Data Sharing with Generative AI: Privacy Calculus, Trust, and Cybersecurity Behaviour in Germany from a Multilateral Security Perspective

Mohammad Mafizul Islam[0000-0001-7623-8988]

Goethe University Frankfurt, Frankfurt am Main, Germany

mafizul.islam@m-chair.de

Abstract. Generative artificial intelligence has entered everyday work and personal life faster than most researchers expected, yet privacy decision-making in such settings is still understood through frameworks designed for bilateral exchanges. The study develops a theoretical framework that treats data sharing with generative AI tools as a multilateral act with consequences for users, providers, and uninvolved third parties. Drawing on privacy calculus theory, trust research, cybersecurity awareness research, and Nissenbaum's contextual integrity, the framework introduces multilateral risk perception as a new construct, defined through recognition of presence, recognition of absent consent, and recognition of onward flow. The paper presents the full research design of a sequential explanatory mixed methods study set in Germany: a quota-based survey of 500 adults to be analysed with PLS-SEM, followed by twenty think-aloud interviews. Because data collection has not yet begun, the contribution lies in the framework, the hypotheses, and the development of the survey instrument, which is reported in full. Germany forms the national setting because it pairs strong data protection law with rapid AI adoption.

Keywords: Privacy calculus · Generative AI · Trust in AI · Cybersecurity awareness · Multilateral security · Contextual integrity · GDPR · Mixed methods · Germany.

1 Introduction

Generative artificial intelligence (AI) has moved into everyday work and personal life at a pace that has caught most researchers off guard. The implications for privacy, however, remain unsettled. When a user types a prompt into a large language model, the data does not stay between the user and the provider. It can flow into the provider's training infrastructure. It can reach third parties whose names appear in the prompt. It can carry proprietary information belonging to an employer that never agreed to the exchange. A single act of disclosure sets several information flows in motion at once, and most of the affected parties have no seat at the table.

Privacy research has long recognised that third parties gain access to personal data, whether through data brokers, advertising networks, or government requests. The dominant decision-theoretic frameworks, however, and above all privacy calculus, have been formulated as bilateral exchanges in which one party weighs the benefits of disclosure against risks to itself [12]. The framing was defensible when the object of study was an e-commerce transaction. Generative AI strains it, because the risks of a disclosure no longer fall on the discloser alone, and because the discloser cannot see where the data goes.

A second feature sharpens the problem. In most online services, a user can close an account, request deletion, and reasonably expect the matter to end there. Once data enters the training pipeline of a generative model, it can become part of the model's weights. Carlini et al. demonstrated that verbatim training data can be extracted from large language models [8], so the concern is practical, and removal after the fact is technically difficult and often impossible. The cost of disclosure is then no longer a temporary loss of control. It is a permanent change in the status of the data itself.

The study responds to these conditions with a theoretical framework and a full research design. The study asks why German users share sensitive data with generative AI tools despite widespread privacy concern, and it approaches the question from the multilateral security tradition as articulated by Rannenberg [30]. The paper makes three contributions. It extends privacy calculus to a setting marked by algorithmic opacity and irreversibility, and treats the bounded rationality of disclosure decisions as a distortion of the calculus rather than a reason to abandon it. It introduces multilateral risk perception, a construct that captures whether users recognise the consequences of their disclosures for parties beyond the immediate exchange, and positions it alongside Nissenbaum's contextual integrity [28] as the behavioural counterpart to a normative theory. It presents the complete design and instrument of a sequential explanatory mixed methods study, including draft items for the new construct, so that the empirical phase can proceed on an openly specified footing.

The paper proceeds as follows. Section 2 reviews the four bodies of literature the framework draws on and states the research gap. Section 3 develops the framework, defines the new construct, and sets out the six hypotheses. Section 4 describes the research design and the ethical safeguards. Section 5 documents the

survey instrument. Section 6 discusses theoretical position, regulatory relevance, and limitations. Section 7 concludes.

2 Related Work

2.1 Privacy Calculus and Its Limits under Uncertainty

Privacy calculus theory holds that people decide about disclosure the way they decide about other trades: they weigh what they expect to gain against what they expect to lose. Dinev and Hart established the model for e-commerce transactions and showed that perceived benefits can outweigh privacy risk beliefs, with trust operating as a countervailing force [12]. A meta-analysis by Baruh, Secinti and Cemalcilar later confirmed across 166 studies that privacy concerns relate negatively to disclosure and to the use of online services, though the effects are modest in size [5].

Modest effects are part of a larger puzzle. Users claim to be very concerned about their privacy yet undertake little to protect their personal data, a gap that has come to be called the privacy paradox [4]. Two readings of the paradox matter here. The first reading treats it as a measurement artefact. Dienlin and Trepte showed that when privacy concerns, attitudes and intentions are modelled as separate steps, concerns shape behaviour indirectly through attitudes and intentions rather than failing to predict it at all [11]. The second reading treats the paradox as a symptom of bounded rationality. Acquisti, Brandimarte and Loewenstein organised the behavioural evidence around three themes: people are uncertain about the consequences of privacy-relevant behaviour and about their own preferences, their concern is context dependent, and their concern is malleable by commercial and governmental actors [1]. Kokolakis, reviewing the debate, notes that privacy calculus rests on the assumption of rational agents in the economic sense, while behavioural research shows disclosure decisions to be shaped by cognitive biases and heuristics, made in limited time and with incomplete information [21].

Bounded rationality shapes the calculus rather than dissolving it. If users cannot know how a generative AI provider stores, reuses, or reproduces their data, the risk side of the calculus becomes a guess, and the framework treats that uncertainty as a measurable feature of the decision rather than a flaw in the theory. Bounded rationality does not remove the trade-off; it distorts the inputs. Algorithmic opacity degrades the quality of the risk estimate, and irreversibility raises the true cost above the perceived cost. Both distortions generate predictions. Users who understand the opacity and the irreversibility should weigh risk differently from users who do not, and the framework in Section 3 builds that difference into the model through cybersecurity awareness and multilateral risk perception. Kokolakis also recommends that future research rely on evidence of actual behaviour rather than self-reports [21], a recommendation the vignette and think-aloud elements of the design take up.

2.2 Trust in AI

Trust enters the calculus as the mechanism that lets people act under uncertainty they cannot resolve. Mayer, Davis and Schoorman define trust as the willingness of a party to be vulnerable to the actions of another party, and trace trustworthiness to three factors: ability, benevolence, and integrity [27]. The triad was developed for organisational relationships, and its transfer to machines is not automatic. Choung, David and Ross examined trust in AI within the technology acceptance model across two studies, one with students and one with a representative sample of the US population [9]. Factor analysis yielded two dimensions rather than three: a human-like dimension formed by the benevolence and integrity items, and a functionality dimension formed by the ability items. Both dimensions raised usage intention indirectly, through perceived usefulness and attitude, and the functionality dimension carried the greater total effect. Trust in what a system can do, in other words, moves adoption more than trust in the system's character, though both matter.

The German situation gives the trust question its urgency. The 2025 global study by Gillespie and colleagues reports that only 32 per cent of Germans are willing to trust AI systems, against a global average of 46 per cent, placing Germany among the least trusting of the advanced economies surveyed. At the same time 71 per cent of Germans believe AI regulation is needed while only 31 per cent consider current safeguards adequate [18]. Low trust paired with strong demand for regulation makes the mediating role of trust an empirical question rather than a settled premise.

2.3 Cybersecurity Awareness as a Distinct Construct

A tempting simplification would treat security awareness as one more expression of privacy concern. The German evidence speaks against it. Biselli and Reuter surveyed 1,219 German private users and found privacy behaviour and security behaviour to be only weakly correlated and to follow different demographic patterns: younger and less educated respondents showed less security behaviour, while no such differences appeared for privacy behaviour, and political ideology influenced neither [7]. The two behaviours are psychologically distinct, so a model that measures only privacy concern will miss the security side.

Generative AI adds threats for which general security intuitions give little guidance. Greshake and colleagues demonstrated indirect prompt injection against deployed LLM-integrated applications, in which instructions hidden in retrieved content redirect the model's behaviour [19]. Carlini and colleagues extracted verbatim training data, including personal information, from a large language model [8]. Awareness of threats of this kind, alongside the concept of training data memorisation itself, forms what the framework calls AI-specific cybersecurity awareness. Such awareness is not a by-product of general privacy attitudes. It is a separate variable, and Section 3 assigns it a moderating role.

2.4 Contextual Integrity and Multilateral Information Flows

Privacy calculus explains disclosure as a trade of benefit against risk, yet it says little about why a particular flow of information should trouble anyone in the first place. Contextual integrity fills that gap. Nissenbaum proposes it as a benchmark for privacy, and its central claim is that privacy depends on whether a flow of information suits the norms of the context in which it happens, and not on whether the information is secret or sensitive in itself [28]. A home address given to a delivery driver is appropriate; the same address passed to an employer is not, even though the address has not changed.

Two kinds of norm define each context. Norms of appropriateness fix what information belongs in a setting, and norms of distribution fix how that information may travel onward. Contextual integrity is kept when both hold and broken when either fails [28]. Nissenbaum is firm that no setting falls outside these norms, and that the idea of a space where anything goes is a fiction. Whether a disclosure counts as a violation then turns on the context, the type of information, the roles of those who receive it, their relationship to the person concerned, and the terms of any onward sharing [28].

Those parameters map closely onto data sharing with generative AI. When an employee pastes a client record into a chat assistant to draft a reply, the information leaves the context it was gathered in and enters the training and processing pipeline of a third party. The data may be accurate and even mundane, so older tests based on secrecy register no problem. Contextual integrity registers one at once, because the flow has crossed a boundary the original context did not sanction. The framework therefore gives a clear account of why multilateral flows trouble people even when nothing confidential is exposed.

Contextual integrity has also been given a formal treatment. Barth, Datta, Mitchell and Nissenbaum express its norms as rules in a logic of information transmission [3]. The formalisation shows the theory can inform the design of systems and the drafting of policy, a point that matters for the regulatory reading in Section 6. Martin and Nissenbaum further show that measuring privacy without context produces confounded results, and they use factorial vignettes to expose the effect of contextual variables on privacy judgements [25], a methodological precedent the design follows.

Contextual integrity and the construct introduced here work as a pair. Contextual integrity is a normative account: it states which flows are appropriate and which breach the norms of a context. Multilateral risk perception is a psychological one: it measures whether users themselves recognise that a disclosure carries consequences for parties beyond the immediate exchange. A user may breach contextual integrity without noticing, precisely because the breach is invisible to the secrecy-based intuitions that still guide much everyday judgement. The gap between the two is where the empirical contribution of this work lies.

2.5 The Research Gap

Each of the four streams above is well developed on its own, and several constructs already gesture at the multilateral problem. Biczók and Chia describe interdependent privacy, in which one user's app permissions expose another user's data [6]. Marwick and boyd describe networked privacy, in which teenagers manage information that flows through social ties they do not control [26]. Petronio's communication privacy management treats disclosed information as jointly owned and governed by negotiated rules [29]. None of these constructs, however, measures an individual's perception of the outward consequences of their own disclosure to an AI system, and none has been integrated with privacy calculus, trust, and security awareness in a single behavioural model. A recent systematic review of conversational AI research confirms that privacy, security, and trust perceptions are studied largely in separation [23]. No existing study, to the

author's knowledge, places these variables within a multilateral security framework and tests them against data-sharing behaviour with generative AI. The study is designed to close that gap.

3 Conceptual Framework and Hypotheses

3.1 Multilateral Security as an Organising Lens

Multilateral security treats every party in a communication system as a holder of legitimate security interests and, at the same time, as a potential attacker, and it aims to balance those competing requirements rather than privilege one side [30]. Rannenberg notes that early security models assumed it was clear who had to be protected against whom, and that the typical conflict runs between the wish for privacy and the interest in cooperation [30]. Generative AI reproduces that conflict at the level of individual behaviour. The user seeks the cooperation benefit of a capable assistant; the provider seeks data; third parties named in prompts seek nothing and risk much. What the provider receives does not stay transient, since verbatim training data, including personal information, can later be extracted from a deployed model [8]. A multilateral lens keeps all of these parties in view, and it motivates the central question of the framework: do users themselves see the other parties at all?

3.2 Multilateral Risk Perception

Multilateral risk perception is defined here as the degree to which a user recognises that a disclosure to a generative AI system carries consequences for parties other than the user and the provider. The definition has three components, and separating them matters because they are likely to behave differently in the data.

The first component is recognition of presence. Prompts routinely contain information about people who are not party to the exchange. A request to redraft an email carries the name and situation of its recipient. A request to summarise meeting notes carries whatever colleagues said in that meeting. A request to check a contract carries the terms a client negotiated in confidence. Recognition of presence is the perception that such material is in the prompt at all, and it cannot be taken for granted, because the user's attention rests on the task rather than on the composition of the text they have pasted.

The second component is recognition of absent consent. Even where a user notices that a third party appears in a prompt, they may treat the disclosure as theirs to make. Employees hold information about clients under conditions that permit some uses and forbid others, and the permission to hold is not a permission to forward. Recognition of absent consent is the perception that the affected party has not agreed and, in most cases, cannot be asked.

The third component is recognition of onward flow. Contextual integrity distinguishes norms of appropriateness from norms of distribution [28], and the second is where generative AI differs most sharply from earlier services. A user may accept that a provider holds their prompt while doubting that the content will travel further. Recognition of onward flow is the perception that the disclosure may reach the training pipeline, may surface in outputs to other users, and cannot reliably be withdrawn.

Several established constructs describe the structural conditions that make multilateral exposure possible, and it is worth stating precisely how the construct proposed here differs from each. Table 1 sets out the comparison.

Table 1. Multilateral risk perception compared with related constructs.

Construct	Unit of analysis	Character	Individual-level measure	Source
Interdependent privacy	Technical dependency between users	Structural	No	[6]
Networked privacy	Social ties through which information travels	Structural and social	Partly, through qualitative accounts	[26]
Communication privacy management	Co-ownership of disclosed information	Normative and dyadic	Yes, for disclosure rules	[29]
Contextual integrity	Information flows against context norms	Normative	No, the theory judges flows rather than perceivers	[28]
Multilateral risk perception	The discloser's awareness of consequences for others	Perceptual	Yes, by design	Proposed here

The pattern in the table is the gap the construct fills. Interdependent privacy and networked privacy describe facts about systems and social structures, and both are agnostic about whether anyone notices. Communication privacy management comes closest, since it treats disclosed information as jointly owned and governed by negotiated rules, but its focus rests on the rules two parties agree between themselves

rather than on a discloser's awareness of parties who never negotiated at all. Contextual integrity supplies the normative test, telling us which flows breach the norms of a context, and says nothing about whether the person producing the flow can see the breach. None of the four measures, at individual level, what a person perceives about the outward reach of their own act.

The distinction carries a practical consequence. A user may breach contextual integrity while behaving in a way that feels entirely reasonable, because the intuitions people bring to privacy decisions are built around secrecy, and nothing in a client's name or a colleague's remark is secret. Under a secrecy test, the disclosure passes. Under a contextual test, it fails. The gap between those two judgements is where multilateral risk perception operates, and it explains why an intervention aimed at raising awareness of confidentiality may leave the problem untouched.

Two further reasons support treating perception as a variable rather than assuming it. Kray and Gonzalez find that people apply different weightings when advising others than when choosing for themselves [22], which suggests that the care a person would demand from someone handling their own data is not the care they apply when handling the data of others. If that asymmetry holds in the AI setting, users will systematically underweight third-party risk, and the underweighting will be invisible to any measure that asks only about risks to the respondent. Beyond that, the workplace setting concentrates the problem. Prompts written at work are more likely to contain material about identifiable others than prompts written at home, so the population in which multilateral risk perception matters most is also the population in which it has never been measured.

The construct enters the model as a moderator rather than a direct predictor. The reasoning is that recognising consequences for others does not, on its own, create an intention to withhold. It changes what the perceived risk of a disclosure means. Where multilateral risk perception is low, perceived risk is read as risk to the self, and the calculus proceeds on that narrow basis. Where it is high, the same perceived risk carries additional weight, because the consequences of getting it wrong extend beyond the person making the decision. H5 states the prediction that follows.

One boundary of the model should be acknowledged. AI literacy is included as an antecedent of multilateral risk perception, on the reasoning that a user who understands how a model ingests and reuses text is better placed to notice that a prompt carries other people's data. The path is not separately hypothesised, and the wider set of determinants, which may include professional training, organisational policy, and prior exposure to a data incident, stays outside the scope of the study reported here and forms part of the broader doctoral project.

3.3 The Contextual Scope of the Study

Generative AI is used for homework, therapy-like conversation, code, and cooking. Privacy behaviour is strongly context dependent [21], so no single study should claim to cover all of these settings, and the study does not. German users have also been shown to hold higher privacy concerns and to behave more cautiously online than users in the United States [31], which makes the workplace setting in Germany a demanding test of whether concern translates into restraint. The scope of the study is workplace and semi-professional use by adults in Germany. The choice follows from the framework: workplace use maximises multilateral exposure, because prompts routinely contain information about clients, colleagues, and the employing organisation, none of whom authorised the disclosure. Context is treated as a design variable rather than a constant. The survey's vignettes vary the contextual parameters that contextual integrity identifies as decisive, the type of information, the parties affected, and the terms of onward flow, following the factorial vignette precedent of Martin and Nissenbaum [25]. Findings will therefore speak to the workplace setting directly and to other settings only by explicit, tested variation.

3.4 Structural Model and Hypotheses

The framework integrates the four streams as follows. Privacy calculus supplies the spine: perceived benefits and perceived risks drive the intention to share data with generative AI tools. Trust in AI, modelled as a second-order construct with the two dimensions Choung et al. report [9] and drawn as a single node in Figure 1, mediates the path from perceived risks to intention. AI-specific cybersecurity awareness moderates the risk path, and multilateral risk perception amplifies it. GDPR awareness and AI literacy stand as antecedents. Figure 1 shows the model, with exogenous and endogenous constructs distinguished by border weight.

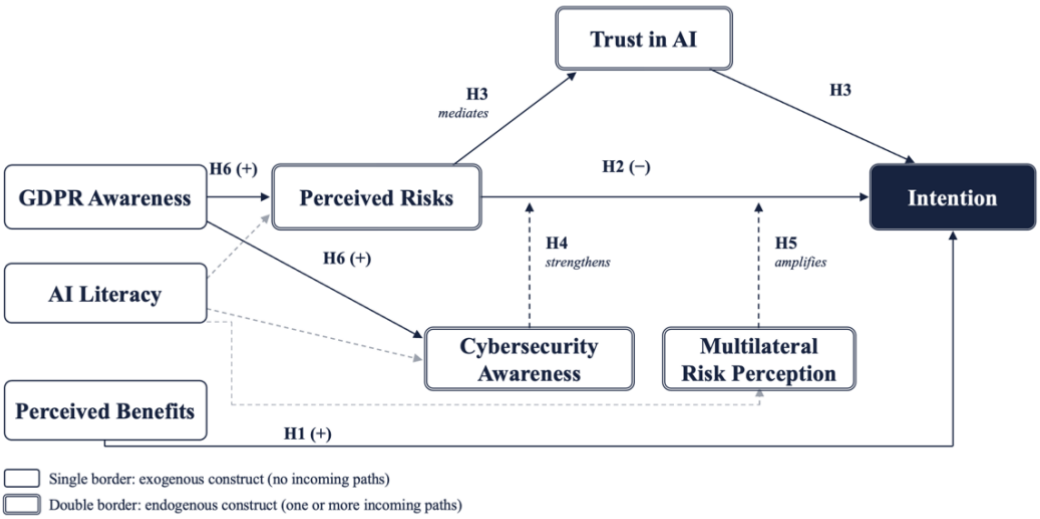

Fig. 1. Structural model of hypothesised relationships. Solid arrows mark hypothesised paths, dashed black arrows mark moderation effects, and dashed grey arrows mark antecedent links that are not separately hypothesised.

One asymmetry in the model is deliberate and should be stated plainly. Perceived risks receives an antecedent path from GDPR awareness while perceived benefits receives none. The reason is theoretical rather than accidental: regulatory knowledge plausibly sharpens a user's sense of what could go wrong without changing their sense of what a tool can do for them. AI literacy enters the model on the same logic, but as an antecedent of multilateral risk perception rather than of perceived risk, since technical understanding bears most directly on whether a user notices that a prompt carries other people's data. Whether benefit perception also responds to those antecedents is a reasonable question, and one the survey data will allow later exploration of, though the model does not predict it.

Six hypotheses give the model its empirical shape, summarised in Table 2. Perceived benefits are expected to raise data-sharing intention (H1), and perceived risks to lower it (H2). Trust in AI is expected to mediate the relationship between perceived risks and intention, so that part of the effect of perceived risks runs through trust (H3). Among users with greater AI-specific cybersecurity awareness, the negative relationship between perceived risk and intention should be stronger (H4). Multilateral risk perception is expected to amplify the negative effect of perceived risk on intention (H5). GDPR awareness, as an antecedent, is expected to raise both perceived risk and cybersecurity awareness (H6). Each hypothesis is directional and each is falsifiable in the survey data. H3 in particular runs against the possibility, raised by the German trust figures [18], that trust and use have decoupled, in which case the mediation would fail to appear.

Table 2. Summary of hypotheses. The Basis column gives the conceptual grounding for each path rather than prior estimates of effect size.

	Path	Type	Direction	Basis
H1	Perceived benefits to intention	Direct	Positive	[5, 12]
H2	Perceived risks to intention	Direct	Negative	[5, 12]
H3	Perceived risks to trust in AI to intention	Mediation	Indirect	[9, 27]
H4	Cybersecurity awareness on the risk path	Moderation	Strengthening	[7, 19]
H5	Multilateral risk perception on the risk path	Moderation	Amplifying	[22, 28]
H6	GDPR awareness to risks and to awareness	Antecedent	Positive	[15, 31]

4 Research Design

4.1 A Sequential Explanatory Mixed Methods Design

The study follows a sequential explanatory mixed methods design [10]. A quantitative phase will establish what the relationships between the constructs are, and a qualitative phase will then explain why they hold, drawing its participants from the survey sample so that the two phases speak about the same population. Integration will follow the joint display technique described by Fetters, Curry and Creswell, in which quantitative and qualitative findings are arrayed against each other to generate meta-inferences [17]. The design answers the recommendation that privacy research move beyond self-report [21] in two ways: the survey embeds behaviour-approximating vignettes, and the interviews will observe reasoning during actual tool use.

4.2 Phase One: Survey

The first phase is a quota-based survey of 500 German adults, sampled on age, gender, education, and federal state so that the achieved sample approximates the adult population on those characteristics. Guidance on minimum sample size in PLS-SEM ties the requirement to the smallest path coefficient a study aims to detect and to the power sought [20], which places the minimum for the paths anticipated here in the region of 300. Interaction effects of the kind H4 and H5 predict carry lower statistical power than direct effects, so the target is set above that minimum. All constructs in the model will be measured with multi-item scales documented in Section 5. Beyond the scales, respondents will work through a set of factorial vignettes describing realistic workplace situations. The vignettes vary three factors that contextual integrity identifies as decisive and that can be manipulated independently of one another: the party affected by the disclosure (client, colleague, or the employing organisation), the stated data-handling terms of the tool (used for training, not used for training, or not stated), and whether the third party is identifiable in the pasted text or has first been stripped of identifying detail. Crossing these factors yields eighteen combinations, which exceed what any one respondent can reasonably assess, so each participant will receive six vignettes under a balanced allocation with the order of vignettes and of the varied elements randomised. Design follows the recommendations of Aguinis and Bradley for experimental vignette methodology [2] and the factorial precedent of Martin and Nissenbaum [25].

> ***Example vignette.*** You are drafting a reply to a client complaint and you are short of time. You consider pasting the client's original message, which contains their name, account number, and a description of their financial situation, into a general-purpose AI assistant so that it can suggest a reply. The provider's terms state that conversations may be used to improve the service. How likely are you to paste the message? Across versions of this scenario the affected party shifts between a client, a colleague, and the employing organisation; the tool's stated terms shift between using the data for training, not using it, and saying nothing; and the third party is either named, as above, or reduced to a non-identifying description before pasting.

Vignette responses will approximate behaviour under specified contextual conditions, which pure attitude items cannot. Each respondent's disclosure ratings will be averaged across their six vignettes to form the behavioural outcome used in the structural model, giving a more reliable measure than a single rating would, while the individual vignette ratings will be analysed separately at vignette level to estimate the effect of each contextual factor. The three-item disclosure-intention scale in Table 3 is a convergent check on the vignette-derived outcome.

4.3 Phase Two: Interviews

The second phase consists of twenty semi-structured interviews of roughly 60 minutes with participants drawn from the survey sample, selected purposively to cover the risk profiles the quantitative phase identifies. Twenty is a planning figure rather than a fixed quota, and the number will be revised upward if coding suggests that new themes are still emerging when the set is complete. Each interview will include a think-aloud episode in which the participant works through a realistic task with a generative AI tool and speaks their reasoning as they go. The protocol is designed to capture hesitation, justification, and the moment at which a third party's data enters or is withheld from a prompt, which is precisely the moment multilateral risk perception is supposed to govern. Interviews will be transcribed and coded thematically, with the coding frame anchored in the model constructs and open to emergent categories.

4.4 Analytical Strategy

Survey data will be analysed with partial least squares structural equation modelling (PLS-SEM), following Hair et al. [20]. PLS-SEM suits the design for two reasons: the model contains a newly developed construct whose measurement is still maturing, and the aim at this stage is prediction and theory building rather than confirmation of an established covariance structure. The measurement model will be assessed for reliability,

convergent validity, and discriminant validity before the structural model is estimated. Moderation (H4, H5) will be tested through interaction terms and mediation (H3) through bootstrapped indirect effects. Multi-group analysis is planned for one pre-specified grouping variable, intensity of AI use, since a sample of this size cannot support reliable comparison across several groupings at once; differences by age and education will be examined descriptively and reported as exploratory. Qualitative material will be integrated with the survey results through joint displays [17].

4.5 Ethics and Data Protection

A study about multilateral privacy risk should not create the harm it examines, and the design carries three safeguards to that end. First, no real third-party data will enter any commercial AI system during the research. Think-aloud tasks will use synthetic material prepared by the researcher, so that participants reason about a realistic scenario without exposing an actual client or colleague. Participants will be asked not to enter genuine work material, and any that appears in a transcript will be removed before analysis.

Second, participation will rest on informed consent under Article 6(1)(a) of the GDPR [15], with the right to withdraw at any point and without giving reasons. Survey responses will be collected through a panel provider under pseudonymous identifiers, and no direct identifiers will be transferred to the research team. Interview audio will be recorded only with separate consent, transcribed, and then deleted, with transcripts pseudonymised at the point of transcription. All research data will be stored on servers operated by Goethe University Frankfurt.

Third, approval will be obtained from the responsible ethics board of Goethe University Frankfurt before any data collection begins, and the hypotheses, measurement model, and analysis plan will be preregistered so that confirmatory and exploratory findings remain distinguishable in later reporting.

5 Instrument Development

The survey instrument combines adapted and newly developed scales. All items use a seven-point Likert format anchored from strongly disagree to strongly agree, with the exception of the GDPR awareness and AI literacy scales, which use knowledge items scored as correct or incorrect. Table 3 documents the full instrument.

Table 3. Operationalisation of model constructs.

Construct	Working definition	Adapted from	Items
Perceived benefits	Expected gains from sharing data with a generative AI tool	[12]	4
Perceived risks	Expected losses to the user from that disclosure	[12]	4
Trust in AI, functionality	Belief that the system performs competently	[9, 27]	4
Trust in AI, human-like	Belief in the benevolence and integrity of the system	[9, 27]	4
Cybersecurity awareness	Knowledge of AI-specific threats and protective practice	[7, 8, 19]	6
Multilateral risk perception	Recognition of consequences for parties beyond the exchange	New, developed here	8
GDPR awareness	Knowledge of rights and obligations under the regulation	[15]	6
AI literacy	Understanding of how generative models process input	New, developed here	5
Data-sharing intention	Stated likelihood of disclosing in a given scenario	[12]	3

Multilateral risk perception requires new items, and their development will follow the construct measurement and validation procedure of MacKenzie, Podsakoff and Podsakoff [24]: definition of the construct, generation of items against that definition, expert review, cognitive pretesting, and pilot testing with exploratory factor analysis. Table 4 gives the draft item pool, organised by the three components set out in Section 3.2 and closing with a summary item.

Table 4. Draft items for multilateral risk perception, by component.

	Component	Draft item
1	Presence	When I use an AI tool for work, I think about whether my prompt contains information about other people.
2	Presence	I notice when text I paste into an AI tool includes details about colleagues or clients.
3	Consent	People whose information appears in my prompts have not agreed to that use.

4	Consent	It is mine alone to decide whether information about others goes into an AI tool. (reversed)
5	Onward flow	Information I enter about other people may be stored by the AI provider.
6	Onward flow	Information I enter about other people may appear in answers given to other users.
7	Onward flow	Once information about another person enters an AI system, it cannot be removed.
8	General	My use of AI tools can create privacy risks for people other than myself.

Three properties of the item pool deserve comment. The pool deliberately mixes items that ask about attention (1 and 2) with items that ask about belief (5 to 7), because the components are conceptually distinct and may not load on a single factor. Should the exploratory analysis return a multi-factor solution, the construct will be modelled as second-order rather than collapsed. Item 4 is reverse-coded to reduce acquiescence bias. Item 8 functions as a summary indicator and provides a check on the coherence of the remaining seven.

Discriminant validity is the central pretest concern, since the construct earns its place only if it measures something its neighbours do not. Three tests will be applied. The items must load more strongly on their own construct than on general privacy concern, which is the nearest established measure. The Fornell-Larcker criterion and the heterotrait-monotrait ratio will be assessed against both trust dimensions and against cybersecurity awareness [20]. And the qualitative phase will provide a face-validity check, since participants who score high on the scale should show the corresponding reasoning in their think-aloud protocols. A pilot with approximately 100 respondents, drawn separately from the main sample, will precede the main survey, and items that fail the criteria will be revised or dropped before the instrument is fielded.

6 Discussion

6.1 Theoretical Position

The theoretical position of the framework can be stated plainly. Behavioural research has questioned whether privacy calculus presumes a rationality that conditions of algorithmic opacity make untenable [1, 21]. The framework keeps the calculus as the spine and treats opacity and irreversibility as measurable distortions of it, with awareness constructs governing how strongly each user's calculus responds to risk. Contextual integrity strengthens rather than replaces that architecture. It explains, normatively, why multilateral flows are violations, while multilateral risk perception measures, psychologically, whether users see them. Should the data show that perceived risk barely moves disclosure even among highly aware users, the calculus reading would weaken and a norm-based or habit-based account would become more plausible. Stating the prediction in that falsifiable form is the purpose of setting out the framework before the data are collected.

6.2 Regulatory Relevance

Germany operates under the GDPR [15] and the EU AI Act [16], two regimes with different aims and stages of implementation, and the European Commission's 2025 Digital Omnibus proposal to simplify digital regulation [14] is under debate as the study is designed. A second proposal in the same package, the Digital Omnibus on AI [13], would postpone the entry into application of the high-risk provisions of the AI Act, and the pressure to conclude it before 2 August 2026 leaves the timing of those obligations unsettled at the very moment the study is designed. The framework connects to that setting in both directions. In one direction, GDPR awareness enters the model as an antecedent, so the study measures whether regulatory knowledge changes perception and behaviour at the individual level, evidence that speaks to whether the law's protective intent reaches the people it protects. Wachter, Mittelstadt and Russell argue that the GDPR falls short of granting a meaningful right to explanation for automated decisions [33], and Veale and Zuiderveen Borgesius identify enforcement and clarity problems in the AI Act's design [32]. Behavioural evidence on what users actually understand complements both critiques.

In the other direction, the multilateral framing exposes a structural limit that simplification debates tend to pass over. Consent-centred instruments assume a data subject who can be asked. In a multilateral flow, the parties most exposed are precisely those who cannot be asked, because they are absent from the interaction and often unaware it took place. Neither the GDPR's consent architecture nor the AI Act's transparency obligations, which attach to providers and deployers rather than to individual users composing prompts, addresses that case directly. The formal reading of contextual integrity [3] suggests one route forward, since appropriateness norms can be expressed as checkable rules about which flows a context

permits, and rules of that kind do not depend on obtaining consent from an absent party. Whether regulatory simplification preserves or erodes room for such an approach is an open question.

6.3 Limitations

The paper presents a framework and a design, and no data. Every hypothesis remains a prediction, and the measurement quality of the new construct is a promise the pilot has yet to keep. The scope decision limits generalisation: findings will concern workplace and semi-professional use in Germany, and travel to other contexts or jurisdictions only as hypotheses. Vignettes approximate behaviour without reproducing its stakes, and think-aloud protocols can alter the reasoning they record. The trust measure inherits a limitation noted by its authors, since path models built on correlations cannot secure causal direction [9]. Finally, the decision to use synthetic material in the think-aloud sessions protects third parties but removes the real consequences that may drive genuine hesitation, so the qualitative phase observes reasoning under reduced stakes. These limits are stated now so that the empirical phase can be judged against them.

7 Conclusion

The paper set out to give a behavioural account of data sharing with generative AI that keeps every affected party in view. It extended privacy calculus into conditions of opacity, irreversibility, and bounded rationality, introduced multilateral risk perception as the psychological counterpart to contextual integrity's normative claims, and specified a two-phase mixed methods design with a documented instrument, set in a country whose combination of strict data protection law and rapid AI adoption makes the tensions visible. The next step is data. The survey and interviews will show whether users see the parties their prompts expose, and whether seeing them changes what users share.

Acknowledgments. The research is conducted at the Chair of Mobile Business and Multilateral Security, Goethe University Frankfurt.

Disclosure of Interests. The author has no competing interests to declare that are relevant to the content of this article.

References

1. Acquisti, A., Brandimarte, L., Loewenstein, G.: Privacy and human behavior in the age of information. Science **347**(6221), 509–514 (2015). https://doi.org/10.1126/science.aaa1465
2. Aguinis, H., Bradley, K.J.: Best practice recommendations for designing and implementing experimental vignette methodology studies. Organizational Research Methods **17**(4), 351–371 (2014). https://doi.org/10.1177/1094428114547952
3. Barth, A., Datta, A., Mitchell, J.C., Nissenbaum, H.: Privacy and contextual integrity: framework and applications. In: 2006 IEEE Symposium on Security and Privacy, pp. 184–198. IEEE, Berkeley (2006). https://doi.org/10.1109/SP.2006.32
4. Barth, S., de Jong, M.D.T.: The privacy paradox – investigating discrepancies between expressed privacy concerns and actual online behavior – a systematic literature review. Telematics and Informatics **34**(7), 1038–1058 (2017). https://doi.org/10.1016/j.tele.2017.04.013
5. Baruh, L., Secinti, E., Cemalcilar, Z.: Online privacy concerns and privacy management: a meta-analytical review. Journal of Communication **67**(1), 26–53 (2017). https://doi.org/10.1111/jcom.12276
6. Biczók, G., Chia, P.H.: Interdependent privacy: let me share your data. In: Sadeghi, A.-R. (ed.) Financial Cryptography and Data Security. LNCS, vol. **7859**, pp. 338–353. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-39884-1_29
7. Biselli, T., Reuter, C.: On the relationship between IT privacy and security behavior: a survey among German private users. In: Ahlemann, F., Schütte, R., Stieglitz, S. (eds.) Innovation Through Information Systems. WI 2021. LNISO, vol. **47**, pp. 388–404. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-86797-3_26
8. Carlini, N., Tramèr, F., Wallace, E., Jagielski, M., Herbert-Voss, A., Lee, K., Roberts, A., Brown, T., Song, D., Erlingsson, Ú., Oprea, A., Raffel, C.: Extracting training data from large language models. In: 30th USENIX Security Symposium, pp. 2633–2650. USENIX Association (2021)
9. Choung, H., David, P., Ross, A.: Trust in AI and its role in the acceptance of AI technologies. International Journal of Human–Computer Interaction **39**(9), 1727–1739 (2023). https://doi.org/10.1080/10447318.2022.2050543
10. Creswell, J.W., Plano Clark, V.L.: Designing and Conducting Mixed Methods Research. 3rd edn. Sage, Thousand Oaks (2018)
11. Dienlin, T., Trepte, S.: Is the privacy paradox a relic of the past? An in-depth analysis of privacy attitudes and privacy behaviors. European Journal of Social Psychology **45**(3), 285–297 (2015). https://doi.org/10.1002/ejsp.2049

12. Dinev, T., Hart, P.: An extended privacy calculus model for e-commerce transactions. Information Systems Research **17**(1), 61–80 (2006). https://doi.org/10.1287/isre.1060.0080
13. European Commission: Digital Omnibus on AI Proposal. COM(2025) 836 final, 2025/0359(COD), Brussels, 19 November 2025
14. European Commission: Digital Omnibus Regulation Proposal. COM(2025) 837 final, 2025/0360(COD), Brussels, 19 November 2025
15. European Parliament and Council: Regulation (EU) 2016/679 (General Data Protection Regulation). Official Journal of the European Union L119, 1–88 (2016)
16. European Parliament and Council: Regulation (EU) 2024/1689 (Artificial Intelligence Act). Official Journal of the European Union L, 2024/1689 (2024)
17. Fetters, M.D., Curry, L.A., Creswell, J.W.: Achieving integration in mixed methods designs – principles and practices. Health Services Research **48**(6pt2), 2134–2156 (2013). https://doi.org/10.1111/1475-6773.12117
18. Gillespie, N., Lockey, S., Ward, T., Macdade, A., Hassed, G.: Trust, Attitudes and Use of Artificial Intelligence: A Global Study 2025. The University of Melbourne and KPMG (2025). https://doi.org/10.26188/28822919
19. Greshake, K., Abdelnabi, S., Mishra, S., Endres, C., Holz, T., Fritz, M.: Not what you've signed up for: compromising real-world LLM-integrated applications with indirect prompt injection. In: Proceedings of the 16th ACM Workshop on Artificial Intelligence and Security, pp. 79–90. ACM, Copenhagen (2023). https://doi.org/10.1145/3605764.3623985
20. Hair, J.F., Hult, G.T.M., Ringle, C.M., Sarstedt, M., Danks, N.P., Ray, S.: An introduction to structural equation modeling. In: Partial Least Squares Structural Equation Modeling (PLS-SEM) Using R, pp. 1–29. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-80519-7_1
21. Kokolakis, S.: Privacy attitudes and privacy behaviour: a review of current research on the privacy paradox phenomenon. Computers and Security **64**, 122–134 (2017). https://doi.org/10.1016/j.cose.2015.07.002
22. Kray, L., Gonzalez, R.: Differential weighting in choice versus advice: I'll do this, you do that. Journal of Behavioral Decision Making **12**(3), 207–218 (1999)
23. Leschanowsky, A., Rech, S., Popp, B., Bäckström, T.: Evaluating privacy, security, and trust perceptions in conversational AI: a systematic review. Computers in Human Behavior **159**, 108344 (2024). https://doi.org/10.1016/j.chb.2024.108344
24. MacKenzie, S.B., Podsakoff, P.M., Podsakoff, N.P.: Construct measurement and validation procedures in MIS and behavioral research: integrating new and existing techniques. MIS Quarterly **35**(2), 293–334 (2011). https://doi.org/10.2307/23044045
25. Martin, K., Nissenbaum, H.: Measuring privacy: an empirical test using context to expose confounding variables. Columbia Science and Technology Law Review **18**, 176–218 (2016)
26. Marwick, A.E., boyd, d.: Networked privacy: how teenagers negotiate context in social media. New Media and Society **16**(7), 1051–1067 (2014). https://doi.org/10.1177/1461444814543995
27. Mayer, R.C., Davis, J.H., Schoorman, F.D.: An integrative model of organizational trust. Academy of Management Review **20**(3), 709–734 (1995). https://doi.org/10.5465/amr.1995.9508080335
28. Nissenbaum, H.: Privacy as contextual integrity. Washington Law Review **79**(1), 119–158 (2004)
29. Petronio, S.: Boundaries of Privacy: Dialectics of Disclosure. State University of New York Press, Albany (2002)
30. Rannenberg, K.: Multilateral security – a concept and examples for balanced security. In: Proceedings of the 2000 Workshop on New Security Paradigms, pp. 151–162. ACM (2001). https://doi.org/10.1145/366173.366208
31. Trepte, S., Reinecke, L., Ellison, N.B., Quiring, O., Yao, M.Z., Ziegele, M.: A cross-cultural perspective on the privacy calculus. Social Media and Society **3**(1), 1–13 (2017). https://doi.org/10.1177/2056305116688035
32. Veale, M., Zuiderveen Borgesius, F.: Demystifying the draft EU Artificial Intelligence Act. Computer Law Review International **22**(4), 97–112 (2021). https://doi.org/10.9785/cri-2021-220402
33. Wachter, S., Mittelstadt, B., Russell, C.: Counterfactual explanations without opening the black box: automated decisions and the GDPR. Harvard Journal of Law and Technology **31**(2), 841–887 (2018)