디지털경제사회연구 2025년 제3권 제1호, pp.165~197
KISDI JDES 25-3-1

Information Disclosure on Digital Platform Services: The Roles of Multidimensional Trust and Perceived Benefits of Recommendation Algorithms

Jae-Eun Oh* · Ki Joon Kim**

This study examines how trust in personal information management capabilities can be associated with the intention to disclose personal information, with the perceived benefit of YouTube algorithm as a mediating factor. Using survey data from 4,437 Korean users, we employed factor analysis and mediation analysis. The results revealed that trust in established entities and in emerging platforms positively predicted the intention to disclose personal information. However, an unexpected algorithm literacy paradox emerged: while trust promoted perceived benefit of YouTube algorithm paradoxically decreased intention to disclose personal information. These findings challenge assumptions about the benefits of algorithmic transparency and extend privacy calculus theory by demonstrating that algorithm awareness can inhibit personal information disclosure. Increased algorithmic understanding may enhance privacy consciousness, leading to more cautious information disclosures. This research contributes to digital privacy literature and provides insights into platform design and transparency policy development.

주제어: *YouTube, recommendation algorithm, information disclose, information management capability*

* Associate Professor, School of Design, Hong Kong Polytechnic University, jae.oh@polyu.edu.hk
** Corresponding author, Associate Professor, Department of Media and Communication, City University of Hong Kong, kj.kim@cityu.edu.hk

1. Introduction

The contemporary digital landscape is increasingly dominated by artificial intelligence systems that fundamentally reshape how individuals interact with digital platforms. AI-driven algorithms, based on vast amounts of user data, now perform complex processes including personalized content curation and public service delivery (Meijer, Lorenz, & Wessels, 2021). This technological evolution has created unprecedented dependency on personal information, as AI systems require extensive user data to function effectively (Gupta, Goyal, & Kumar, 2024). Unlike conventional digital services with transparent data collection, modern AI systems operate through complex processes that analyze personal information in ways often opaque to users (Albahri et al., 2023). This creates a paradoxical situation where users must provide detailed personal information to access AI-enhanced services while lacking clear understanding of data processing.

YouTube serves as an adequate basis for examining algorithmic transparency and privacy behavior. Its recommendation algorithm drives approximately 70% of platform content consumption, influencing billions of users worldwide (NBC News, 2024. June 18). However, investigations reveal concerning aspects including tendencies to recommend extreme content to maximize engagement. Mozilla's research indicates that YouTube's user control features have "negligible" effects, suggesting users experience minimal agency over their algorithmic experiences.

Current digital environments present a critical paradox– while users increasingly understand that personal information powers algorithmic systems, the relationship between this awareness and intention to disclose personal information remains poorly understood. Existing privacy research has focused on traditional privacy calculus models, assuming users make rational cost-benefit calculations when deciding information disclosure (Dinev & Hart, 2006). However, these models may inadequately explain user behavior in algorithmic contexts.

The emergence of algorithm awareness as a mediating factor represents a significant theoretical gap. While research has examined relationships between institutional trust and information disclosure (Pavlou, 2011), limited attention addresses how algorithmic system perceptions influence this relationship. This gap is problematic given that algorithmic mediation has become the dominant mode of digital interaction. Furthermore, users interact with diverse digital platforms from traditional banking to emerging AI technologies. The assumption that users apply uniform trust criteria across all digital services may be flawed, yet most research treats digital service trust as monolithic.

This study addresses these gaps through three primary objectives. First, this research develops understanding of how users evaluate institutional trustworthiness across different digital service categories. Second, this study extends privacy calculus theory by incorporating algorithmic awareness, examining how YouTube algorithm perceptions influence trust-disclosure relationships. Third, this research investigates the "algorithm literacy paradox"- whether increased algorithmic understanding decreases intention to disclose personal information, challenging conventional transparency assumptions.

2. Personal information management capabilities

Artificial intelligence, including generative AI, requires personal information. Basically, AI learns based on large amounts of user data, including personal information, and uses personal information to provide customized services to users (Bhuiyan, 2024). Recently, as AI, such as generative AI, has become more sophisticated, AI models must be able to explain how they reached their conclusions in order to be trusted in real-world environments (Albahri et al., 2023). To this end, AI inevitably requires the use and analysis of user personal information. In such situations, users often have to provide their personal information to utilize AI-based services from companies, public institutions, or even government services (Willems et al., 2023). Naturally, the capabilities of companies, institutions, or governments to manage personal information becomes a critical issue for users who have provided their information. This is because if these entities manage personal information poorly or lack the necessary capabilities, the consequences of data breaches, such as hacking, fall squarely on the individual (Lee & Lee, 2012). As such, the management of personal information by relevant businesses, institutions, and governments is a very important issue that is linked to the convenience of users and the trust of businesses and institutions.

In response, related businesses, institutions, and government are paying close attention to the protection and management of user personal information. For example, traditional financial institutions such as banks have established conservative personal information protection systems under the strict regulations of the Financial Supervisory Service (Financial Stability Board, 2024, July 16). On the other hand, internet-only banks such as Kakao Bank and Toss are facing new challenges in balancing innovative services with personal information protection (Kim, Kwon, & Kim, 2023). In particular, social media platforms such as KakaoTalk, Facebook, and TikTok operate in a complex environment where they must comply with various personal information protection regulations such as GDPR (General Data Protection Regulation) and CCPA (California Consumer Privacy Act) due to their global service characteristics. Kakao has also raised concerns about the possibility of anonymity being compromised or personal information being exposed in open chats due to the linkage of member serial numbers and temporary IDs in the design and operation of the KakaoTalk service (Maeil Business, 2024, May

23). Generative AI (i.e., ChatGPT, Gemini, Copilot) and metaverse platforms (i.e., Fortnite, Zepeto) are raising new risks that existing personal information protection regulations did not anticipate. Microsoft provides practical support for public sector organizations when considering generative AI services, guiding them on how to utilize generative AI functions while complying with GDPR obligations (Microsoft Public Sector Center of Expertise, 2024, April). However, issues such as the inclusion of personal information in training data and the generation of false personal information due to hallucinations remain unresolved challenges.

The government and other public entities are also working hard to manage personal information. The Personal Information Protection Commission, an independent central administrative agency established to independently handle matters related to the processing and protection of personal information, is performing strong supervisory functions, such as imposing fines on domestic social media platforms that have caused controversy over personal information. In addition, public services such as Government 24 and Minwon 24 are setting examples for personal information protection by minimizing the processing of resident registration numbers and providing various means of identity verification (Yoon, 2017). However, there are significant disparities in personal information management capabilities among local governments. In educational institutions, issues related to student personal information protection have emerged due to the spread of online classes, while in medical institutions, finding a balance between the use and protection of personal information in AI diagnosis and infectious disease response processes has become a major challenge (Anjum, 2025, January 10). Across the public sector as a whole, the shortage of specialized personnel in personal information protection and the need to address disparities in capabilities between institutions remain key challenges in the process of implementing digital government and adopting new technologies.

Based on these discussions, differences in the use of personal information between services or fields suggest that users may apply different standards of trust to each area. Accordingly, this study sets the following research questions.

RQ1. What sub-dimensions comprise the level of trust in the personal information management capabilities of companies, public institutions, and government?

3. Relationship between user trust in personal information management capabilities and the intention to disclose personal information

1) The concept of trust

Trust and reliability may appear to be similar concepts. However, they require clear conceptual distinction. Trust represents a psychological state based on individuals' subjective evaluations, whereas reliability refers to the objective and technical consistency of systems. Trust in personal information management capabilities is defined as individuals' subjective confidence and expectations that specific institutions will appropriately and securely handle their personal information. This conceptualization is grounded in established theoretical frameworks. Mayer, Davis, and Schoorman (1995) conceptualized trust as formed through evaluations of the trustee's ability, benevolence, and integrity.

In a similar vein, Dinev and Hart (2006) also demonstrated that institutional trust plays a pivotal role in individuals' information disclosure decisions. Son and Kim (2008) further confirmed that trust mitigates perceived risks of information disclosure in online environments. Based on these theoretical foundations, this study employs the concept of "trust" rather than "reliability."

2) Cross-domain trust transfer and algorithm awareness

While general institutional trust and specific platform algorithm awareness may appear conceptually distant, their relationship is theoretically grounded in trust transfer theory and cross-domain trust mechanisms. Trust transfer theory posits that users' general institutional trust can be transferred to specific platforms or services through the interaction between platform-based mechanisms and general institutional mechanisms (Chen et al., 2015). This theoretical framework suggests that individuals' trust evaluations are not confined to isolated domains but operate through interconnected trust networks.

Cross-domain trust transfer has been empirically validated in various digital contexts. Research on robo-advisor services demonstrates that general technology trust, vendor trust, and specific service trust operate through transfer mechanisms across different domains (Fang et al., 2019). Similarly, studies on AI-based algorithms confirm that users' general institutional and technological trust influences the formation of trust in specific algorithmic systems (Glikson & Woolley, 2020). YouTube algorithm awareness can be understood within this framework, where users' perceptions of a specific algorithmic system are informed by their broader trust in digital institutions and information management capabilities.

3) Trust in personal information management capabilities and the intention to disclose personal information

Trust in an organization's ability to manage personal information directly influences decisions to intention to disclose personal information (Mutimukwe, Kolkowska, & Grönlund, 2020). Intention to disclose personal information is defined as individuals deliberate propensity or psychological tendency to provide their personal information to service providers upon request when using digital services or applications. This conceptualization is grounded in established theoretical frameworks from communication and behavioral sciences. The Theory of Planned Behavior (TPB) emphasizes that behavioral intentions represent the immediate antecedent of actual behavior, formed through the interplay of attitudes, subjective norms, and perceived behavioral control (Ajzen, 1991). Within privacy research, Norberg, Horne, and Horne (2007) demonstrated that disclosure intentions serve as critical predictors of actual information disclosure behaviors, while highlighting the distinction between what individuals intend to do versus what they actually do -- a phenomenon known as the privacy paradox.

Privacy Calculus Theory positions disclosure intention as the outcome of individuals' cost-benefit calculations regarding intention to disclose personal information. Individuals will disclose personal information if they perceive that the overall benefits of disclosure are at least balanced by, if not greater than, the assessed risk of disclosure (Dinev & Hart, 2006). Here, trust is a key factor that directly influences the outcome of this risk-benefit assessment. When a website is perceived as trusted users are willing to accept their vulnerability in exchange for the benefit of receiving more useful personalized recommendations (Seckler et al., 2015). This means that trust provides a direct motivation for disclosing personal information. Trust forms the basis of social exchange and relationships, indicating a willingness to depend on a partner in an exchange and confidence in that partner (Shanka, & Buvik, 2019).

Numerous empirical studies have confirmed a direct positive relationship between trust and personal information disclosure. Some studies (e.g., Dienlin

& Metzger, 2024; Huang, 2015; Lee & Jhou, 2025; Taddei & Contena, 2013) have shown that trust has a positive effect on self-disclosure, meaning that users disclose more information on websites when they have a high level of trust in those websites. Research has also shown that users' trust in a platform affects the extent to which they disclose their personal information (Dienlin & Metzger, 2024). Individuals' decisions to disclose personal information depend on their trust in an organization's ability to manage personal information. It is important to explain how an organization's personal information protection mechanisms can shape individual users' privacy concerns, privacy awareness, trust, and information disclosure behavior (Xu et al., 2011). This means that an organization's personal information management capabilities form users' trust through comprehensive capabilities at the technical, policy, procedural, and organizational levels, and that this trust determines the intention to disclose personal information.

Summarizing the above theoretical discussion, the hypothesis that trust in personal information management capabilities will have a positive effect on the intention to disclose personal information is based on the following logical grounds. Therefore, we have established the following hypothesis.

H1. Trust in personal information management capabilities will have a positive effect on the intention to disclose personal information.

4. The mediating role of perceived benefit of YouTube algorithm

YouTube's recommendation algorithm possesses distinctive features that amplify its mediating effect between institutional trust and information disclosure. The algorithm drives approximately 70% of platform content consumption and shapes information consumed by billions of users worldwide

(NBC News, 2024), creating heightened user sensitivity to its operations. More critically, research indicates that YouTube's user control features have negligible effects (Hattingh, 2021), suggesting users experience substantial perceived lack of control over algorithmic processes. In this context, the relationship between trust in personal information management capabilities and the intention to disclose personal information may operate through a more sophisticated mechanism.

The pathway from trust in personal information management capabilities to perceived benefit of YouTube algorithm can be theoretically explained through trust transfer theory, as mentioned above. Trust transfer theory posits that trust developed in one domain can influence initial trust in other domains that are perceived to have certain connections to the trusted domain (Stewart, 2003; McKnight et al., 2002). In digital contexts, research on government surveillance technologies has demonstrated that individuals' trust in one domain can be transferred to serve as initial trust in related technological systems when certain linkages are perceived between the domains (Lu et al., 2011; Oldeweme et al., 2021).

Furthermore, this trust transfer is reinforced by institutional proximity effects. When users encounter YouTube's recommendation algorithm, they cognitively associate it with their existing trust frameworks regarding digital service providers' data management capabilities. Research on digital trust formation reveals that user perceptions and expectations indirectly affect digital trust through user satisfaction, creating a cognitive pathway that facilitates trust transfer from general institutional capabilities to specific algorithmic systems (Guo, 2022). That is, users who perceive that traditional digital service providers (such as banks, search engines, and e-commerce platforms) manage personal information competently are more likely to develop positive perceptions of YouTube algorithm systems operated by similar entities.

On the other hand, the relationship between perceived benefit of YouTube algorithm and the intention to disclose personal information operates through a counterintuitive mechanism. Unlike conventional assumptions that greater algorithmic transparency leads to increased information sharing (Molina & Sundar, 2024), empirical evidence suggests that enhanced algorithm awareness may actually reduce users' willingness to disclose personal information.

This paradoxical effect can be explained through the extended privacy calculus framework. Algorithm awareness acts as a potential antecedent factor in the privacy decision-making process, influencing how consumers assess privacy risks and make self-disclosure decisions (Tian, Zhang, & He, 2024). When users develop comprehensive understanding of how YouTube's recommendation algorithms collect, analyze, and utilize their personal information, this awareness triggers a more sophisticated risk-benefit recalculation process.

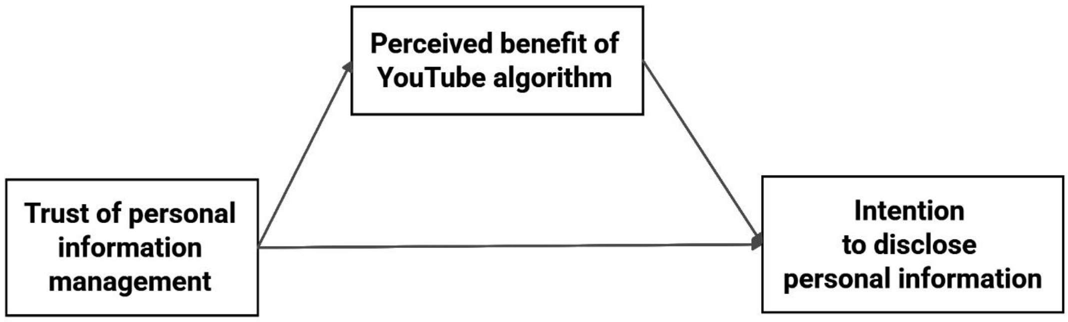
Specifically, users who better understand YouTube's algorithmic operations become more cognizant of the extent and implications of data collection practices. This heightened awareness leads to enhanced privacy consciousness, prompting users to adopt more cautious intentions to disclose privacy. Research on algorithm awareness in e-commerce contexts demonstrates that as users' understanding of algorithmic processes increases, they engage in more deliberate privacy calculus, often resulting in reduced information disclosure (Shin, Kee, & Shin, 2022).

In light of the preceding theoretical discourse, the following hypothesis was formulated.

H2. Perceived benefit of YouTube algorithm will mediate the relationship between trust in personal information management capabilities and the intention to disclose personal information.

In sum, the current study investigates associations among the constructs shown in Figure 1.

[Figure 1] Proposed research model



5. Methods

This study was based on data from the "2024 Intelligent Information Society User Panel Survey" conducted by the Korea Communications Commission and carried out by the Korea Information Society Development Institute (KISDI). This data was publicly released by the Korea Communications Commission in July 2025 and was collected from 4,437 respondents aged 17 to 71 residing nationwide who own smartphones and use the internet at least once a day. Considering that the data (panel) samples respondents with a certain level of digital literacy (Shim & Park, 2024), it aligns with the purpose of this study, which aims to examine the relationship between digital service users' management capabilities, perceptions of YouTube's recommendation algorithm, and intention to disclose personal information. In addition, the data includes demographic characteristics such as gender, age, and educational level, which serve as control variables in this study, as well as key variables that form the research model, such as perceptions of YouTube's recommendation algorithm, intention to disclose personal information, and digital service management capabilities.

The demographic characteristics of the panel survey sample are as follows. First, among the 4,437 samples, 43.9% were male and 56.1% were female. The age distribution was as follows: 5.7% were in their teens, 17.2% in their 20s, 17.5% in their 30s, 20.6% in their 40s, 21.7% in their 50s, and 17.4% in their 60s. In terms of educational attainment, 37.5% had a high school diploma or lower, while 62.5% had a college degree or higher. Finally, based on monthly income, respondents' perceptions of their standard of living were as follows: upper level 2.3%, upper-middle level 4.2%, middle level 44.5%, lower-middle level 43.6%, and lower level 5.5% (see Table 1). When analyzing gender and educational attainment, the gender and educational attainment variables were converted into dummy variables (male = 1, college degree or higher = 1) for analysis.

〈Table 1〉 Socio-demographic characteristics (n=4,437).

Variables	Statistics
Gender	Male (1,946, 43.9%), Female (2,491, 56.1%)
Age	10s (208, 5.7%), 20s (627, 17.2%), 30s (1,054, 17.5%), 40s (804, 20.6%), 50s (905, 21.7%), 60s(839, 17.4%)
Education	High school diploma or lower (1,662, 37.5%) College degree or higher (2,776, 62.5%)
Average Monthly Income	Upper level (100, 2.3%) Upper-middle level (186, 4.2%) Middle level (976, 44.5%) Lower-middle level (934, 43.6%) Lower level (242, 5.5%)

6. Measures

Intention to disclose personal information was measured by the extent to which individuals consented to disclosure or share their personal information to service manufacturers, operators, and other personal information collectors while using digital services or apps. To this end, the following measurement question was used: “When using a service or app, would you provide the following information if the manufacturer or operator requested it as necessary?” This question consists of a total of 15 detailed questions. Examples of sub-questions include “real name,” “age,” “mobile phone number,” and “current location.” All questions were measured using a 5-point Likert scale (1: absolutely cannot provide, 5: can provide completely). This study measured the extent to which users are willing to disclose their personal information. Based on the research by Shim and Cho, 2024, the 15 sub-questions were converted to 1 point (4–5 points) and 0 points (1–3 points) according to the intention to disclose personal information, and the response values of the 15 sub-items were summed. The total score ranges from 0 to 15 points (M = 7.01, SD = 4.30). A higher total score indicates a broader range of personal information disclosure. Additionally, it suggests a higher intention to disclose personal information, as respondents are willing to provide not only sensitive personal information such as home addresses but also digital information such as mobile phone numbers.

Trust in personal information management capabilities refers to the degree to which users perceive the ability of platform service companies and various institutions to safely protect their personal information during the process of collecting, storing, processing, and utilizing it. This variable was measured using the question, “How well do you think the following institutions or companies manage personal information?” The reliability of personal information management capabilities is measured using 14 detailed items. For example, it includes items such as “traditional financial institutions (e.g., XX Bank, YY Bank, etc.),” “government or public institutions,” “virtual world platforms (e.g., Fortnite, Zepeto, etc.),” and “generative AI (e.g., xGPT, Coxxx, etc.).” Responses were collected using a 5-point Likert scale (1 = not managed at all, 5 = managed very well, 9 = don't know). In each question, “don't know” responses were treated as missing values (Denman et al., 2018) to clearly reflect the respondents' actual intentions and were excluded from the analysis.

The trust in personal information management capabilities is composed of various detailed items, ranging from traditional services such as banking to trust in emerging platforms such as generative AI. Therefore, it cannot be ruled out that this concept may be composed of one or more sub-concepts. Accordingly, this study conducted exploratory factor analysis and then measured the reliability and validity of the scale based on the extracted factors. As shown in the research results below, the detailed items were found to be loaded on the factors of "*Trust in established entities* (M = 3.73, SD = .82, Cronbach's α = .88)" and "*Trust in emerging platforms* (M = 4.29, SD = .09, Cronbach's α = .89)." The results of the construct validity verification using the chi-square test were statistically significant (χ^2 = 27,664, $p < .001$), and the KMO value was .889, indicating suitability for factor analysis.

Perceived benefit of YouTube algorithm was measured using 6 detailed items based on the question, "How do you feel about YouTube's automatic recommendation service?" Among these, a total of six detailed items were used to measure perceptions of YouTube's recommendation algorithm. For example, YouTube's automatic recommendation service was composed of detailed items such as "It is well suited to my tastes," "It is useful to me," "It is well suited for watching movies or videos," "It is unbiased and objective," "It is generally satisfactory," and "It provides diverse information rather than uniform information." Responses were collected using a 5-point Likert scale (1: Not at all, 5: Very much so) (M = 3.72, SD = .46, Cronbach's α = .69).

Among the five sub-items excluded from the measurement, one was a question about the intention to continue using YouTube's recommendation algorithm, and the remaining four items were found to impair internal trust. In addition, exploratory factor analysis was conducted to confirm whether the six detailed items showed consistent construct validity. Principal component analysis was performed, and the eigenvalue loaded on the first factor was 2.36, with an explained variance of 39.4%. The factor loading the detailed items ranged from a minimum of .718 to a maximum of .793. In addition, Bartlett's sphericity test and KMO sample adequacy were conducted to confirm the suitability of factor analysis. The results showed that the sphericity test was significant (χ^2 = 3002, $p < .001$), and the KMO value was .761, confirming the suitability of factor analysis.

7. Statistical Analysis

This study conducted descriptive statistical analysis, factor analysis (personal information management capability trust), and correlation analysis (Pearson's Correlation Analysis) between variables to identify the characteristics of the main variables. To verify RQ 1, exploratory factor analysis (EFA) was conducted on "personal information management capability trust." Next, to verify the research hypothesis, we analyzed how YouTube algorithm users' perceptions mediate the relationship between personal information management capability trust and intention to disclose personal information.

SPSS version 28.0 was used as the statistical package for analysis, and the Process Macro Model 4 proposed by Hayes (2012) was utilized to verify the mediating effect. At this point, bootstrapping was set with a sample size of 10,000 and a confidence interval of 95% to verify the significance of the mediating effect. Table 2 shows the correlations and descriptive statistics of all variables included in the factor analysis and mediating effect verification to validate this research model.

〈Table 2〉 Zero-Order Correlation Matrix of Key Variables

Variables	1. Gender	2. Age	3. Edu.	4. TEE	5. TEP	6. PBYA	7. IDPI
1	1						
2	-.03	1					
3	.110***	-.351***	1				
4	-.034*	.102***	-.028	1			
5	-.02	.036*	-.029	.310***	1		
6	.01	-.021	-.012	-.070***	.182***	1	
7	.02	-.024	.048***	.110***	.050***	-.077***	1
M	.44	44.2	5.03	3.73	4.29	3.72	7.01
SD	.49	14.5	.96	.82	.09	.46	4.30

Note: TEE (Trust in established entities), TEP (Trust in emerging platforms), PBYA (perceived benefit of YouTube algorithm), IDPI (intention to disclose personal information), *** $p < .001$; ** $p < .01$; * $p < .05$

8. Results

1) Trust in personal information management capabilities

This study sought to explore the subfactors contained in the various detailed items that constitute the trust in personal information management capabilities through RQ1. Before extracting the factors, Bartlett's sphericity test and KMO sample adequacy verification were conducted to confirm whether the 14 items were suitable for factor analysis. The results showed that, as indicated in Table 3, both the sphericity test results and the sample adequacy results (〉80, Kaiser & Rice, 1974) met the criteria for factor analysis. Therefore, principal component analysis was used to extract the factors.

The factor analysis results showed that the 14 detailed items loaded onto two factors. The factor labeling reflected not only all the loaded items but also the unique characteristics that clearly distinguished them from other factors. Established entities derive their legitimacy from long-standing institutional frameworks, regulatory structures, and social contracts that have been developed over time (Markard et al., 2016). In contrast, emerging platforms must establish legitimacy through technological innovation and user acceptance, often operating in tension with existing institutional environments (Coşkun & Arslan, 2024). This institutional perspective recognizes that legitimacy depends on the congruence of organizational structure, activities, and outcomes with the regulatory, normative, and cognitive processes of the institutional environment. Digital trust, as it has evolved from traditional interpersonal and institutional trust, reflects these different legitimacy-building processes in the digital society (Albinson et al., 2019). Based on this discussion, the first factor was named "trust in established entities." The items loaded onto the trust in established entities factor consist of 13 detailed items, including search sites, traditional financial institutions, online shopping malls, and delivery apps. These services can be considered established digital services. In other words, they are services that are already familiar to users and have been socially verified. Accordingly, this study referred to the first factor as the trust of everyday digital service management capabilities.

In contrast, the second factor, "trust in emerging platforms," is composed of innovative platforms based on recently emerged technologies. Specifically, the three sub-items are virtual world platforms, generative AI, and online games demonstrating metaverse-like scalability. The factor was named based on the characteristic that these services are relatively new and innovative but have not yet been sufficiently verified.

〈Table 3〉 Exploratory factor analysis of personal information management capability

KMO goodness-of-fit test	MSA = .889	
Bartlett's Test of Sphericity	χ^2	27,664
	df	91
	p	〈.001

Variables	Communality	Factor loading	
		1	2
Traditional financial institutions (xxx Bank etc.)	.531	.692	
Search engines (Naver, Daum, Google, etc.)	.500	.726	
Internet financial institutions (xxkao Bank, x Bank, etc.)	.594	.634	
Social media (xxaoTalk, xxbook, xxagram, xTok, etc.)	.546	.684	
Government or public institutions	.646	.586	
Non-profit organizations (charities, foundations, etc.)	.649	.587	
Online shopping malls (xxupang, 11st, xMarket etc.)	.565	.659	
Telecommunications companies (SKT, KT, LGU+)	.514	.702	
Delivery apps/sites (xinjok, xgiyo, xpang xts, etc.)	.577	.629	
Medical institutions	.691	.526	
OTT service providers (xflix, xisneyx, xxG, xavve, etc.)	.645	.435	
Online game companies (Xgue of Legends, PUBG, etc.)	.361		.787
Virtual world platforms (Fortnite, Zepeto, etc.)	.112		.952
Generative AI (ChatGPT, Gemini, Copilot, etc.)	.275		.847

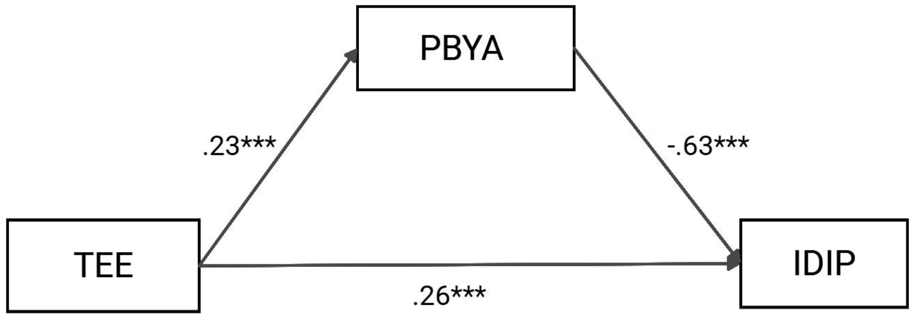
Factor	TEE	TEP
Eigen value	4.40	2.40
% of variance	31.4	17.1
Cumulative %	31.4	48.5
Cronbach's α	.88	.89

Note: TEE (Trust in established entities), TEP (Trust in emerging platforms)

Mediating effect between trust in personal information management capabilities and intention to disclose personal information in perceived benefit of YouTube algorithm

This study sought to predict the direct effect of user trust in the personal information management capabilities of companies or institutions on the intention to disclose personal information through H1. In RQ1 mentioned above, personal information management capabilities were classified into factors such as trust in established entities and trust in emerging platforms. Accordingly, H1 separately verified the association with trust in established entities on the intention to disclose personal information and the association with trust in emerging platforms on the intention to disclose personal information.

[Figure 2] Relationship between trust in established entities and the intention to disclose personal information in perceived benefit of YouTube algorithm



As shown in Figure 2 and Table 4 below, the effect of trust in established entities on the intention to disclose personal information was significant ($b = .26$, $p < .001$). In other words, the higher the level of trust users have in the trust in established entities, the higher the level of personal information they are willing to provide.

〈Table 4〉 Indirect Effect of trust in established entities for intention to disclose personal information through perceived benefit of YouTube algorithm

DV	IV	b	SE	t	95% CI		F	R square
					LL	UL		
PBYA	TEE	.23	.03	7.56***	.17	.29	17.1***	.015
	gender	.11	.13	.88	−.14	.37		
	age	−.006	.01	−1.29	−.02	.003		
	edu	.38	.14	2.65**	.10	.66		
IDPI	TEE	.26	.03	7.74***	.13	.33	18.9***	.025
	PBYA	−.63	.14	−4.24***	−.07	−.34		
	gender	.16	.13	1.12	.02	.43		
	age	−.01	.01	−2.84**	−.05	−.01		
	edu	.25	.15	1.64	.03	.56		

	b	SE	Z	95% CI	
				LL	UL
Total effect	.12	.02	8.61***	.17	.29
Direct effect	.26	.03	8.62***	.20	.32
Indirect effect	−.14	.003	3.10***	.004	.02

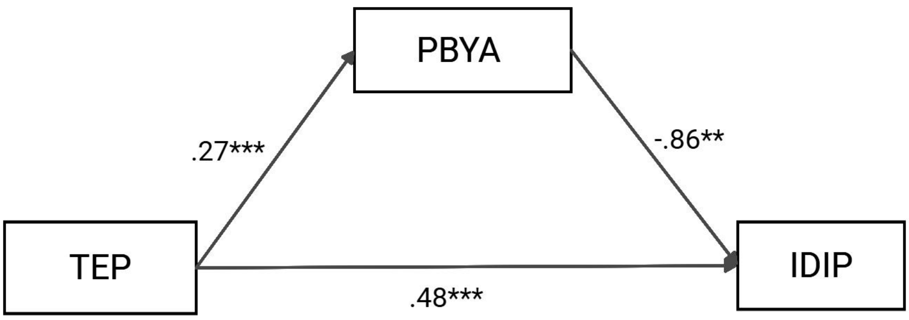
Note: TEE (trust in established entities), PBYA (perceived benefit of YouTube algorithm), IDPI (intention to disclose personal information), *** $p < .001$; ** $p < .01$; * $p < .05$

Through H2, we sought to verify how perceived benefit of YouTube algorithm mediate trust in established entities and intention to disclose personal information. The analysis results showed that trust in established entities had a positive effect on perceived benefit of YouTube algorithm ($b = .23$, $p < .001$). This suggests that the higher users' trust in their management capabilities for established entities they use on a daily basis, the more positive their perception of YouTube's recommendation algorithm is likely to be. Additionally, when analyzing the effect of perceived benefit of YouTube algorithm on the intention to disclose personal information, it was found that the more positively users perceived benefit of YouTube algorithm, the lower intention to disclose personal information ($b = -.63$, $p < .001$).

Next, we verified the mediating effect of perceived benefit of YouTube algorithm on the relationship between trust in established entities and the intention to disclose personal information. The analysis showed a significant indirect effect ($b = -.14$, $p < .001$). When comparing the total effect ($b = .12$, $p < .001$) and the direct effect ($b = .26$, $p < .001$), the direct effect remained significant even after incorporating the perceived benefit of YouTube algorithm. However, the direction of the indirect effect was negative, indicating that perceived benefit of YouTube algorithm negatively inhibits the relationship between trust in established entities and the willingness to provide personal information.

Similarly, this study verified the direct effect of trust in emerging platforms on the intention to disclose personal information.

[Figure 3] Relationship between trust in emerging platforms and the intention to disclose personal information in perceived benefit of YouTube algorithm



The verification results showed that trust in emerging platforms had a significant impact on the intention to disclose personal information, as shown in Figure 3 and Table 5. In other words, the more users trust in emerging platforms faithfully, the more likely they are to provide a high intention to disclose personal information.

〈Table 5〉 Indirect effect of trust in emerging platforms for intention to disclose personal information through perceived benefit of YouTube algorithm

DV	IV	b	SE	t	95% CI		F	R square
					LL	UL		
PBYA	TEE	.27	.08	3.45***	.12	.42	5.70***	.010
	gender	.09	.13	.69	−.17	.35		
	age	−.003	.005	−.64	−.01	.01		
	edu	.40	.14	2.79**	.17	.68		
IDPI	TEE	.48	.09	5.53**	.31	.65	13.0***	.018
	PBYA	−.86	.15	−5.70**	−1.2	−.56		
	gender	.15	.14	1.10	−.13	.42		
	age	−.01	.01	−2.22*	−.02	−.001		
	edu	.27	.16	1.75	−.03	.58		

	b	SE	Z	95% CI	
				LL	UL
Total effect	.25	.08	3.07**	.10	.43
Direct effect	.48	.10	5.00***	.30	.68
Indirect effect	−.23	.02	−4.53***	−.13	−.05

Note: TEP (trust in emerging platforms), PBYA (perceived benefit of YouTube algorithm), IDPI (intention to disclose personal information), *** p 〈 .001; ** p 〈 .01; * p 〈 .05

In addition, we verified the mediating effect of perceived benefit of YouTube algorithm on the relationship between trust in emerging platforss and the intention to disclose personal information. The verification results showed that trust in emerging platforms had a positive effect on the perceived benefit of YouTube algorithm (b = .27, p 〈 .001). In other words, the higher users evaluated the management capabilities of companies or institutions regarding trust in emerging platforms, the more positive the benefits YouTube’ s recommendation algorithm were. We also analyzed the effect of the perceived benefit of YouTube algorithm on the intention to disclose personal information. The analysis results showed that the more positively users perceived benefit of YouTube algorithm, the lower the intention to disclose personal information was (b = −.86, p 〈 .01).

Meanwhile, when we verified the mediating effect of perceived benefit of YouTube algorithm on the relationship between trust in the management capabilities of trust in emerging platforms and the intention to disclose personal information, we found a statistically significant indirect effect ($b = -.23$, $p < .001$). When comparing the total effect ($b = .25$, $p < .01$) and the direct effect ($b = .48$, $p < .001$), the direct effect remained significant even after introducing perceived benefit of YouTube algorithm as a mediating variable. However, since the indirect effect showed a negative effect, it can be interpreted that the perceived benefit of YouTube algorithm inhibits the relationship between trust in emerging platforms and the intention to disclose personal information.

9. Discussion

This study examined the impact of user trust in the personal information management capabilities of companies or government agencies on the intention to disclose personal information by users. Furthermore, it explored how perceived benefit of YouTube algorithm mediate the relationship between these two factors. To this end, factor analysis was first used to define the sub-dimensions of user trust in personal information management capabilities. We then assessed the effects of each dimension of trust in personal information management capabilities on the intention to disclose personal information and analyzed the indirect effects of each dimension through perceived benefit of YouTube algorithm.

First, user trust in personal information management capabilities was composed of two dimensions: “trust in established entities” and “trust in emerging platforms.” Trust in established entities were composed of digital services and institutional organizations that are already widely established in our society and widely used in daily life, such as traditional financial institutions, search sites, online shopping malls, and delivery apps. On the other hand, trust in emerging platforms included innovative platforms that provide next-generation digital environments and new user experiences and interaction methods based on relatively recent innovative technologies, such as virtual world platforms, generative AI, and online games. These results suggest that users perceive the maturity or familiarity of services as the primary criteria when evaluating the personal information management capabilities of companies, governments, and public institutions. This demonstrates that the definition of trust proposed by Mayer, Davis, and Schoorman (1995), namely, "the willingness to be vulnerable to another's actions based on positive expectations about their intentions or behaviors, regardless of whether one can monitor or control them," is formed differently depending on the type of service. In particular, the distinctive characteristics between the two factors imply that digital service providers must establish differentiated strategies for each service to build user trust. In the traditional service domain, it is important to continuously strengthen security and improve transparency based on existing trust foundations, while in the new service domain, active communication regarding the safety and ethical use of technology is necessary.

Second, both trust in established entities and trust in emerging platforms were found to have a positive effect on the intention to disclose personal information. This means that the more users trust a specific service provider's personal information management capability, the more willing they are to provide personal information to that service. Interestingly, the effect of trust in emerging platforms was greater than trust in established entities. These results can be interpreted as reflecting users' ambivalent attitudes toward new technologies. Given the high level of uncertainty surrounding trust in emerging platforms, it can be interpreted that once trust is established, its influence becomes even more significant. Additionally, users of trust in emerging platforms may be a group that is relatively innovative and willing to take risks. This result emphasizes the importance of building initial trust when introducing new technologies.

Third, the mediating effect of perceived benefit of YouTube algorithm appeared in a direction different from expectations. The higher the trust in personal information management capabilities, the more positive perceived benefit of YouTube algorithm. However, paradoxically, positive perceived benefit of YouTube algorithm was found to reduce the intention to disclose personal information. As a result, the overall mediating effect was confirmed to be negative. These findings demonstrate the "algorithm literacy paradox," in which users become more cautious about providing personal information as their understanding of algorithms increases. This is a new discovery that extends the concept of the privacy paradox proposed by Hughes-Roberts (2013) and Taddicken (2014) to the realm of algorithm transparency. Unlike the privacy paradox defined in previous studies as "the phenomenon of observed discrepancies or inconsistencies between attitudes toward privacy and actual behavior," this study identified a new paradoxical relationship between positive perceptions of algorithms and intention to disclose personal information.

Fourth, the inhibitory effect of perceived benefit of YouTube algorithm was observed in both dimensions of personal information management capability trust. However, the magnitude of this effect was greater in emerging platforms. This suggests that users of emerging platforms are more sensitive to algorithms. Groups with higher levels of interest and understanding of new technologies are more likely to be aware of the potential risks of algorithms. The results of this study provide several important practical implications. Digital service providers should go beyond simply improving technical security levels and instead understand the process of building user trust and establish differentiated trust-building strategies tailored to each service type.

Especially for emerging platforms clear communication about the safety and ethical use of technology is necessary.

Fifth, the negative mediating effect of perceived benefit of YouTube algorithm on intention to disclose personal information represents more than a simple inhibitory mechanism. Our findings suggest that this phenomenon involves complex psychological processes including learned helplessness, psychological reactance, and perceived loss of informational control. As users develop greater perceived benefit of YouTube algorithm, they may simultaneously experience a sense of powerlessness regarding their ability to control these systems, leading to defensive information-sharing behaviors (Lewandowsky, et al., 2024). This is consistent with psychological reactance theory, which suggests that when individuals perceive their autonomy as being threatened by algorithmic control, they engage in resistance behaviors to restore their sense of freedom (Schlund & Zitek, 2024). Furthermore, the opacity and complexity of recommendation algorithms may create a situation analogous to learned helplessness, where users recognize the system's influence but feel unable to meaningfully control or influence its operations (Lu, 2024). This dual awareness may create a psychological tension that manifests as reduced willingness to provide personal information, even when users acknowledge the functional benefits of algorithmic personalization.

Finally, when establishing personal information protection policies, it is necessary to consider the complex relationship between users' perceptions and behaviors. Unlike the traditional negative relationship between privacy concerns and intention to disclose personal information suggested in studies by Dinev and Hart (2006), Malhotra, Kim, and Agarwal (2004), and Son and Kim (2008), this study identified a more complex relationship in the context of algorithms. Rather than simply increasing or decreasing the provision of personal information, it is important to create an environment where users can make rational decisions based on sufficient information. To achieve this, technical security measures must be complemented by user education, transparent information disclosure, and the provision of user–centric personal information management tools.

This study is not without limitations. First, this study targeted only respondents aged 17 to 71 who owned a smartphone and used the internet at least once a day, fundamentally limiting the sample to a group with a certain level of digital literacy. As the researchers noted, this was an intentional choice, but it excluded intention to disclose personal information and algorithm awareness of digitally marginalized groups or those with low digital capabilities. In particular, the perspectives of older adults or low–income groups with limited experience using digital services were omitted, limiting the generalizability of the research results to the entire Korean society. Furthermore, due to the nature of panel surveys, respondents who continuously participate in surveys are more likely to have a higher intention to participate in surveys and a higher level of interest in digital services than the general population, so this selection bias may have caused systematic errors in the research results.

Second, this study used a cross–sectional design to collect data at a single point in time, which fundamentally limits the ability to establish true causal relationships between variables. In particular, it is impossible to clearly establish the temporal sequence between trust in personal information management capabilities and the intention to disclose personal information, and the possibility of spurious relationships caused by third variables cannot be completely ruled out. Additionally, to explain the "algorithm literacy paradox" phenomenon, where the mediating effect of perceived benefit of YouTube algorithm appeared in the opposite direction of expectations, longitudinal analysis of the process of user awareness change is necessary. However, the current study design does not allow for sufficient exploration of such dynamic processes.

Third, another limitation of this study is the conceptual ambiguity in the perceived benefit of YouTube algorithm construct, which conflates functional usefulness and critical control awareness without clear distinction. Future research should develop separate measures for algorithmic functionality perceptions versus user agency and control to better understand the algorithm literacy paradox. Qualitative studies could also provide deeper insights into how users differentiate between these dimensions of algorithm awareness.

References

Albahri, A. S., Duhaim, A. M., Fadhel, M. A., Alnoor, A., Baqer, N. S., Alzubaidi, L., ... & Deveci, M. (2023). A systematic review of trustworthy and explainable artificial intelligence in healthcare: Assessment of quality, bias risk, and data fusion. *Information Fusion*, 96, 156-191.

Albinson, N., Balaji, S., & Chu, Y. (2019). Building digital trust: Technology can lead the way. Deloitte Insights.

Anjum, N. (2025, January 10). Artificial intelligence in education: Striking a balance between innovation & privacy. Edly. https://edly.io/blog/artificial-intelligence-in-education-and-privacy-concerns/.

Bhuiyan, M. S. (2024). The role of AI-Enhanced personalization in customer experiences. *Journal of Computer Science and Technology Studies*, 6(1), 162-169.

Coşkun, R., & Arslan, S. (2024). The role of organizational language in gaining legitimacy from the perspective of new institutional theory. *Journal of Management & Organization*, 30(6), 2003-2024.

Dienlin, T., & Metzger, M. (2024). Who needs privacy? Exploring the relations between need for privacy and personality. *Collabra: Psychology*, 10(1).

Dinev, T., & Hart, P. (2006). An extended privacy calculus model for e-commerce transactions. *Information Systems Research*, 17(1), 61-80.

Financial Stability Board. (2024, July 16). Recommendations for regulating and supervising bank and non-bank payment service providers offering cross-border payment services: Consultation report. https://www.fsb.org/uploads/P160724-2.pdf.

Guo, Y. (2022). Digital trust and the reconstruction of trust in the digital society: An integrated model based on trust theory and expectation confirmation theory. *Digital Government: Research and Practice, 3*(4), 1-19.

Gupta, S., Goyal, N. K., & Kumar, A. (2024). A historical, present, and prospective review of artificial intelligence's role in securing personal information and private data. In Artificial Intelligence and Communication Techniques in Industry 5.0 (pp. 257-272). CRC Press.

Hattingh, M. (2021). The dark side of YouTube: A systematic review of literature. IntechOpen.

Huang, L. S. (2015). Trust in product review blogs: The influence of self-disclosure and popularity. *Behaviour & Information Technology*, 34(1), 33-44.

Kim, S., Kwon, H. J., & Kim, H. (2023). Mobile banking service design attributes for the sustainability of internet-only banks: A case study of Kakao Bank. *Sustainability,* 15(8), 6428.

Lee, M., & Lee, J. (2012). The impact of information security failure on customer behaviors: A study on a large-scale hacking incident on the internet. *Information Systems Frontiers*, 14, 375-393.

Lee, C. W., & Jhou, Y. T. (2025). Understanding self-disclosure in social networking sites: The influence of trust and perceived privacy risk. *Journal of Applied Finance & Banking*, 15(1), 111-131.

Lewandowsky, S., Robertson, R. E., & DiResta, R. (2024). Challenges in understanding human-algorithm entanglement during online information consumption. *Perspectives on Psychological Science, 19*(5), 758-766.

Lu, B., Wang, Z., & Zhang, S. (2021). Platform-based mechanisms, institutional trust, and continuous use intention: The moderating role of perceived effectiveness of sharing economy institutional mechanisms. *Information & Management, 58*(7), 103504.

Lu, W. (2024). Inevitable challenges of autonomy: ethical concerns in personalized algorithmic decision-making. *Humanities and Social Sciences Communications, 11*(1), 1-9.

Malhotra, N. K., Kim, S. S., & Agarwal, J. (2004). Internet users' information privacy concerns (IUIPC): The construct, the scale, and a causal model. *Information Systems Research*, 15(4), 336-355.

Markard, J., Wirth, S., & Truffer, B. (2016). Institutional dynamics and technology legitimacy–A framework and a case study on biogas technology. *Research Policy,* 45(1), 330-344.

Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An integrative model of organizational trust. *Academy of management review, 20*(3), 709-734.

McKnight, D. H., Choudhury, V., & Kacmar, C. (2002). Developing and validating trust measures for e-commerce: An integrative typology. *Information Systems Research*, 13(3), 334-359.

Meijer, A., Lorenz, L., & Wessels, M. (2021). Algorithmization of bureaucratic organizations: Using a practice lens to study how context shapes predictive policing systems. *Public Administration Review,* 81(5), 837-846.

Microsoft Public Sector Center of Expertise. (2024, April). GDPR & generative AI: A guide for the public sector. https://wwps.microsoft.com/wp-content/uploads/2024/04/GDPR-and-Generative-AI-A-Guide-for-the-Public-Sector-FINAL.pdf.

Molina, M. D., & Sundar, S. S. (2022). When AI moderates online content: effects of human collaboration and interactive transparency on user trust. *Journal of Computer-Mediated Communication, 27*(4), zmac010.

Mutimukwe, C., Kolkowska, E., & Grönlund, Å. (2020). Information privacy in e-service: Effect of organizational privacy assurances on individual privacy concerns, perceptions, trust and self-disclosure behavior. *Government Information Quarterly*, 37(1), 101413.

NBC News. (2024, June 18). YouTube's algorithm more likely to recommend users right-wing and religious content, research finds. https://www.nbcnews.com/tech/tech-news/youtubes-algorithm-recommends-users-right-wing-religious-content-resea-rcna155478.

Oldeweme, A., Mä rtins, J., Westmattelmann, D., & Schewe, G. (2021). The role of transparency, trust, and social influence on uncertainty reduction in times of pandemics: empirical study on the adoption of COVID-19 tracing apps. *Journal of medical Internet research, 23*(2), e25893.

Pavlou, P. A. (2011). State of the information privacy literature: Where are we now and where should we go?. *MIS Quarterly*, 977-988.

Schlund, R., & Zitek, E. M. (2024). Algorithmic versus human surveillance leads to lower perceptions of autonomy and increased resistance. *Communications Psychology, 2*(1), 53.

Seckler, M., Heinz, S., Forde, S., Tuch, A. N., & Opwis, K. (2015). Trust and distrust on the web: User experiences and website characteristics. *Computers in Human Behavior*, 45, 39-50.

Shanka, M. S., & Buvik, A. (2019). When does relational exchange matters? Social bond, trust and satisfaction. *Journal of Business-to-Business Marketing*, 26(1), 57-74.

Shim, H., & Park, J. (2024). Relationship between YouTube's recommendation algorithm and continuous usage intention: The mediating effect of regulatory attitudes toward algorithmic transparency and accountability principles. *Korean Journal of Journalism & Communication Studies*, 68(5), 165-195.

Shin, D., Kee, K. F., & Shin, E. Y. (2022). Algorithm awareness: Why user awareness is critical for personal privacy in the adoption of algorithmic platforms?. *International Journal of Information Management*, 65, 102494.

Son, J. Y., & Kim, S. S. (2008). Internet users' information privacy-protective responses: A taxonomy and a nomological model. *MIS Quarterly*, 503-529.

Stewart, K. J. (2003). Trust transfer on the world wide web. *Organization science, 14*(1), 5-17.

Taddei, S., & Contena, B. (2013). Privacy, trust and control: Which relationships with online self-disclosure?. *Computers in Human Behavior*, 29(3), 821-826.

Tian, S., Zhang, B., & He, H. (2024). Role of algorithm awareness in privacy decision-making process: a dual calculus lens. *Journal of Theoretical and Applied Electronic Commerce Research, 19*(2), 899-920.

Willems, J., Schmid, M. J., Vanderelst, D., Vogel, D., & Ebinger, F. (2023). AI-driven public services and the privacy paradox: Do citizens really care about their privacy?. *Public Management Review*, 25(11), 2116-2134.

Xu, H., Dinev, T., Smith, J., & Hart, P. (2011). Information privacy concerns: Linking individual perceptions with institutional privacy assurances. *Journal of the*

Association for Information Systems, 12(12), 1.

Yoon, J. W. (2017). e-Government for better civil services: How the Korean government implemented the e-Registration system. Global Delivery Initiative Case Study, World Bank, Washington, DC.

논문 투고일: 2025. 6. 30.
논문 수정일: 2025. 7. 29.
게재 확정일: 2025. 8. 13.