were supported.

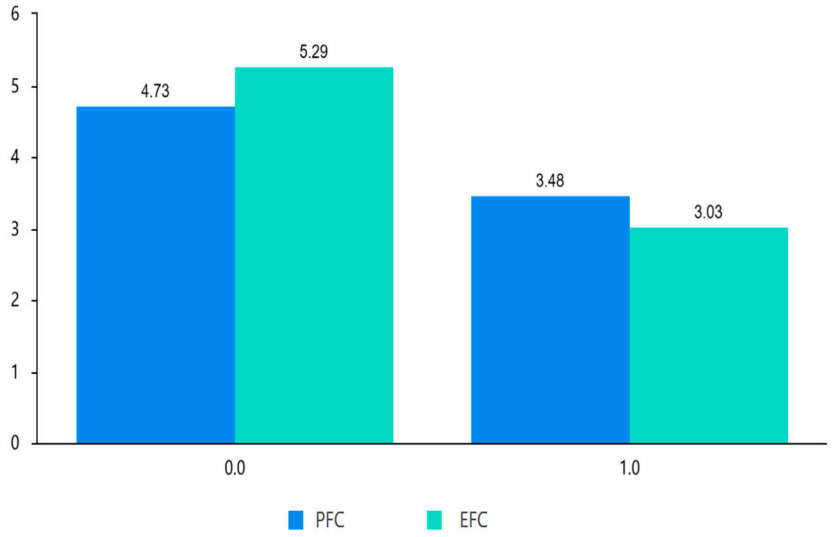


Figure 3. T-test results for the supervisibility of the recommendation algorithm throughput and audiences' coping behaviors. Notes: PFC = problem-focused coping behaviors; EFC = emotion-focused coping behaviors.

4.3.2. Testing the Mediating Role of Perceived Threat

An independent samples t-test was conducted to verify H2. The results showed that perceived threat was significantly higher in the unsupervised condition (M = 5.12, SD = 0.97) than in the supervised condition (M = 2.98, SD = 0.84), t (296) = 20.327, $p < 0.001$, Cohen's d = 2.353. Thus, H2 was supported.

To verify H3, separate regression analyses were conducted for perceived threat and audiences' coping behaviors. The results showed that perceived threat positively impacted audiences' coping behaviors (F (1296) = 727.626, $p < 0.001$), with a standardized regression coefficient of β = 0.843, supporting H3.

H4 was verified by conducting a regression analysis on the relationships between the supervisibility of the recommendation algorithm throughput and perceived threat and between perceived threat and audiences' coping behaviors. The results showed that the supervisibility of the recommendation algorithm throughput had a significant total effect

on audiences' coping behaviors ($c = -1.752$, $p < 0.001$). After adding the mediating variable of perceived threat, the direct effect of the supervisibility of the recommendation algorithm throughput on audiences' coping behaviors became smaller and not significant ($c' = 0.022$, $p > 0.05$). The path analysis revealed that the supervisibility of the recommendation algorithm throughput significantly negatively predicted perceived threat ($a = -2.144$, $p < 0.001$), while perceived threat significantly positively predicted audiences' coping behaviors ($b = 0.828$, $p < 0.001$). The indirect effect of perceived threat was −1.775, with a 95% bootstrap confidence interval of [−2.036, −1.522], which did not include 0, indicating that this mediating effect was statistically significant. To summarize, perceived threat fully mediated the relationship between the supervisibility of the recommendation algorithm throughput and audiences' coping behaviors, supporting H4.

4.3.3. Testing the Moderating Effects of Privacy Concerns and Self-Efficacy

To verify H5, the independent and moderating variables were first mean-centered. The moderation analysis included three models (Table 7). The first model included the independent variable (the supervisibility of the algorithm throughput), the second model additionally included the moderating variable (privacy concerns), and the third model additionally included the interaction term between the independent and moderating variables.

Table 7. Moderating effect of privacy concerns (n = 298).

	Model 1					**Model 2**					**Model 3**				
	B	**SE**	**t**	***p***	**β**	**B**	**SE**	**t**	***p***	**β**	**B**	**SE**	**t**	***p***	**β**
Constant	4.057	0.053	76.870	0.000 ***	-	4.057	0.040	102.559	0.000 ***	-	4.050	0.039	104.196	0.000 ***	-
Supervisibility of the recommendation algorithm throughput	−2.144	0.106	−20.307	0.000 ***	−0.763	−2.079	0.079	−26.243	0.000 ***	−0.740	−2.077	0.078	−26.715	0.000 ***	−0.739
Privacy concerns						0.615	0.040	15.228	0.000 ***	0.429	0.642	0.040	15.912	0.000 ***	0.448
Supervisibility of the recommendation algorithm throughput × Privacy concerns											−0.285	0.081	−3.538	0.000 ***	−0.100
R^2	0.582					0.766					0.776				
Adjusted R^2	0.581					0.764					0.773				
F	F (1296) = 412.389, p = 0.000					F (2295) = 482.991, p = 0.000					F (3294) = 338.739, p = 0.000				
ΔR^2	0.582					0.184					0.010				
ΔF	F (1296) = 412.389, p = 0.000					F (1295) = 231.901, p = 0.000					F (1294) = 12.519, p = 0.000				

Note: Model 1 includes only the independent variable (algorithm supervisibility); Model 2 adds the moderator (privacy concerns); Model 3 adds the interaction term. B = unstandardized coefficient; SE = standard error; β = standardized coefficient; ΔR^2 = change in R^2; ΔF = change in F value. Dependent variable: perceived threat. The interaction term (supervisibility × privacy concerns) is shown in the last row of Model 3. *** $p < 0.001$.

In the first model, the effect of the independent variable (the supervisibility of the algorithm throughput) on the dependent variable (perceived threat) was examined without including the influence of the moderating variable (privacy concerns). As shown in Table 7, the supervisibility of the algorithm throughput was significant ($t = -20.307$, $p < 0.001$). The interaction term between the supervisibility of the algorithm throughput and privacy concerns was significant ($t = -3.538$, $p < 0.001$), indicating that privacy concerns moderated the relationship between algorithm supervisibility and perceived threat.

The moderating effect of privacy concerns was further analyzed by combining it with a simple slope graph (Figure 4). As privacy concerns increased, the influence of the supervisibility of the algorithm throughput on perceived threat strengthened; that is, privacy concerns positively moderated the influence of the supervisibility of the algorithm throughput on perceived threat. Thus, H5 was supported.

To verify H6, the independent and moderating variables were first mean-centered. The moderation analysis included three models (Table 8). The first model included the independent variable (perceived threat), the second model additionally included the moderating variable (self-efficacy), and the third model additionally included the interaction term between the independent and moderating variables.

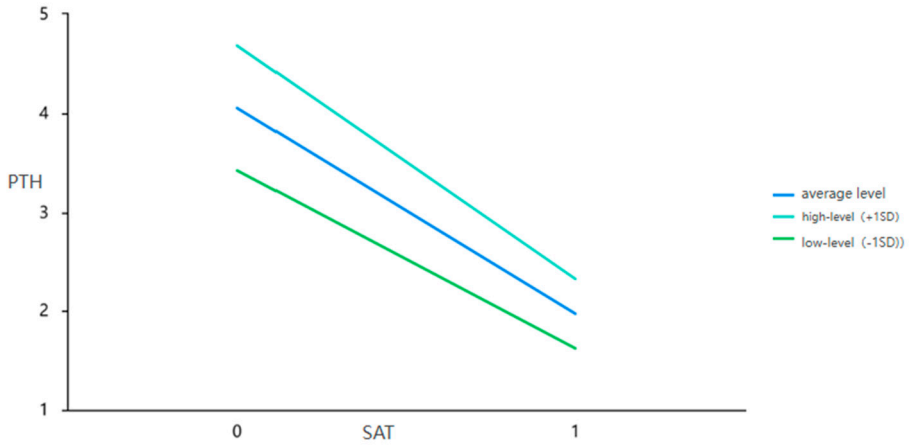


Figure 4. The moderating effect of privacy concerns.

Table 8. Moderating effect of self-efficacy (n = 298).

	Model 1					Model 2					Model 3				
	B	SE	t	*p*	β	B	SE	t	*p*	β	B	SE	t	*p*	β
Constant	4.136	0.043	96.645	0.000 ***	-	4.136	0.043	96.482	0.000 ***	-	4.124	0.043	96.437	0.000 ***	-
Perceived threat	0.822	0.030	26.975	0.000 ***	0.843	0.822	0.031	26.737	0.000 ***	0.843	0.802	0.031	25.547	0.000 ***	0.823
Self-efficacy						0.001	0.048	0.030	0.976	0.001	−0.010	0.048	−0.216	0.829	−0.007
Perceived threat × Self-efficacy											0.085	0.033	2.549	0.011 *	0.082
R^2	0.711					0.711					0.717				
Adjusted R^2	0.710					0.709					0.714				
F	*F* (1296) = 727.626, *p* = 0.000					*F* (2295) = 362.586, *p* = 0.000					*F* (3294) = 248.392, *p* = 0.000				
ΔR^2	0.711					0.000					0.006				
ΔF	*F* (1296) = 727.626, *p* = 0.000					*F* (1295) = 0.001, *p* = 0.976					*F* (1294) = 6.495, *p* = 0.011				

Note: Model 1 includes only the independent variable (perceived threat); Model 2 adds the moderator (self-efficacy); Model 3 adds the interaction term. B = unstandardized coefficient; SE = standard error; β = standardized coefficient; ΔR^2 = change in R^2; ΔF = change in *F* value. Dependent variable: audiences' coping behaviors. The interaction term (perceived threat × self-efficacy) is shown in the last row of Model 3. * $p < 0.05$, *** $p < 0.001$.

In the first model, the influence of the independent variable (perceived threat) on the dependent variable (audiences' coping behaviors) was investigated without considering the influence of the moderating variable (self-efficacy). As shown in Table 8, perceived threat was significant ($t = 26.975$, $p < 0.001$), indicating that it had a significant impact on audiences' coping behaviors. Additionally, the interaction term between perceived threat and self-efficacy was significant ($t = 2.549$, $p < 0.05$), indicating that self-efficacy moderated the relationship between perceived threat and audiences' coping behaviors.

The moderating effect of self-efficacy was further analyzed by combining it with a simple slope graph (Figure 5). As self-efficacy increased, the influence of perceived threat on audiences' coping behaviors strengthened; that is, self-efficacy positively moderated the influence of perceived threat on audiences' coping behaviors, supporting H6.

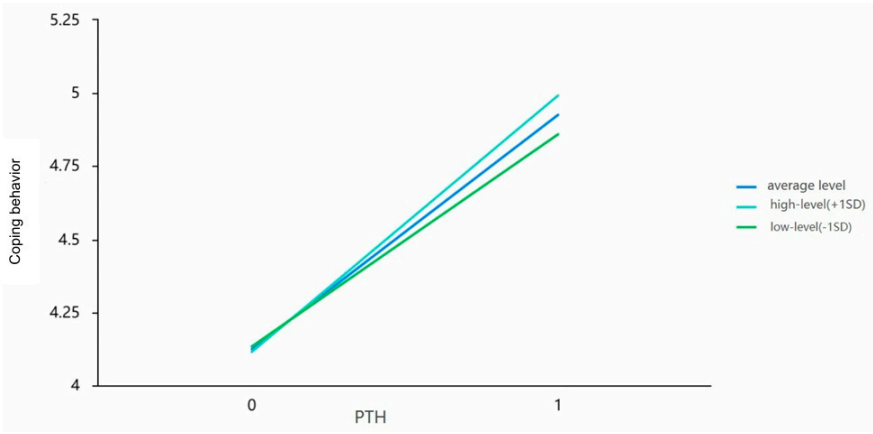


Figure 5. The moderating effect of self-efficacy.

5. Conclusions

5.1. Summary of the Findings

This study examines how recommendation algorithm throughput affects audiences' coping behaviors and explores the mediating effect of perceived threat and the moderating effects of self-efficacy and privacy concerns. Based on a scenario-based experiment including 298 participants, this study draws the following conclusions. First, the supervisibility of the recommendation algorithm throughput significantly impacts audiences' coping behaviors. Unsupervised algorithms increase audiences' coping behaviors more than supervised algorithms. Further, supervised and unsupervised recommendation algorithms cause targeted advertising audiences to adopt problem- and emotion-focused coping behaviors, respectively.

Second, the impact of the supervisibility of the recommendation algorithm throughput on audiences' perceived threat varies depending on the type of algorithm. Targeted advertising audiences are more likely to perceive threats from unsupervised algorithms and this perception positively impacts their coping behaviors. Moreover, perceived threat mediates the influence of the supervisibility of the recommendation algorithm throughput on audiences' coping behaviors.

Third, the moderating effect of privacy concerns on the relationship between the supervisibility of the recommendation algorithm throughput and perceived threat is significant. At a high level of privacy concerns, unsupervised algorithms lead to higher perceived threat; that is, privacy concerns play a positive moderating role. Moreover, the moderating effect of self-efficacy on the relationship between perceived threat and audiences' coping behaviors is also significant. At high levels of self-efficacy, as perceived threat increases, audiences are more likely to adopt coping behaviors; that is, self-efficacy plays a positive moderating role.

Nevertheless, the use of weight-related products in the scenarios may have independently influenced perceived threat; future research using neutral product categories is needed to isolate the pure algorithm effect.

5.2. Contributions

From a theoretical perspective, the findings are significant for future research. Most previous studies on consumers' negative behaviors such as avoidance of and coping with algorithm recommendations have been based on privacy management theories, focusing on how privacy attitudes affect behaviors. Discussions on algorithm recommendations have largely concentrated on purely technical aspects such as optimizing algorithm models to increase adoption. In contrast, we find that the supervisibility of algorithm throughput determines audiences' attitudes and behaviors toward recommendation algorithms, offering a new theoretical lens beyond privacy and technical perspectives. Specifically, this study extends TTAT in three ways. First, it introduces algorithm throughput supervisibility (supervised vs. unsupervised) as a novel exogenous antecedent of threat appraisal, moving beyond the traditional focus on security threats. Second, it clarifies that users' perceived loss of control—rather than data breach risk—can trigger threat avoidance processes, thereby broadening TTAT's applicability to algorithmic design contexts. Third, it identifies privacy concerns and self-efficacy as moderators that amplify the threat–coping linkage, offering a more nuanced understanding of individual differences. These extensions demonstrate that TTAT is a versatile framework for studying human–AI interaction beyond its original security domain.

From a practical perspective, this study offers three tentative recommendations for platforms and advertisers seeking to reduce users' perceived threat and negative coping behaviors. First, platforms may consider disclosing the level of algorithm supervisibility

(e.g., indicating whether the algorithm is supervised or unsupervised). For unsupervised algorithms, which in our study induced significantly higher perceived threat (Cohen's $d = 2.35$), an additional explanation about controllability (e.g., clearing browsing history or adjusting privacy settings) could be helpful. Second, when unsupervised algorithms are used, platforms could offer user control options such as turning off cross-category inference, deleting browsing histories, or resetting preferences with a single click. These measures directly address the perceived loss of control that we identified as the core mechanism behind threat appraisal. Third, for product categories that are inherently sensitive (e.g., health, weight, finance), platforms could combine algorithm disclosure with reassuring messages that emphasize user control (e.g., "You can adjust your preferences at any time"). Our results suggest that such messages may be particularly effective for users with high privacy concerns and high self-efficacy. By implementing these strategies, platforms may help shift users from emotion-focused coping (avoidance and dissatisfaction) toward problem-focused coping (constructive actions such as adjusting settings), potentially improving user trust and advertising effectiveness.

5.3. Limitations and Scope for Future Research

Although this study makes several theoretical and practical contributions, it is not without limitations.

First, generalizability of the findings. The experimental scenario was based on a virtual shopping platform, and the sample, while covering 31 provincial-level regions in China, was skewed toward female (67.8%), younger (mean age 30.98, 54.4% aged 18–30), and highly educated participants (70.1% with a bachelor's degree or higher). Owing to the heterogeneity of potential users across platforms and demographic groups, the applicability of our findings to other settings may require further verification. Moreover, users' algorithm literacy varies, which may influence their attitudes toward algorithm supervisibility and coping behaviors. Future research should include users from multiple platforms, use more diverse samples (e.g., older adults, less educated populations, more balanced gender samples), and incorporate algorithm literacy as a potential moderator.

Second, methodological constraints. Although the scenario-based experiment largely reflects behavioral tendencies, reactions in simulated environments may differ from real-world settings. This study relied on self-reported coping intentions and did not capture actual behaviors. Furthermore, we manipulated the independent variables through scenarios and verified correlations via questionnaires at a single time point. Future research could: (a) use experimental techniques such as event-related potentials or behavioral tracking to capture more objective responses; (b) collect actual behavioral traces (e.g., click-through rates, time spent on privacy settings, account deletion, or switching to alternative platforms); (c) conduct within-subjects experiments; (d) adopt longitudinal designs to examine whether repeated exposure to low-supervisibility algorithms leads to habituation, increased threat, or changes in coping strategies over time.

Third, data recency. The data were collected between August 2022 and January 2023, a period preceding the widespread public and academic focus on generative AI (e.g., ChatGPT). While the core concept of algorithm throughput supervisibility (supervised vs. unsupervised) remains theoretically unchanged, the rapid evolution of AI technologies may have influenced consumers' general awareness of algorithmic systems. Future research should replicate our findings using more recent data to ensure the temporal generalizability of the results.

Fourth, product sensitivity. As explained in the research design, the scenarios deliberately included weight-related products (plus-size clothing and weight-loss items) to simulate realistic cross-category inferences by algorithms—a feature intended to activate

participants' awareness of algorithmic inference. This design choice was part of our manipulation, not a confound. Nevertheless, such products may independently evoke emotional discomfort and perceived threat. This is a significant limitation of the current study: the observed differences between algorithm conditions reflect the combined influence of algorithm type and product sensitivity. The core conclusions that unsupervised algorithms increase perceived threat and emotion-focused coping remain valid, but the effect sizes should be interpreted with this product-related influence in mind. Future research should employ a clearer experimental design, for example by using neutral product categories (e.g., books, electronics) or holding product category constant across conditions, to isolate the pure effect of algorithm characteristics.

Fifth, manipulation wording and inherent entanglement of manipulation elements. The unsupervised condition description "not fully controlled by the developers" may have primed participants to perceive not only algorithmic autonomy but also a lack of human accountability. Specifically, the phrase "not fully controlled" could evoke a sense that the algorithm might behave unpredictably, make erroneous inferences, or operate without any mechanism for correction or redress—all of which are likely to heighten perceived threat independently of participants' conceptual understanding of unsupervised learning. While this wording was necessary to credibly convey low supervisibility, it may have introduced an additional affective cue that could influence perceived threat. Future research could test alternative phrasings that convey algorithmic autonomy without explicitly mentioning developer control to isolate the effect of perceived controllability more cleanly. Meanwhile, in our manipulation, we used multiple descriptors, such as basic-level vs. high-level autonomous and controllable vs. not fully controlled, to operationalize supervised vs. unsupervised. These elements were deliberately aligned because in our conceptualization, algorithmic autonomy and controllability are integral to the degree of supervisibility. They are not extraneous confounds but rather natural expressions of the construct. Nevertheless, because these elements were presented together, the current design does not allow us to fully separate their individual effects. We acknowledge this as a limitation. Future research could attempt to vary perceived human oversight while keeping other descriptors more constant to better isolate the core dimension of supervisibility.

Author Contributions: Conceptualization, L.C. and R.S.; methodology, L.F.; software, L.F. and L.C.; validation, L.F. and L.Z.; formal analysis, L.F.; investigation, L.C. and L.F.; resources, L.Z.; data curation, L.C.; writing—original draft preparation, L.C.; writing—review and editing, L.C. and L.F.; visualization, L.Z.; supervision, R.S.; project administration, R.S.; funding acquisition, R.S. and L.F. All authors have read and agreed to the published version of the manuscript.

Funding: This research project was supported by Humanities and Social Sciences Planning Project of the Ministry of Education, PRC (Grant No. 20YJA630054), the National Social Sciences Fund of China (Grant No. 22BGL006), and the Humanities and Social Sciences Foundation of Fujian Province of China (Grant Number: FJ2024C082).

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki, and approved by the Ethics Committee of Jimei University(protocol code JMU202206033 and date of approval 1 June 2022). The research design fully took into account various potential risks, and detailed explanations were given to the participants before the study began. The research data were processed in accordance with privacy protection regulations and were anonymized.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data that support the findings of this study are available from the corresponding author upon reasonable request. The data are not publicly available due to privacy and ethical restrictions regarding participant anonymity.

Acknowledgments: We also thank the anonymous reviewers who provided valuable comments on the manuscript.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Usama, M.; Qadir, J.; Raza, A.; Arif, H.; Yau, K.A.; Elkhatib, Y.; Hussain, A.; Al-Fuqaha, A. Unsupervised machine learning for networking: Techniques, applications and research challenges. *IEEE Access* **2019**, *7*, 65579–65615. [CrossRef]
2. Langer, M.; Landers, R.N. The future of artificial intelligence at work: A review on effects of decision automation and augmentation on workers targeted by algorithms and third-party observers. *Comput. Hum. Behav.* **2021**, *123*, 106878. [CrossRef]
3. Antonio, A.B.; Dela Cruz, K.L.; Jimenez, I.A.; Pantoja, E. Invasion or personalization: An overview on user attitudes towards the privacy issues in targeted advertising in NCR and its effect in consumer purchase behavior. *J. Bus. Manag. Stud.* **2022**, *4*, 38–47. [CrossRef]
4. Peng, L. Illusion, algorithmic prisoners, and rights transfer: New risks in the era of data and algorithms. *J. Northwest. Norm. Univ. Soc. Sci. Ed.* **2018**, *55*, 20–29. [CrossRef]
5. Zhang, H.; Quan, D.; Lu, T. Research on the paradox of users' response to privacy threats in online shopping context: From the perspective of response behavior theory. *J. Inf. Sci.* **2019**, *38*, 141–148.
6. Zhang, J.; Liu, J.; Zhong, W. Advertising precision and advertising effectiveness: A field experiment based on privacy concern. *J. Manag. Sci.* **2019**, *32*, 123–132.
7. Li, Z.; Sun, R. Research on the influence of recommendation explanations on changing users' behavioral intentions: From the perspective of communication and persuasion theory. *China Soft Sci.* **2019**, *6*, 176–184.
8. Yang, Y.; Li, Y.; Lian, X. A systematic review and prospects of algorithm avoidance behavior. *China Hum. Resour. Dev.* **2024**, *41*, 112–127. [CrossRef]
9. Culnan, M.J.; Armstrong, P.K. Information privacy concerns, procedural fairness, and impersonal trust: An empirical investigation. *Organ. Sci.* **1999**, *10*, 104–115. [CrossRef]
10. Sun, R.; Luo, Y. A study on consumers' privacy paradox behavior from the perspective of self-perception theory: Evidence from ERPs. *Nankai Bus. Rev.* **2021**, *24*, 153–162. [CrossRef]
11. Wang, Y.; Xi, Y.; Jiang, Y. Research on the nonlinear relationship between privacy salience and avoidance behavior of internet targeted advertising: The mediating role of perceived threat. *Nankai Bus. Rev.* **2023**, *26*, 135–147.
12. Yang, M.; Wen, X. The mediating effect of privacy protection willingness: Privacy concern, privacy protection self-efficacy, and targeted advertising avoidance. *Press Circ.* **2020**, *7*, 41–52. [CrossRef]
13. Zhu, Q.; Wang, X.; Xin, L. Research on the influence mechanism of privacy concern on click-through intention of targeted online advertising: The roles of consumers' risk perception and self-efficacy. *Soft Sci.* **2018**, *32*, 115–119. [CrossRef]
14. Peng, H.P.; Chen, Y. Algorithmic manipulation in short video recommendations and its collaborative governance. *Chin. Ed. J.* **2023**, *3*, 86–90.
15. Xu, J.B.; Wang, J.F. Algorithm security: The invisible mechanism of pseudo public opinion and risk governance. *Mod. Commun. J. Commun. Univ. China* **2022**, *44*, 138–146. [CrossRef]
16. Yang, K.K.; Chen, L. AI-empowered ideological governance and its risk response. *World Social. Stud.* **2022**, *7*, 98–106, 135.
17. Zhang, C.C. The artificial public opinion manipulated by algorithms is the natural enemy of the real world. *Young J.* **2022**, *7*, 111. [CrossRef]
18. Shin, D. The effects of explainability and causability on perception, trust, and acceptance: Implications for explainable AI. *Int. J. Hum. Comput. Stud.* **2021**, *146*, 102551. [CrossRef]
19. Miller, T. Explanation in artificial intelligence: Insights from the social sciences. *Artif. Intell.* **2019**, *26*, 1–38. [CrossRef]
20. Doshi-Velez, F.; Kim, B. Towards a rigorous science of interpretable machine learning. *arXiv* **2017**, arXiv:1702.08608. [CrossRef]
21. Turilli, M.; Floridi, L. The ethics of information transparency. *Ethics Inf. Technol.* **2009**, *11*, 105–112. [CrossRef]
22. Mehrabian, A.; Russell, J.A. *An Approach to Environmental Psychology*; MIT Press: Cambridge, MA, USA, 1974.
23. Liang, H.; Xue, Y. Understanding security behaviors in personal computer usage: A threat avoidance perspective. *J. Assoc. Inf. Syst.* **2010**, *11*, 394–413. [CrossRef]
24. Shin, D.; Park, Y.J. Role of fairness, accountability, and transparency in algorithmic affordance. *Comput. Hum. Behav.* **2019**, *98*, 277–284. [CrossRef]
25. Qiu, Y.Q.; Chen, C.F. Choice of "Black Box" Opening Methods Based on Artificial Intelligence and Algorithmic News Transparency. *J. Zhengzhou Univ. (Philos. Soc. Sci. Ed.)* **2018**, *51*, 84–88. Available online: https://www.cnki.com.cn/Article/CJFDTotal-ZZDX201805016.htm (accessed on 5 April 2026).
26. Praveena, M.; Jaiganesh, V. A literature review on supervised machine learning algorithms and boosting process. *Int. J. Comput. Appl.* **2017**, *169*, 32–35. [CrossRef]

27. Shin, D. User perceptions of algorithmic decisions in the personalized AI system: Perceptual evaluation of fairness, accountability, transparency, and explainability. *J. Broadcast. Electron. Media* **2020**, *64*, 541–565. [CrossRef]
28. Zhang, H. Normative and technical analysis of autonomous algorithm privacy protection. *Lanzhou Acad. J.* **2021**, *3*, 120–135.
29. Qadir, J.; Yau, K.L.A.; Imran, M.A.; Ni, Q.; Vasilakos, A.V. IEEE Access special section editorial: Artificial intelligence enabled networking. *IEEE Access* **2015**, *3*, 3079–3082. [CrossRef]
30. Suthaharan, S. Big data classification: Problems and challenges in network intrusion prediction with machine learning. *SIGMETRICS Perform. Eval. Rev.* **2014**, *41*, 70–73. [CrossRef]
31. Lazarus, R.S.; Folkman, S. *Stress, Appraisal, and Coping*; Springer Publishing Company: New York, NY, USA, 1984.
32. Shin, D. How do people judge the credibility of algorithmic sources? *AI Soc.* **2022**, *37*, 81–96. [CrossRef]
33. Henderson, M.J.; Shade, L.R.; Mackinnon, K. Every click you make: Algorithmic literacy and the digital lives of young adults. *AoIR Sel. Pap. Internet Res.* **2020**. [CrossRef]
34. Noble, S.U. *Algorithms of Oppression: How Search Engines Reinforce Racism*; New York University Press: New York, NY, USA, 2018.
35. Xu, H.; Luo, X.; Carroll, J.M.; Rosson, M.B. The personalization privacy paradox: An exploratory study of decision making process for location-aware marketing. *Decis. Support Syst.* **2011**, *51*, 42–52. [CrossRef]
36. Jiang, Y.; Wang, Y. Research on the impact of Internet targeted advertising push methods on consumers' attitudes: The mediating role of perceived threat. *Mod. Manag.* **2020**, *40*, 64–67. [CrossRef]
37. Hong, W.; Thong, J.Y.L. Internet privacy concerns: An integrated conceptualization and four empirical studies. *MIS Q.* **2013**, *37*, 275–298. [CrossRef]
38. Miltgen, C.L.; Henseler, J.; Gelhard, C.; Popovič, A. Introducing new products that affect consumer privacy: A mediation model. *J. Bus. Res.* **2016**, *69*, 4659–4666. [CrossRef]
39. China Academy of Information and Communications Technology (CAICT). *Telecom and Internet Users' Rights and Interests Protection White Paper*; CAICT: Beijing, China, 2018.
40. Bandura, A. Self-efficacy: Toward a unifying theory of behavioral change. *Psychol. Rev.* **1977**, *84*, 191–215. [CrossRef]
41. Li, Y. Theories in online information privacy research: A critical review and an integrated framework. *Decis. Support Syst.* **2012**, *54*, 471–481. [CrossRef]
42. Esmaeili, M. Assessment of Users' Information Security Behavior in Smartphone Networks. Doctoral Dissertation, University of British Columbia, Vancouver, BC, Canada, 2014.
43. Johnston, A.C.; Warkentin, M. Fear appeals and information security behaviors: An empirical study. *MIS Q.* **2010**, *34*, 549–566. [CrossRef]
44. Workman, M.; Bommer, W.H.; Straub, D. Security lapses and the omission of information security measures: A threat control model and empirical test. *Comput. Hum. Behav.* **2008**, *24*, 2799–2816. [CrossRef]
45. Yang, Q. A Study on Information Disclosure and Protection Behaviors in Social Networks. Doctoral Dissertation, University of Science and Technology of China, Hefei, China, 2021.
46. Malhotra, N.K.; Kim, S.S.; Agarwal, J. Internet users' information privacy concerns (IUIPC): The construct, the scale, and a causal model. *Inf. Syst. Res.* **2004**, *15*, 336–355. [CrossRef]
47. Wang, C.K.; Hu, Z.; Liu, Y. Reliability and validity of the General self-efficacy Scale. *Appl. Psychol.* **2001**, *7*, 37–40.
48. Schwarzer, R.; Jerusalem, M. Generalized self-efficacy scale. In *Measures in Health Psychology: A User's Portfolio. Causal and Control Beliefs*; Weinman, J., Wright, S., Johnston, M., Eds.; NFER-Nelson: Windsor, UK, 1995; pp. 35–37.
49. Shi, H. Why Do Users Remain Unconcerned in the Face of Information Security Threats? An Empirical Study in the Cloud Computing Context. Master's Thesis, Jiangsu University of Science and Technology, Zhenjiang, China, 2016.
50. Chen, X. An Empirical Study on Consumers' Responses to Merchant Collusion on Reputation. Master's Thesis, Guangdong University of Technology, Guangzhou, China, 2017.
51. Podsakoff, P.M.; MacKenzie, S.B.; Lee, J.-Y.; Podsakoff, N.P. Common method biases in behavioral research: A critical review of the literature and recommended remedies. *J. Appl. Psychol.* **2003**, *88*, 879–903. [CrossRef] [PubMed]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.