

Invisible Influence and the Autonomy-Recognition Gap in AI Paternalism and Psychological Reactance

Khalil Gibran Kadir¹

¹ Universitas Hasanuddin, Indonesia

Correspondence: khalilgibrank@gmail.com

Article Info

Article History:

Received Aug 15th, 2026

Revised Aug 21th, 2026

Accepted Aug 30th, 2026

Keyword:

AI paternalism; Autonomy recognition; Digital choice architecture; Human-AI interaction; Psychological reactance.

ABSTRACT

This conceptual study examined how AI-mediated paternalistic influence could shape digital choice without being recognized by users as a threat to autonomy. The analysis used integrative literature synthesis and conceptual theory building across psychological reactance, autonomy, technology paternalism, digital nudging, and human-AI interaction. Published peer-reviewed studies were treated as secondary scholarly evidence rather than as an exhaustive dataset. The analysis distinguished system-level autonomy-relevant features from user recognition and perceived freedom threat. It developed the Autonomy-Recognition Gap as the discrepancy between identifiable features of an AI-mediated choice arrangement and the extent to which users recognized how those features bore on their freedom to choose. The synthesis showed that opacity and low recognition were related but conceptually distinct, while low reactance could not by itself demonstrate that autonomy had been preserved. Psychological legibility was positioned as a recognition condition, and reactance deficit was interpreted as one possible downstream outcome of recognition failure. Five theoretical propositions were formulated to guide later empirical testing without making causal or population-level claims.



© 2026 The Authors. Published by Punggawa Legacy Center. This is an open access article under the CC BY license (<https://creativecommons.org/licenses/by/4.0/>)

INTRODUCTION

AI now enters decisions at a stage where it can change what people encounter before they make an explicit choice. Recommender systems reorder alternatives, generative tools frame advice, and more agentic applications can execute parts of a decision after receiving only broad instructions. Digital nudging therefore operates in environments that are adaptive and often personalized rather than through a single visible interface cue. Weinmann et al. (2016) established the importance of digital choice architecture, while Jesse and Jannach (2021) showed how recommender systems extend that logic. Duane et al. (2025) later documented a wider shift toward AI-powered and personalized nudges. The practical issue is no longer only whether technology offers advice. It is also whether the structure around the advice changes what remains easy to notice, compare, refuse, or undo.

Paternalism matters when an intervention is justified by the person's welfare while interfering with choice or action. Technology complicates that picture because influence can be embedded in automated rules and interface design rather than a visible human decision. Rochi et al. (2024) distinguish limiting, overruling, and welfare-oriented dimensions of technology paternalism. AI adds a further difficulty because apparently independent system behavior still reflects institutional aims and design choices. The article therefore uses AI-mediated paternalistic influence and does not treat AI as an intentional moral agent.

Reactance theory becomes relevant after a restriction is experienced as bearing on a valued freedom. Dillard and Shen (2005) linked reactance to anger and negative cognition, while Rains (2013) confirmed the broader association between threatened freedom and reactance. Rosenberg and Siegel (2018) also show that the response depends on how restrictions are interpreted. Recent meta-analysis

reaches the same general conclusion. Li and Shi (2026) found that freedom-threatening communication increased anger, negative cognition, and reactance. AI therefore creates a prior question that standard reactance models can easily leave implicit. A choice environment may change before the user recognizes that the change matters for freedom. The technical novelty of AI does not by itself require a wholly new behavioral theory. Theoretical portability depends on whether an explanatory mechanism survives contextual transfer without losing its original boundary (Kadir, 2026c).

Recent AI studies make that distinction difficult to ignore. Oh et al. (2025) found that perceived AI autonomy was associated with greater threat to freedom and reactance, although personalization could improve attitudes toward the same technology. Reynolds-Tylus and Martinez Gonzalez (2026) likewise showed that AI attribution can be studied within a reactance framework without assuming that an AI source produces one uniform response. The sharper evidence comes from Görlitz and Rosenthal-von der Pütten (2026). In their experiment, covert restrictions initially generated comparatively weak reactance and disclosure changed autonomy perceptions. They described this pattern as a reactance deficit. Their finding does not leave an empty field. It creates a narrower theoretical problem about what, exactly, lies between an autonomy-relevant system arrangement and the user's representation of that arrangement as a threat.

Opacity cannot answer that question by itself. It describes limits on access to a system's operation, while recognition concerns what a user understands about the system's bearing on choice. A disclosure that recommendations are personalized may reveal little about omitted alternatives. An opaque system can produce a restriction when it blocks an action. Perceived autonomy differs again because interpretation has already occurred. The unresolved relation lies between autonomy-relevant features that can be identified in the choice arrangement and the user's recognition of what those features mean for choosing otherwise.

The Autonomy-Recognition Gap names that relation. It is the discrepancy between identifiable autonomy-relevant features of an AI-mediated choice arrangement and the extent to which the user recognizes how those features affect the ability to choose, refuse, revise, or act otherwise. The construct is not covert restriction, opacity, perceived autonomy, or reactance deficit under another label. Those concepts sit at different analytical levels. The gap connects a system-side basis for autonomy evaluation with a user-side recognition state. The study asks how this discrepancy can arise and when it may weaken the route from AI-mediated influence to perceived freedom threat and reactance.

RESEARCH METHODS

The study uses conceptual theory building through integrative literature synthesis. This design fits a problem in which the intended contribution is a clearer distinction among constructs and a more precise account of how they may relate, rather than an estimate of effect size or prevalence. Jaakkola (2020) describes conceptual research as suitable for theory synthesis and construct development. Torraco (2005, 2016) similarly treats integrative review as a way to connect bodies of scholarship that do not yet share one explanatory architecture. The present study therefore works with published research as secondary scholarly evidence and does not present a new empirical dataset.

Relevant literature was selected purposively from psychological reactance, autonomy and agency, technology paternalism, digital nudging, and human-AI interaction. Foundational studies were retained when they defined constructs that remain necessary for the argument. More recent work was used where AI changes the form or visibility of influence. The aim was conceptual coverage rather than exhaustive retrieval. Accordingly, the literature base is not treated as a representative sample from which frequencies, prevalence, or field-wide effect estimates can be inferred.

Analysis focused on distinctions that are easily collapsed in AI discussions. System features that alter the practical choice arrangement were separated from awareness that AI is present, recognition of how AI bears on choice, perceived threat to freedom, reactance, and later behavior. Studies were then read for evidence or argument about when those levels converge and when they do not. Particular attention was given to visibility of intervention, reversibility, explanation, personalization, and user control because these features recur across the relevant literatures. The resulting synthesis was used to refine the Autonomy-Recognition Gap, identify its boundary with adjacent concepts, and formulate a process framework with theoretical propositions.

The claims remain deliberately limited. Empirical studies support the plausibility of particular links or show that a distinction matters in observed settings, but they are not observations generated

here. The framework does not claim that recognition causes reactance, that one form of AI influence is more harmful, or that a recognition gap has a known prevalence. Its contribution is explanatory and classificatory. Later empirical work would be needed to estimate magnitude, test direction, and determine whether the proposed relations hold across tasks and populations.

RESULTS AND DISCUSSION

Restriction is not recognition

AI-mediated influence can be autonomy-relevant without being paternalistic in every instance. A recommendation that faithfully serves a user's stated goal, keeps alternatives available, and can be rejected at little cost may support rather than diminish self-direction. Formosa (2021) is useful here because machine autonomy and human autonomy are not inverse quantities. More independent machine action does not automatically mean less human autonomy. Hofmann (2026) also shows that artificial autonomy can affect understanding, competence, and voluntariness in different ways. The analysis should therefore begin with identifiable features of the choice arrangement rather than assume autonomy loss from the mere presence of automation. Filtering, difficult reversal, delegated execution, or concealed option suppression become relevant because they change the conditions under which a user can exercise choice, not because all automation is paternalistic by definition. A recent analysis of AI chatbot interaction makes a related distinction by treating conversational influence as behaviorally relevant without displacing human agency or reducing action to a single technological cause (Kadir, 2026b).

Recognition concerns a different level. A user may know that a recommendation came from AI and still misunderstand how the system organized the surrounding alternatives. Awareness of AI presence answers who or what participated in a decision. Autonomy recognition asks whether the person sees how the intervention bears on choosing otherwise. Mittelstadt et al. (2016) made opacity central to the ethics of algorithms, yet psychological recognition is not technical transparency. The user needs intelligible information that connects system behavior with its practical consequences for choice. Psychological legibility names that condition. It is a quality of the situation that can make autonomy-relevant influence easier or harder to interpret.

The Autonomy-Recognition Gap appears when autonomy-relevant features are identifiable in the choice arrangement but their significance is not recognized by the user. This avoids assuming that the researcher has already established an objective loss of autonomy. Covert restriction can produce such a gap, though it is only one route. A visible ranking may also do so when omitted alternatives remain unknown. Conversely, a technically opaque system may leave little gap when its practical restriction is obvious. Opacity may therefore contribute to recognition failure, but it is neither necessary nor sufficient for the gap. That distinction provides the construct with its main discriminant value.

Recognition conditions in digital choice

An explicit restriction usually leaves a visible trace. A blocked transaction or automated reversal is hard to mistake for neutral information even when the model remains opaque. Personalized ranking blends more easily into the environment the user takes as given. Weinmann et al. (2016) and Jesse and Jannach (2021) show why arrangement and salience matter. AI-powered nudges can also be personalized in real time (Duane et al., 2025). Recognition becomes harder when users see only the produced ranking and cannot compare it with what was omitted.

Table 1 Conceptual Architecture of AI Paternalism and Recognition

<i>Construct</i>	<i>Working meaning</i>	<i>Boundary in the present framework</i>
AI-mediated paternalistic influence	System-enabled steering or restriction that is presented or justified as serving the user	Describes an intervention arrangement and does not treat AI as an intentional moral agent
Autonomy-relevant feature	A feature that changes the practical conditions for choosing, refusing, revising, or acting otherwise	Does not by itself establish an objective loss of autonomy
Autonomy recognition	User recognition of how a system feature bears on freedom of choice	Differs from simple awareness that AI is present
Psychological legibility	The degree to which available information makes the system's bearing on choice understandable	A condition of recognition rather than a separate outcome

<i>Construct</i>	<i>Working meaning</i>	<i>Boundary in the present framework</i>
Autonomy-Recognition Gap	Discrepancy between identifiable autonomy-relevant features and user recognition of their bearing on choice	Differs from opacity, perceived autonomy, and covert restriction
Perceived freedom threat	Subjective experience that a valued freedom is threatened or removed	A proximal condition for reactance
Reactance deficit	Comparatively weak freedom-restoration response despite autonomy-relevant influence	A downstream outcome already identified in prior research

Control changes the interpretation of influence. Shneiderman (2020) argues that capable automation can coexist with meaningful human control. An accessible override makes the scope of automated authority easier to understand because the user can see that the system's action is contestable. A hidden or costly override leaves formal choice intact while making practical control difficult. Reversibility therefore matters in two ways. It can preserve user agency, and it can make the distribution of decision authority more recognizable. Neither function proves that autonomy was preserved, but both give the user a clearer basis for evaluating the intervention.

Disclosure creates a separate problem because information can be present without being useful. A notice that content is personalized does not necessarily reveal what was filtered, which criteria mattered, or whether the user can recover omitted alternatives. Görlitz and Rosenthal-von der Pütten (2026) found that disclosure changed autonomy perceptions after covert restriction, which shows that new information can revise the psychological representation of an intervention. That result should not be generalized into a claim that transparency solves paternalism. Explanation has value here when it improves psychological legibility. Dense technical detail may satisfy a formal transparency requirement while leaving the practical choice architecture just as difficult to evaluate.

Personalization can make influence feel benign even when it is consequential. Oh et al. (2025) found that perceived AI autonomy related to freedom threat while personalization was associated with more favorable attitudes. Choubey and Chakraborty (2026) similarly report that consumers may read agentic AI nudges through trust, persuasiveness, and ethical acceptance rather than through coercion alone. Such findings do not show that personalization hides autonomy loss. They show that attribution matters. Users can interpret intervention as assistance because it seems aligned with their interests. Sometimes that interpretation will be accurate. At other times the same benign framing may reduce attention to filtering, delegation, or constraints that would matter if they became salient. Recognition is therefore partly a problem of how influence is categorized, not simply whether it is noticed.

These conditions should be treated as recognition problems rather than levels of risk. Secondary literature does not justify labeling particular interventions low, moderate, or high in recognition risk. Table 2 therefore describes why influence may be difficult to interpret without ranking the forms. The change is also methodological. A conceptual article can compare mechanisms and specify plausible conditions, but it should not manufacture measurement precision that its evidence cannot support.

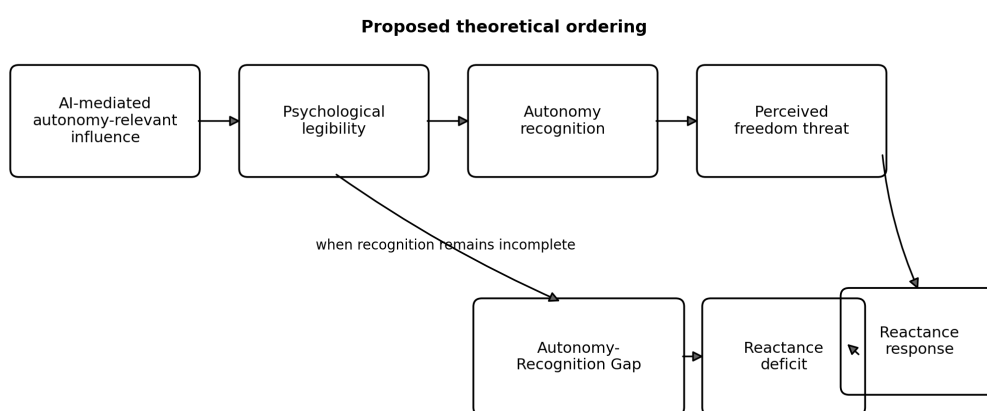
Table 2 Forms of AI-Mediated Influence and Recognition Problems

<i>Form of AI-mediated influence</i>	<i>Autonomy relevance</i>	<i>Recognition problem</i>	<i>Information or control that may improve evaluation</i>
Explicit recommendation	Directs attention while leaving alternatives available	Users may focus on the advice and not on how alternatives were selected	Clear source, purpose, and access to alternatives
Personalized ranking	Changes salience and ordering	The user may not see the baseline set or ranking rule	Explanation of ranking logic and access to unranked options
Adaptive default	Makes one option easier to accept than others	Convenience can obscure the effect of the default	Visible default status and easy reversal
Option filtering	Changes which alternatives are practically visible	Omitted options are difficult to recognize by definition	Disclosure of filtering criteria and recoverable alternatives
Automated overruling	Replaces or reverses a user-selected action	The restriction may be obvious while the basis remains unclear	Reason for intervention and immediate override where appropriate
Delegated agentic action	Transfers part of decision execution to AI	The scope of delegation may expand beyond what the user expected	Clear delegation boundary, audit trail, and revocation

<i>Form of AI-mediated influence</i>	<i>Autonomy relevance</i>	<i>Recognition problem</i>	<i>Information or control that may improve evaluation</i>
Covert restriction	Constrains action without a salient intervention	The user may not represent the restriction as system-imposed	Disclosure that reveals what was constrained and how

Reactance deficit as an interpretive outcome

Reactance theory normally begins once a valued freedom is represented as threatened. In AI-mediated choice, that starting point can conceal an earlier transition. The proposed ordering moves from an autonomy-relevant intervention to recognition of its bearing on choice and then to perceived freedom threat and possible reactance. Picture 1 presents that ordering without claiming established causal effects. Other conditions remain important. The value attached to the freedom, legitimacy of restriction, prior expectations, and personal agency can all change the response. Sittenthaler et al. (2015), Steindl and Jonas (2015), and Steindl et al. (2015) show that restrictions are interpreted through context rather than translated mechanically into reactance.



Picture 1 Autonomy-Recognition Framework

Note Arrows indicate proposed theoretical ordering rather than empirically established causal effects.

Reactance deficit can be interpreted as one possible consequence of poor recognition rather than proof that the user lacks a protective motivation. Görlitz and Rosenthal-von der Pütten (2026) already established the relevance of weak reactance under covert restriction. The present contribution is narrower. It places that outcome within a relation that can also arise outside covert restriction and can vary even when opacity is similar. If users do not connect a system feature to their freedom to act otherwise, the psychological trigger described by reactance theory may remain weak. The argument remains theoretical and does not make recognition the only route into reactance.

Behavior gives weak evidence about autonomy when recognition is unknown. Compliance after a transparent recommendation may reflect considered agreement, but similar behavior can follow convenience, trust, or limited awareness of alternatives. Glikson and Woolley (2020) show that trust in AI depends on several cues. Logg et al. (2019) found substantial willingness to use algorithmic judgment, while Dietvorst et al. (2015) showed that visible error can reverse that willingness. Neither compliance nor resistance therefore identifies autonomous endorsement by itself. The inference is especially weak when the analyst does not know whether the person understood how the system structured the available choice.

Low reactance is equally ambiguous. It may reflect genuine agreement, a legitimate intervention, low importance of the affected freedom, or an autonomy-supportive system. Ryan and Deci (2000) treat autonomy as volitional self-regulation rather than reflexive independence from influence, which prevents a simple equation between autonomy and resistance. A user can recognize an intervention and still endorse it. The concern raised by the Autonomy-Recognition Gap is therefore procedural rather than anti-influence. People need a meaningful basis for evaluating consequential

system behavior. Reactance is useful because it reveals one protective response to perceived threat, but the absence of reactance cannot settle whether the conditions for autonomous endorsement were present.

Theoretical contribution and boundary conditions

The framework changes the unit of explanation. Work on algorithmic paternalism often asks whether AI undermines autonomy, while reactance studies ask how users respond once a threat is perceived. The Autonomy-Recognition Gap sits between those questions and explains why they cannot be collapsed. Hofmann (2026) analyzes how AI can affect conditions of autonomy, and Görlitz and Rosenthal-von der Pütten (2026) show that covert restriction can evade a normal reactance response. The present framework does not compete with either contribution. It separates the system feature, the user's recognition of its relevance, and the later subjective threat. That separation makes it possible to discuss a recognition failure without treating opacity as the psychological state or weak reactance as evidence that the intervention went unnoticed.

Ethical and psychological claims should also remain separate. An ethically questionable intervention can be unobtrusive, while a justified restriction may provoke reactance. Schmauder et al. (2023) argue for interdisciplinary oversight of algorithmic nudging because user acceptance cannot replace ethical evaluation. Rudra et al. (2026) make a related point for LLM-mediated clinical trial participation, where autonomy and consent still matter when communication is persuasive. The Autonomy-Recognition Gap adds a narrower bridge by asking whether users have recognized enough of an intervention to evaluate the steering or authority being exercised. The same separation between algorithmic output and legitimate authority appears in predictive policing, where risk ranking does not by itself establish a sufficient basis for coercive action (Kadir, 2026a).

The construct has clear limits. Low perceived autonomy does not establish a recognition gap because that judgment may follow full recognition. Opacity is also insufficient since an opaque system can impose a visible practical restriction. Nor is the gap synonymous with manipulation. A user may misunderstand a benign ranking rule whose autonomy implications are minor. The construct becomes useful when identifiable features of the choice arrangement plausibly matter for choosing otherwise and recognition of that relevance is incomplete. Keeping this threshold modest prevents the idea from expanding until almost every imperfectly understood digital system is treated as invisible paternalism.

User capability is another boundary. The same interface need not create the same degree of recognition across people. Domain expertise, digital literacy, familiarity with recommender systems, dependence on a platform, and prior expectations can alter what a person infers from identical information. The framework does not convert these differences into moderators because no empirical model is estimated here. They instead warn against describing recognizability as a fixed property of a system. Psychological legibility is relational. It depends on what the system makes available and what the user can reasonably understand from that material in the decision context.

Table 3 states five propositions without repeating them as five parallel paragraphs. Together they formalize the theoretical claim that intervention and perceived threat are distinct, that psychological legibility can narrow the distance between system features and user recognition, and that low recognition may create conditions for reactance deficit. The propositions also protect the interpretation of compliance. Acceptance of AI guidance should not be called autonomous simply because resistance is absent. Evidence about recognition, meaningful control, and reversibility is needed before that stronger interpretation becomes defensible.

Empirical testing does not require a heavy design. A simple vignette could hold the recommendation constant while varying whether users are told what the system filtered or how they can reverse its action. Measures of recognition should be kept separate from perceived autonomy and reactance. Another study could compare awareness that AI was used with awareness of what AI changed in the choice arrangement. These designs would test the framework more cleanly than a large model that treats trust, satisfaction, personalization, autonomy, and reactance as interchangeable indicators. The conceptual work here is useful precisely because it tells later studies what not to collapse.

Design implications follow the same logic. Transparency is not valuable merely because a disclosure exists. It matters when users can understand consequential filtering, delegation, or restriction well enough to decide whether they accept it. Meaningful control also requires more than a theoretical

option to opt out. A reversal path that is obscure or disproportionately costly can preserve formal choice while weakening practical control. Human-centered AI should therefore be assessed with attention to what users can recognize and contest, not only whether they remain satisfied or continue using the system.

Table 3 Theoretical Propositions of the Autonomy-Recognition Framework

<i>Proposition</i>	<i>Theoretical statement</i>	<i>Interpretive consequence</i>
P1	AI-mediated paternalistic influence need not become a perceived freedom threat	System intervention and subjective threat should be analyzed separately
P2	Lower psychological legibility can widen the Autonomy-Recognition Gap	Opacity can contribute to recognition failure without being identical to it
P3	Meaningful explanation and reversible user control can improve recognition of system influence	Recognition depends on information that is usable for evaluation, not disclosure alone
P4	A wider Autonomy-Recognition Gap may weaken the route from autonomy-relevant influence to perceived freedom threat	Reactance deficit can be interpreted as one possible downstream outcome of low recognition
P5	Compliance under low-recognition conditions should not be treated as sufficient evidence of autonomous endorsement	Behavioral acceptance requires process-sensitive interpretation

CONCLUSION

Autonomy cannot be inferred from the mere presence of choice when an AI system also shapes which alternatives become visible, actionable, or easy to reverse. The Autonomy-Recognition Gap identifies a specific discrepancy between autonomy-relevant features that can be located in a digital choice arrangement and the user's recognition of how those features bear on freedom. That relation is conceptually different from system opacity, perceived autonomy, and reactance deficit. The distinction matters because a user may know that AI is present yet still misunderstand the practical structure of influence. Equally, a transparent system can impose a visible restriction that leaves little recognition gap. Treating those situations as equivalent would obscure the psychological step that reactance theory requires before a freedom threat can be experienced.

Reactance deficit may follow recognition failure, but it does not prove that autonomy was preserved and it is not a concept introduced here. Evidence already shows weak initial reactance under covert technological restriction and changed autonomy perceptions after disclosure. The framework extends that result by separating intervention, psychological legibility, recognition, and perceived threat without assigning risk scores or causal effects. Until empirical tests are available, compliance remains a weak proxy for autonomous endorsement when users have not been shown to understand what changed or how they could reverse it.

REFERENCES

- Choubey, V., & Chakraborty, D. (2026). How do consumers perceive digital nudging by agentic AI in E-commerce? *Journal of Retailing and Consumer Services*, 93, 104918. <https://doi.org/10.1016/j.jretconser.2026.104918>
- Dietvorst, B. J., Simmons, J. P., & Massey, C. (2015). Algorithm aversion: People erroneously avoid algorithms after seeing them err. *Journal of Experimental Psychology: General*, 144(1), 114–126. <https://doi.org/10.1037/xge0000033>
- Dillard, J. P., & Shen, L. (2005). On the nature of reactance and its role in persuasive health communication. *Communication Monographs*, 72(2), 144–168. <https://doi.org/10.1080/03637750500111815>
- Duane, J.-N., Ericson, J., & McHugh, P. (2025). Digital nudges: A systematic narrative review and taxonomy. *Behaviour & Information Technology*, 44(13), 3250–3270. <https://doi.org/10.1080/0144929X.2024.2440116>
- Formosa, P. (2021). Robot autonomy vs. human autonomy: Social robots, artificial intelligence (AI), and the nature of autonomy. *Minds and Machines*, 31, 595–616. <https://doi.org/10.1007/s11023-021-09579-2>

- Glikson, E., & Woolley, A. W. (2020). Human trust in artificial intelligence: Review of empirical research. *Academy of Management Annals*, 14(2), 627–660. <https://doi.org/10.5465/annals.2018.0057>
- Görlitz, M., & Rosenthal-von der Pütten, A. M. (2026). Technology paternalism and the reactance deficit—Are our natural protective mechanisms failing us? *Computers in Human Behavior: Artificial Humans*, 8, 100290. <https://doi.org/10.1016/j.chbah.2026.100290>
- Hofmann, B. (2026). Artificial autonomy and algorithmic paternalism: AI shaping human autonomy and decision-making. *Frontiers in Artificial Intelligence*, 9, 1860239. <https://doi.org/10.3389/frai.2026.1860239>
- Jaakkola, E. (2020). Designing conceptual articles: Four approaches. *AMS Review*, 10, 18–26. <https://doi.org/10.1007/s13162-020-00161-0>
- Jesse, M., & Jannach, D. (2021). Digital nudging with recommender systems: Survey and future directions. *Computers in Human Behavior Reports*, 3, 100052. <https://doi.org/10.1016/j.chbr.2020.100052>
- Kadir, Z. K. (2026a). Dapatkah pembunuhan diprediksi? Apa yang diprediksi, risiko kesalahan, dan batas penggunaan algoritma (algorithmic pre-crime). *Triwikrama: Jurnal Ilmu Sosial*, 12(4). <https://doi.org/10.2238/xj4x0s47>
- Kadir, Z. K. (2026b). From delusion to lethal action: A criminological analysis of AI chatbots and the conversational production of homicidal intent. *Jurnal Ilmu Sosial dan Humaniora*, 2(3), 3177–3190. <https://doi.org/10.63822/a25za715>
- Kadir, Z. K. (2026c). Old explanations, new crimes? Testing the portability of criminological theory in the digital age. *Al-Zayn: Jurnal Ilmu Sosial & Hukum*, 4(4), 20463–20475. <https://doi.org/10.61104/alz.v4i4.12119>
- Li, Z., & Shi, J. (2026). Message effects on psychological reactance: Meta-analyses. *Human Communication Research*, 52(1), 38–52. <https://doi.org/10.1093/hcr/hqaf016>
- Logg, J. M., Minson, J. A., & Moore, D. A. (2019). Algorithm appreciation: People prefer algorithmic to human judgment. *Organizational Behavior and Human Decision Processes*, 151, 90–103. <https://doi.org/10.1016/j.obhdp.2018.12.005>
- Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. *Big Data & Society*, 3(2), 1–21. <https://doi.org/10.1177/2053951716679679>
- Oh, J., Nah, S., & Yang, Z. D. (2025). How autonomy of artificial intelligence technology and user agency influence AI perceptions and attitudes: Applying the theory of psychological reactance. *Journal of Broadcasting & Electronic Media*, 69(3), 161–182. <https://doi.org/10.1080/08838151.2025.2485319>
- Rains, S. A. (2013). The nature of psychological reactance revisited: A meta-analytic review. *Human Communication Research*, 39(1), 47–73. <https://doi.org/10.1111/j.1468-2958.2012.01443.x>
- Reynolds-Tylus, T., & Martinez Gonzalez, A. (2026). Artificial intelligence and psychological reactance: Examining responses to health recommendations attributed to a human or AI source. *Review of Communication*, 1–17. <https://doi.org/10.1080/15358593.2026.2633198>
- Rochi, M., Rauschnabel, P. A., Renner, K.-H., & Ivens, B. S. (2024). Technology paternalism: Development and validation of a measurement scale. *Psychology & Marketing*, 41(5), 1172–1188. <https://doi.org/10.1002/mar.21971>
- Rosenberg, B. D., & Siegel, J. T. (2018). A 50-year review of psychological reactance theory: Do not read this article. *Motivation Science*, 4(4), 281–300. <https://doi.org/10.1037/mot0000091>
- Rudra, P., Balke, W.-T., Kacprowski, T., Ursin, F., & Salloch, S. (2026). Algorithmic nudging for clinical trial participation: Autonomy and consent in the era of large language models. *Research Ethics*, 1–16. <https://doi.org/10.1177/17470161261451591>
- Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist*, 55(1), 68–78. <https://doi.org/10.1037/0003-066X.55.1.68>
- Schmauder, C., Karpus, J., Moll, M., Bahrami, B., & Deroy, O. (2023). Algorithmic nudging: The need for an interdisciplinary oversight. *Topoi*, 42, 799–807. <https://doi.org/10.1007/s11245-023-09907-4>

- Shneiderman, B. (2020). Human-centered artificial intelligence: Reliable, safe & trustworthy. *International Journal of Human-Computer Interaction*, 36(6), 495–504. <https://doi.org/10.1080/10447318.2020.1741118>
- Sittenthaler, S., Steindl, C., & Jonas, E. (2015). Legitimate vs. illegitimate restrictions—A motivational and physiological approach investigating reactance processes. *Frontiers in Psychology*, 6, 632. <https://doi.org/10.3389/fpsyg.2015.00632>
- Steindl, C., & Jonas, E. (2015). The dynamic reactance interaction—How vested interests affect people's experience, behavior, and cognition in social interactions. *Frontiers in Psychology*, 6, 1752. <https://doi.org/10.3389/fpsyg.2015.01752>
- Steindl, C., Jonas, E., Sittenthaler, S., Traut-Mattausch, E., & Greenberg, J. (2015). Understanding psychological reactance: New developments and findings. *Zeitschrift für Psychologie*, 223(4), 205–214. <https://doi.org/10.1027/2151-2604/a000222>
- Torraco, R. J. (2005). Writing integrative literature reviews: Guidelines and examples. *Human Resource Development Review*, 4(3), 356–367. <https://doi.org/10.1177/1534484305278283>
- Torraco, R. J. (2016). Writing integrative literature reviews: Using the past and present to explore the future. *Human Resource Development Review*, 15(4), 404–428. <https://doi.org/10.1177/1534484316671606>
- Weinmann, M., Schneider, C., & vom Brocke, J. (2016). Digital nudging. *Business & Information Systems Engineering*, 58(6), 433–436. <https://doi.org/10.1007/s12599-016-0453-1>