A Dual-Method Investigation of Data Sharing Determinants in GenAI-applications Through Systematic Review and Structural Equation Modelling

Julian Held[1]

Abstract

This paper investigates the determinants influencing individuals' data sharing decisions through a two-pronged approach combining systematic literature review and empirical analysis. First, a systematic review of 53 studies across multiple fields and technologies identifies and synthesizes key determinants into a comprehensive framework, addressing the absence of unified definitions and closing a critical literature gap. Second, an online survey (n=357) examines relationships between privacy concerns, perceived risks and benefits, trust, prior disclosure behavior, and willingness to share personal data specifically in the GenAI context. Using partial least squares structural equation modeling, the study reveals trust significantly reduces privacy concerns, consistent with existing literature. However, trust's effects on data sharing willingness and perceived risks proved statistically insignificant. Surprisingly, both perceived risks and benefits showed negative (though insignificant) relationships with sharing willingness, contradicting traditional assumptions. These unexpected findings are attributed to ChatGPT's still novel technological environment, potential survey framing effects, and the unique nature of AI-driven platforms that may reshape conventional decision-making processes. The study provides evidence of the privacy paradox in ChatGPT contexts, where stated privacy concerns diverge from actual sharing behaviors. This research contributes a comprehensive framework for understanding data sharing determinants, offers insights into privacy decision-making in emerging AI technologies, and provides practical implications for businesses deploying large language models. Despite limitations in sampling methods and external validity, this work advances theoretical understanding and practical applications while establishing foundations for future research in AI-mediated data disclosure behaviors.

Keywords
Data Sharing Determinants, Privacy Paradox, Structural Equation Modeling, Systematic Literature Review

---

[1]University of Neuchâtel, Switzerland

Corresponding Author:
Julian Held, University of Neuchâtel, Institute of Management, Rue A.-L. Breguet 2, 2000 Neuchâtel, Switzerland. Email: julian.held@unine.ch

1 Introduction

In an increasingly connected and digitized world, the use of digital solutions is inherently connected to the collection, storage, and analysis of personal data as a by-product of use (Cichy et al., 2021, p. 1863). Individuals share their personal data with companies to get for instance personalized content, targeted advertisements, improved services and products (De Schaepdrijver et al., 2022; Leppäniemi et al., 2017; Sarathy and Muralidhar, 2006; Treiblmaier and Pollach, 2007). Exemplarily, Internet of Things devices and their manufacturers need unrestricted access to their customers' data as a precondition to create customer value (Cichy et al., 2021, p. 1866). Further, companies benefit from their customers' data from a lot of perspectives. For instance, innovation today is highly driven by customer data, by providing valuable insights for companies (Cichy et al., 2021, p. 1866). Those customer insights also help businesses to make marketing more efficient, predict customers' behaviour, enable price discrimination and help to optimize supply chains and production (Acquisti, 2010, p. 13; De Nijs, 2017; Sagiroglu and Sinanc, 2013, p. 46). Thus, customer data can be a competitive advantage for companies, stressing the importance of better understanding the determinants of data sharing as the key for data collection (De Nijs, 2017, pp. 1-4).

Hereby, previous studies examined challenges and determinants of data sharing in different contexts and technologies. For example Cichy et al. (2021) studying Internet of Things devices, Ayyagari et al. (2011) examining technostress or studies investigating determinants of online privacy concerns and how these affect the privacy protection behaviour of teenagers (Citron, 2006; Daigle and Khan, 2020; Morey et al., 2015; Youn, 2009).

However, besides this variety of conducted studies, what is missing is an overarching and comprehensive framework on the determinants of data sharing, as most studies examine very specific contexts or technologies. Likewise, an evaluation of the importance of these determinants is missing. Thus, the first research question posed is: *What are key determinants in individuals' willingness to share personal data?* To close this literature gap determinants are collected and analysed in an extensive literature review, based on the Prisma (2023) guidelines, covering the most important databases for information systems literature. While several studies stressed the importance of trust or control, others in different contexts stressed for instance, brand familiarity, or the importance of an equity-oriented relationship (Ackermann et al., 2022; Adams and

Freedman, 1976; Leon et al., 2013; Leppäniemi et al., 2017; Morey et al., 2015; Ur et al., 2012, p. 7; Wottrich et al., 2018). Therefore, publications covering a wide range of contexts, fields and technologies, also present very different aspects, viewpoints, and ambiguous determinants' definitions. Thus, one of the key challenges of the first research question is to consider a broad range of relevant studies while only gathering the most important determinants to not overload the final framework. During this process, results are additionally quantified and presented graphically to clarify the selection. Several specifically relevant determinants are then highlighted and further examined in the empirical part of the paper.

Building up on the knowledge of this analysis and selection of determinants, the second research question investigates: What determinants are more relevant in data sharing decisions, and how are they interrelated? For this purpose, an online survey via Qualtrics is conducted to formally test the relevance and interrelations of different determinants in the context of a relatively novel and relevant technology, ChatGPT. A partial least squares structural equation model in smartPLS (2022) subsequently reveals relationships and tests several hypotheses regarding privacy concerns, perceived risks, perceived benefits, trust, prior disclosure behaviour and the willingness to share personal data in ChatGPT. This contributes to the literature by introducing the so far under-investigated context of ChatGPT to the field of data sharing and privacy.

In an era where data is essential for digital operations, this understanding is especially important for both, businesses and governments employing LLM such as ChatGPT. Companies understanding the motivations behind data sharing and how to implement this knowledge into LLM-driven business models or processes, enable a key factor for success by for instance facilitating the customization of products, services, and marketing strategies ultimately enhancing customer satisfaction and loyalty (Acquisti, 2010, p. 13; Cichy et al., 2021, p. 1866; De Nijs, 2017; Sagiroglu and Sinanc, 2013, p. 46). Likewise, building on the insights of this study, businesses will be able to navigate the ethics surrounding data privacy and consent, ensure compliance with data regulations and ultimately build trust with individuals (Chellappa and Sin, 2005, pp. 196–198). The study also helps governments to meet the requirements for their roles as thoughtful guardians of citizen data by understanding why people share data, ultimately allowing effective policies that balance public interest with individual privacy rights. This is

fundamental to avoid, for instance, potential misuse of data and stimulating citizen' trust in the government (see for example Bansal and Nah, 2020; Pongratz, 2023).

Finally, whether it is in the commercial or public sector, understanding the drivers of data sharing is crucial for responsible and effective collection and use of data. Fostering trust and ensuring the ethical implementation and adoption of LLM in our increasingly data-driven world is of great value to the society. Additionally, to these insights, the study also contributes to the ongoing debate regarding the privacy paradox.

2 Theoretical and Technical Foundations

The following sections provide theoretical and technical foundations required for the later parts of the paper. First, the characteristics of data and various values of data sharing for individuals as well as companies are presented. Further, basics of LLM are introduced and their characteristics regarding data sharing. After that, sharing behaviour theories and the determinants of willingness to share personal data are extensively investigated.

2.1 Characteristics of Data

Data refers to unprocessed, actually meaningless character strings, that only transform into information or knowledge once they are organized and thereby given meaning (Mertens et al., 2005, pp. 36–37; Moreira et al., 2018, p. 7). It can be classified on various dimensions, such as data or character type, form of manifestation, format, or purpose (Mertens et al., 2005, pp. 36–37). Character and data types can be then further classified as numerical, alphabetical or alpha-numerical (Mertens et al., 2005, pp. 36–37; Moreira et al., 2018, p. 7). Data formats and forms of manifestation include for instance textual, numerical, visual, auditory, or audio-visual forms (Mertens et al., 2005, pp. 36–37; Moreira et al., 2018, p. 7). Data needs typically to be organized and structured to facilitate analysis, and allow interpretation and extraction of meaningful insights in all forms (Moreira et al., 2018, p. 7). For instance, a simple dataset like a contact list, containing information such as names and telephone numbers, does not provide easily understandable information in its character strings format, until the user opens the file in the according program, such as SQL, thereby organizing the character strings into readable, interpretable information.

Simplified, data can further be classified, besides semi-structured data, into two primary categories, structured and unstructured data (Eberendu, 2016, p. 48). Structured data is the type of data possessing a predetermined organization and format, which enables easier searchability and analysability (Eberendu, 2016, p. 48). It is of a quantitative nature and usually stored in relational databases or data warehouses and can be analysed for instance in relational databases applying SQL (Eberendu, 2016, p. 48). Alternatively, it can be represented in tabular form with rows and columns, such as Excel to depict transaction records, customer information or inventory data (Eberendu, 2016, p. 48).

Unstructured data on the other hand, as the name indicates, lacks a specific format and organization in its raw format and can transmit information in various forms, such as text documents, emails, social media posts, audio recordings, images, videos, or sensor data (Inmon and Nesavich, 2007, chap. 2). Due to the great variety of different formats, processing and analysing unstructured data poses greater challenges compared to structured data (Inmon and Nesavich, 2007, chap. 2).

One distinguishing criterion in data collection is the question of whether users voluntarily provide their data or whether it is collected implicitly without the user's direct consent (Lavie et al., 2010, pp. 3–5). Explicit data collection requires the user for voluntary provision of information through methods such as user questionnaires or keywords to personalize a website or recommendation mechanism (Lavie et al., 2010, pp. 3–5). However, users do not always disclose their data voluntarily or might provide false information to protect their privacy, inducing companies to implicitly collect data (Lavie et al., 2010, pp. 3–5). Implicit data collection therefore refers to the gathering of information through, for instance sensor data, emotion sensing technologies or traceable browsing behaviours disclosing users' preferences (Lavie et al., 2010, pp. 3–5). The data types companies are collecting can vary greatly depending on several parameters such as industry, e. g. healthcare versus financial services, or purposes like research-orientation compared to marketing (Syed et al., 2013, pp. 2446–2450). Thus, it is difficult to give a comprehensive overview of the exact types and information explicitly and implicitly collected. However, several, commonly mentioned data types are being collected.

Personally identifiable information is a broad term mainly including data such as home addresses, telephone numbers, names, email addresses, or social security numbers,

simply any information that can directly identify an individual (Krishnamurthy et al., 2011, pp. 3–5). Demographic information can be summarized as data related to age, marital status, race, ethnicity, gender, education, or other demographic characteristics (Gumbus and Grodzinsky, 2015, pp. 119–121). Further types of data enabling identification of a person are IP addresses, device IDs, cookies, or other tracking technologies that record users' online activities and thereby reveal patterns or proxy variables such as income (Gumbus and Grodzinsky, 2015, p. 120). Similar to the previously mentioned online identifiers, behavioural data itself includes information about users' interactions with company websites, mobile applications, products or services, such as browser or purchase histories, search queries, preferences and engagement patterns and thus is considered a distinctive data type itself (Gumbus and Grodzinsky, 2015, p. 119). Companies also collect information about users' physical locations through GPS, Wi-Fi, or IP addresses (Minch, 2015, pp. 4–6). Through the rise of platforms like Facebook, Instagram or Twitter, user-generated content, including reviews, comments, posts, photos, or videos, is explicitly collected and stored (Moura and Serrão, 2015, pp. 7–8). Nevertheless, social media companies also collect data implicitly, such as friends lists, privately shared information or interactions with contents and users (Moura and Serrão, 2015, pp. 7–8).

Further, one of the frequently collected data types that regularly raises concerns is financial information involving credit card details, bank account information or payment history (Bin Sulaiman et al., 2022, pp. 56–58). Through technological advancements, such as wearables or fitness trackers, health-related data has become more relevant in the data sharing and privacy literature (Esmaeilzadeh, 2019; Solino-Fernandez et al., 2019). Hereby, next to the personally identifiable information, health-related information such as stress levels, moods, habits, sleep or sports patterns are collected (Banerjee et al., 2018, pp. 1–2).

Combined, all this sensitive information empowers an organization of precisely analysing, predicting and ultimately manipulating individuals, implying that the transparent individual has already become reality (Bäumler and Mutius, 2013, pp. 27–29). However, it must be noted that this is only an extract of various, commonly collected data types and is therefore neither depicting the whole variety of all contexts, nor industries, purposes, or technologies. The ongoing innovation of digital technologies, for instance Internet of Things devices, often gives rise to new, or at least more granular data types,

such as new sensors' data, inducing a constant change in the state of knowledge in this field (Cichy et al., 2021, pp. 1865-1866).

2.2 Values of Data Sharing

The fast-paced environment of data collection has made it necessary for governments to establish various frameworks and regulations to give individuals more transparency and protection in the digital age. To comply with official data privacy guidelines, such as the OECD Guidelines on the Protection of Privacy and Transborder Flows of Personal Data or the EU General Data Protection Regulations, businesses have to be more transparent on data handling practices (Cate and Mayer-Schönberger, 2013, p. 67; Goddard, 2017, p. 703). The legislations for instance introduced the requirement for individuals to give explicit consent before sharing data and thereby enable to make better-informed disclosure decisions (Cate and Mayer-Schönberger, 2013, p. 67; Goddard, 2017, p. 703). Considering the number of challenges, the question arises: What value does data sharing add to individuals and organizations and why individuals should disclose data at all if faced with numerous threats to their privacy? For an excerpt on common values of data sharing see figure 2-1.

Values of Data Sharing

Data collecting party (e. g. companies, organizations)	**Data sharing party** (e. g. individuals)
▪ Improve products or services ▪ Personalize marketing ▪ Improve operations ▪ Price discrimination ▪ Better informed decisions ▪ Resell data	▪ Personalized services or products ▪ Less irrelevant marketing or recommendations ▪ Lower prices (if lower costs are passed on) ▪ Enhanced customer experience

Fig. 2-1: Excerpt on values of data sharing

According to economic theory, one characteristic of an efficient market, e. g. a financial market, is the requirement of reflecting all available, relevant information to all market participants (Fama, 1970, p. 383). This allows for the following thought-experiment. What is relevant in financial market theory, can be similarly thought of information asymmetry in any market, also regarding the exchange of information in terms of data, between two parties. A company that knows the purchase behaviour, advertisement

preferences or browsing habits of its potential customers should be enabled to make better-informed decisions on how to interact and finally, gain a competitive advantage compared to a company missing this information (De Schaepdrijver et al., 2022; Leppäniemi et al., 2017; Sarathy and Muralidhar, 2006; Treiblmaier and Pollach, 2007). While companies' usage possibilities of customer data are nearly unlimited, according to the literature, they can be separated into three broader categories for simplification (Morey et al., 2015, p. 7). Collecting data to improve products or services, personalize marketing and advertising or resell the collected data to other companies and third parties (Morey et al., 2015, p. 7).

First, through the collection and analysis of customer data, companies can improve their operations, supply-chains, minimize their costs as well as optimize pricing strategies (Piccoli and Watson, 2008, p. 116; Vassakis et al., 2018, pp. 11–12). For example based on a broad set of variables, such as user's location, device brand or transaction frequency, first-degree price discrimination is no longer a hypothetical construct from economics but standard in the digital age, enabling companies to charge consumers, individual prices (Shiller, 2014, pp. 1–2; Xiao, 2022, pp. 1–2). Further, customers' data enables companies to improve their products and services and thereby, drives innovation (Cichy et al., 2021, p. 1864; Trabucchi and Buganza, 2019, pp. 5–6).

Second, data such as click-logs, search-logs or other information disclosing individuals' behaviour and psychological traits, enables companies primarily to personalize interactions, especially in marketing and advertising with the ultimate goal of higher conversion rates (Matz and Netzer, 2017, pp. 7–8; Trabucchi and Buganza, 2019, pp. 5–6).

Third, individual data introduced a new industry. After collection, the secondary market for reselling individuals' data to third parties is lucrative, although control is lost and ethical considerations arise (Martin, 2015, pp. 73–75).

While companies benefit from efficiency increases in marketing through personalization, customers on the other hand benefit from not being overloaded with irrelevant advertisements, offers or recommendations not meeting their needs and preferences and thus can decrease transaction costs (Anshari et al., 2019, p. 95; Madhok, 2002, pp. 535–540). Further, data-driven efficiency-increases, in companies' operations, induce lower costs, offering the possibility to either lower prices and thereby increase

consumer surplus, or alternatively increase producer surplus (Brown et al., 2011, p. 11). As mentioned earlier, customer data also drives innovation and aligns services and products better with individuals' requirements (Brown et al., 2011, p. 11). In healthcare for instance, insurance companies may improve products and services by analysing customers' health-related data to offer e. g. monetary incentives to support a healthier lifestyle and ultimately lower prices for sporty individuals (Canhoto and Arp, 2017, p. 20).

Overall, individuals' benefits of data sharing could be summarized in terms of an enhanced customer experience, as well as an increased customer value (Cichy et al., 2021, p. 1866). However, it should be noted that the above-assessed values are mainly referring to data sharing between companies and individuals. There are countless other domains or contexts in which data sharing could be investigated but would go beyond the scope of this paper. To mention a few, the educational and academic sector, governments and authorities, law enforcement and security or urban planning are likewise benefitting from individuals' shared data (Martin, 2015, p. 68). While the values of data sharing in these domains should overlap with the mentioned ones there may be specifics to each of them.

2.3 Large Language Models and Data Sharing

This paper will expand the research of determinants in data sharing to a new, trending and currently academically underexplored technology, namely, LLM. This focus enriches the paper with explorative character in the second research question with ChatGPT being one of the larger, publicly available, best known and highly debated LLM, making it an attractive subject of research for the empirical investigation (Brown et al., 2020; Dilmegani, 2023).

LLM are advanced artificial intelligence systems designed to understand and "generate human-like text" (Kasneci et al., 2023, p. 1). LLM are trained on huge amounts of data, allowing them to learn patterns, contexts, and subtleties of language (Kasneci et al., 2023, p. 1; Tamkin et al., 2021, pp. 2–3). Referring to OpenAI's GPT-3, the training data consists of approximately 570 gigabytes of text (Cascella et al., 2023, p. 1). While this data amount does not necessarily sound immense, it corresponds to approximately 385 million pages in Microsoft Word (Coiera et al., 2023, p. 98). LLM are based on deep learning techniques and use neural networks to process and generate text

(Tamkin et al., 2021, pp. 2–3). Simplified they are probabilistic models building sentences based on the question, which word or token is the next most likely one, by deriving relationships from (un-)structured training data (Carlini et al., 2021, p. 2633). Derived and learned probability weights assigned to different tokens are also called parameters and indicate capabilities and skills of the model, for instance amounting to 175 billion parameters for version GPT-3 (Cascella et al., 2023, p. 1).

LLM can be utilized to automate nearly any task which involves language processing, such as content generation, journalistic angle ideation, summarization, translation, chatbots or customer support (Eggmann et al., 2023, pp. 2–3; Kasneci et al., 2023, pp. 1–2; Petridis et al., 2023, sec. 5.2; Sallam, 2023, pp. 10–17). In the healthcare industry, LLM assist for example in medical research and education or patient care by analysing big amounts of medical literature and patient records (Sallam, 2023, pp. 10–17). In finance, they are used for sentiment analysis of market data or generating standardized, financial reports (Araci, 2019). Further, they are also used in fields like education, legal services, journalism and entertainment (Kasneci et al., 2023, pp. 2–3).

Besides capacity limitations, LLM application scenarios had been mainly limited in terms of data input or availability (Zhao et al., 2023, p. 5). Though, recently released plugins reduce these limitations by allowing the AI for instance to browse on the web or access external data (Zhao et al., 2023, p. 5). With all the advantages introduced, LLM have the potential to positively transform industries, however ethical considerations, biases, privacy, and responsible use must be addressed likewise to ensure the positive impact and adoption in future (Brown et al., 2020, pp. 10–14; Teubner et al., 2023, pp. 97–99).

Since LLM require huge amounts of data input and rely on extensive data sharing, they are prone to cyber-attacks (Brown et al., 2022, pp. 1315–1317; Greshake et al., 2023, p. 3). For instance, while inverted attacks aim at manipulating the model to generate unintended or malicious output, prompt injection exploits weaknesses in the model's response generation process (Brown et al., 2022, pp. 1315–1317; Greshake et al., 2023, p. 3).

Prompt injection simply put means, that the user’s entered personal information can be retrieved and remotely accessed by an attacker using indirect prompts (Brown et al., 2022, pp. 1315–1317; Greshake et al., 2023, p. 3). These risks are leveraged by

the newly LLM-integrated plugins and therefore, the potential data leakages are not limited to direct interactions with LLM itself but to all indirect interactions with its plugins (Greshake et al., 2023, pp. 3–4). Figure 2-2 exemplarily shows several such plugins.

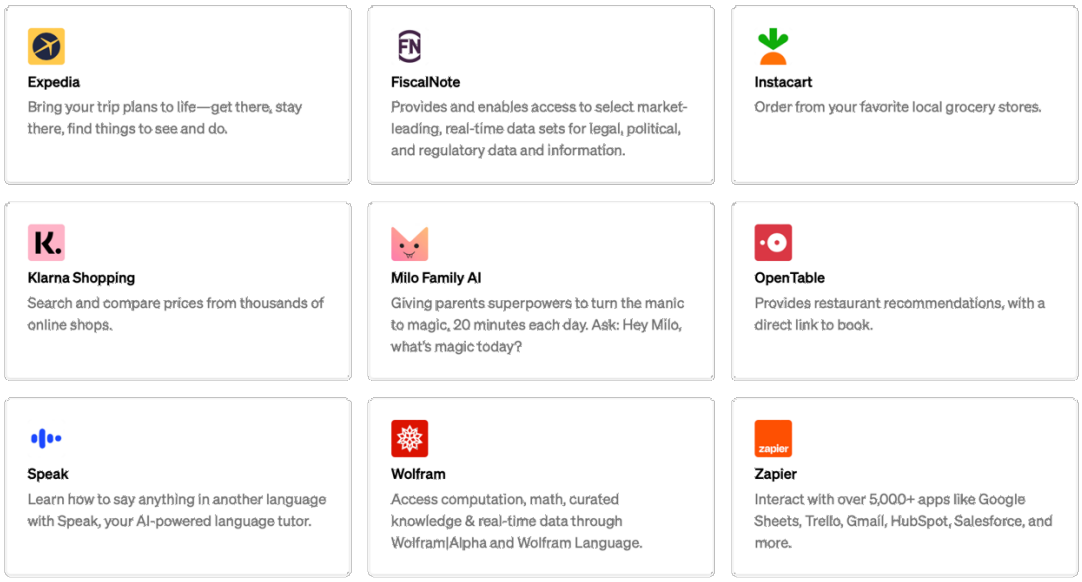


Fig. 2-2: OpenAI's ChatGPT Plugins (OpenAi, 2023a)

Plugins are downloaded and installed on the user's device and connect with ChatGPT to combine its computing power with the access to specific external databases or the web (OpenAi, 2023a). In figure 2-3 the functioning of the plugin Instacart is exemplarily depicting following imaginable application scenario. First a user who installed the plugin asks ChatGPT for customized recipes for a dinner with friends based on the user's available ingredients at home (Zhuang, 2023). ChatGPT subsequently intermediates between user and plugin provider, Instacart. Thereby it accesses Instacart's external database to customize and improve results based on relevant data (Zhuang, 2023). Instacart offers for example the option to fully-automatized deliver all necessary ingredients for the proposed meal within an hour from a participating grocery store (Zhuang, 2023).

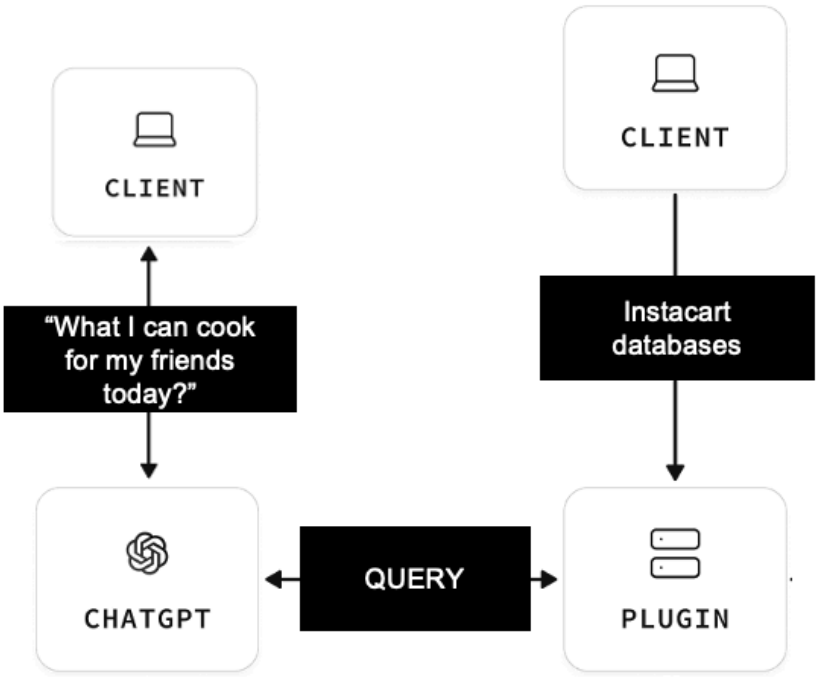


Fig. 2-3: Exemplary ChatGPT Plugin architecture (Romanov and Copplestone, 2023)

While OpenAI promotes the induced leveraged utility of the LLM, it also admits increased risks for users and accessed external data (OpenAi, 2023a). Attacks like the previously mentioned ones could thus pose a risk to personal data because they may extract sensitive information from the plugins' and model's in- and output (OpenAi, 2023b, p. 13). For instance, through carefully created requests, sensitive data like personally identifiable information, addresses or payment information that someone entered on Instacart, or confidential business information entered by an employee, could unintentionally be leaked by an attacker (Carlini et al., 2021, pp. 2636–2639). While a seemingly harmless plugin such as Instacart may sound as it is not requiring lots of data disclosure, the privacy policy of Instacart tells a different story. The following personal information Instacart (2023) collects from its users:

- Contact information
- Account information
- Order information
- Age and identity verification information
- Vehicle information
- Payment information
- Interaction with your personal shopper

- Information you make public
- Other information provided to Instacart
- Information you provide about others
- Device and usage information
- Location information
- Cookies, Pixels, and other tracking technologies
- Personal Health Information.

While most of the collected information might be nowadays considered as status quo, the last point, personal health information, may surprises. Hereby they explain: “Some of the information that we collect about you may relate to: (i) your past, present or future physical or mental health or condition, (ii) the provision of health care to you, or (iii) your past, present, or future payment for the provision of health care” (Instacart, 2023, sec. 5). This demonstrates a significant privacy intrusion by plugins, which users likely not anticipate when using applications supposedly only helping with usual tasks such as cooking, recipes, or grocery shopping.

With hundreds of verified ChatGPT plugins it is difficult to estimate the potential of privacy intrusion for all of them, however it can be assumed that most available plugins are not collecting less data than the provided example Instacart but the same or even more.

In a recent paper, which OpenAI (2023b, p. 13) published itself, it investigated potentials and risks of its new version GPT-4, giving a vivid example for a potential privacy breach. It is explained that data sources which the LLM is trained on, may include personal information, as well as knowledge about individuals with a significant online presence, such as celebrities, politicians or academics (OpenAi, 2023b, p. 13). Since the model can synthesize diverse information types and engage in multi-step reasoning, it can execute tasks involving personal and geographic information, such as associating a phone number with a geographic location of the according individual (OpenAi, 2023b, p. 13). Thereby it could potentially identify individuals when supplemented with external data (OpenAi, 2023b, p. 13). The following list shows an excerpt of personal data that OpenAI (2023c) declares to collect:

- Account information (name, contact information, account credentials, payment information, transaction history)
- User content (input, file uploads, feedback)
- Log data (IP address, browser information, interaction behaviour)
- Usage data (viewed or engaged content, usage behaviour, computer type, mobile device, connection type, settings)
- Device information (device name, operating system, browser information)
- Cookies (when visiting the website)
- Analytics (interaction behaviour)
- Communication information (name, contact information, content of messages)
- Social media information (contact details, interaction behaviour, social media activity).

This list is likely expandable with a rising number of plugins used and gives a hint on the amount and sensitivity of shared data.

A very telling example is the ban of ChatGPT by the Italian government in March 2023 (Pongratz, 2023). The data-protection authorities' main reasons for the ban were lacking legal base for mass collection and storage of users' personal data by OpenAI, incomplete information on the type of stored information and lastly, inadequate filters for young users (Pongratz, 2023). Under the threat of a monetary penalty and market exclusion, OpenAI adapted the age verification process and implemented an option to file an objection for collection and usage of personal data (Supantha et al., 2023). Therefore, this highly debated topic offers a practically relevant opportunity to study data sharing decisions.

2.4 Sharing Behaviour Theories

This section provides an overview of several important theories explaining sharing behaviour, depicted in figure 2-4. Thereby, each theory and the concept are explained and subsequently applied to the context of data sharing. This approach demonstrates the range of potential explanations for disclosure or sharing decisions in literature. The end of this section discusses which theories are further investigated in the paper.

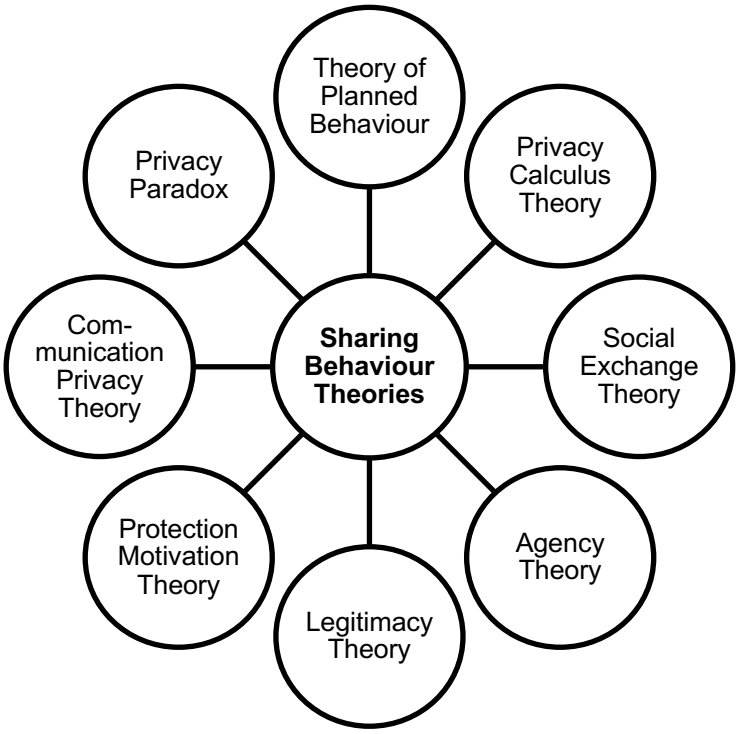


Fig. 2-4: Overview of sharing behaviour theories

The theory of planned behaviour developed by Ajzen (1991) is a social psychological theory that tries to explain human behaviour based on an individual's intention (Ajzen, 1991, p. 179). It identifies three key determinants of intention: “Attitudes, subjective norms, and perceived behavioural control” (Ajzen, 1991, p. 179). Attitudes refer to an individual's personal evaluation of an action, subjective norms involve the social pressure and expectations from the social environment and perceived behavioural control relates to one's perceived ability to successfully execute the behaviour (Ajzen, 1991, p. 188).

In the next step the theory is applied to the context of individuals' decisions to share their personal data. Attitudes, in terms of individuals' attitudes towards data sharing, play an important role in shaping intentions, for instance, some individuals may view data sharing as risky, fearing potential privacy loss or misuse of their personal data (Ajzen, 1991, p. 188; Smith et al., 2011, pp. 997–1002). On the other hand, some may perceive benefits such as personalized services or targeted advertisements resulting from data sharing positively (Smith et al., 2011, p. 1001). Additionally, individuals may hold the belief that companies sell their data or provide access to third parties, which can influence their attitudes towards data sharing (Smith et al., 2011, p. 1001). Further, social influences or subjective norms also contribute to individuals' intentions regarding

a certain behaviour, here data sharing (Ajzen, 1991, p. 188). If individuals hear for example frequently of data breaches or privacy violations, it may create a negative social norm around data sharing, making them reluctant to share their personal data (Bélanger et al., 2021, p. 5). Another potential scenario could be friends repeatedly expressing concerns or warning of data collection practices of organizations or companies, finally reinforcing own reservations about sharing personal data (Bélanger et al., 2021, p. 5). The third determinant, perceived behavioural control in the context of data sharing, would refer to an individual's perceived control over shared personal data (Ajzen, 1991, p. 179). Individuals who believe they have control over their shared personal data, for instance by controlling access, the right to erase, or manage privacy settings, finally might feel more comfortable engaging in data sharing (Ajzen, 1991, pp. 182–183; Bansal and Nah, 2020, p. 3; Bélanger et al., 2021).

Compared to the previously assessed theory, the privacy calculus theory is a frequently applied framework that explains decision making based on the trade-off between weighing benefits of disclosing information versus costs of privacy (Keith et al., 2013, p. 1165; Kokolakis, 2017, pp. 129–130). Thereby, it suggests a rational solution based on expected value (Keith et al., 2013, p. 1165; Kokolakis, 2017, pp. 129–130). Thus, the privacy calculus theory is basically an expected utility theory, claiming that rational individuals are always trying to maximize their expected utility from a certain behaviour or decision, by asserting probabilities to all potential outcomes (Friedman and Savage, 1952; Von Neumann and Morgenstern, 2007). Accordingly, the determinants are costs, such as privacy concerns or perceived risks, weighing negatively and perceived benefits as positive counterweights (Keith et al., 2013, p. 1165). Privacy calculus theory suggests that individuals engage in a cost-benefit analysis when deciding to share personal data, weighing the perceived privacy risks, such as location data disclosure, against the expected benefits, e. g. personalization (Keith et al., 2013, p. 1168). This simplicity brings a wide range of possible application scenarios and is one of the causes making the theory accepted and applied across the investigated literature in the field. For this reason, the theory is also chosen for the conceptualization of the research model in the second research question of this paper and is further discussed in the later parts.

Social exchange theory reaches back to the 1950’s and has been revised and extended since then various times (Cropanzano and Mitchell, 2005; Redmond, 2015;

Thibaut and Kelley, 1959). The concept is mainly applied to social relationships but has also often been successfully applied to other constellations such as organizational-, job- or economic-relationships (Cropanzano and Mitchell, 2005; Redmond, 2015; Thibaut and Kelley, 1959). The theory provides an understanding of when individuals engage in relationships and when they try to reduce engagement (Redmond, 2015). Sharing personal data with a company to benefit from personalized services in return is such an economic relationship, or the interaction between a physician asking his patient for consent to share personal (Redmond, 2015, p. 20). Both can therefore be interpreted as an economic relationship, based on a social exchange, allowing the application of the social exchange theory (Redmond, 2015, p. 20).

The theory is based on the following five key principles. Individuals, trying to maximize rewards and minimize costs in their social interactions, evaluate the fairness in terms of equity of an exchange by comparing the ratio of their inputs and outcomes with those of others and may engage in reciprocity (Redmond, 2015, pp. 14–16). Consequently, the duration and stability of relationships are influenced by perceived rewards and costs, and individuals are motivated to search for alternatives if the existing relationship fails to provide sufficient rewards or involves too high costs (Redmond, 2015, pp. 16–20). Applied to the data sharing context, the individual weighs the expected benefits of sharing personal data, such as personalized services or discounts, against the perceived risks and costs, like privacy concerns or potential misuse of data, comparable to the privacy calculus theory (Redmond, 2015, pp. 14–20). However, in terms of reciprocity the individual may additionally consider previous experiences, such as a history of positive interactions and transactions with the company or organization, possibly enhancing the individual's willingness to engage in data sharing (Redmond, 2015, pp. 14–20).

Like the before-mentioned theories, one more theory describing the sharing process in an economic, rational and calculative way, is the principal agent theory (Fischer, 1993). The principal, e. g. customer, hires and pays the agent, e. g. a company, for getting a product or service in return, based on the assumptions of rationality, utility and profit maximization (Braun and Guston, 2003, pp. 303–304; Fischer, 1993, sec. 2). Differences of opinion, different levels of risk aversion, conflicting priorities and targets between the two actors are shaping this relationship (Braun and Guston, 2003, pp. 303–304; Fischer, 1993, sec. 2).

In the context of individual data sharing, the principal agent theory suggests individuals acting as principals who delegate the task of data management to companies, acting as agents, to receive benefits such as personalized products or services (Braun and Guston, 2003, pp. 303–304; Fischer, 1993, sec. 2). Thereby, the principal must assess for instance the company's competence or trustworthiness and other determinants, disclosing capabilities to ultimately reduce information asymmetries and make an informed decision (Braun and Guston, 2003, p. 303). This is due to the assumption that the agent, here the data collecting party, tries to maximize profits from the shared data by e. g. selling it to third parties or unintended marketing use, while the principal wants his or her personal data protected and used only for agreed purposes (Braun and Guston, 2003, pp. 303–304). Typical solutions to this problem could include monitoring mechanisms or contracts (Braun and Guston, 2003, pp. 303–304).

Different to these previous theories the protection motivation theory is a psychological framework that explains how individuals respond to potential threats and make decisions to protect themselves (Norman et al., 2015, pp. 81–85). Hereby, individuals evaluate the severity and vulnerability of a threat and effectiveness of potential protective measures, in terms of individual's self-efficacy, response efficacy and response costs (Norman et al., 2015, pp. 81–85). This evaluation then is ultimately influencing the motivation whether to engage in protective behaviours or not (Norman et al., 2015, pp. 81–85). According to the theory, individuals are more likely to adopt protective actions when they perceive threat and vulnerability, recommended coping actions and self-efficacy as significantly high (Norman et al., 2015, pp. 81–85). Applying the theory exemplarily to individuals deciding whether to share personal data with a healthcare company or a physician, e. g. via wearables as it has previously been investigated in some studies, gives further insights in the psychological processes (Norman et al., 2015, pp. 81–104). Individuals hereby would assess the threat of potential risks associated with sharing their personal data, like privacy breaches or misuse of information and their vulnerability to these risks, such as the sensitivity or amount of data and potential harm it could cause, if lost (Norman et al., 2015, pp. 88–104). Subsequently, individuals would evaluate for instance the effectiveness of the data protection measures and their confidence in the company's ability to protect their data (Norman et al., 2015, pp. 81–104). If individuals perceive the threat as significant, their vulnerability as high, but believe in the healthcare company’s data protection measures, they

would be more likely to share their personal data and not to engage in protection behaviour (Norman et al., 2015, pp. 81–85). Response costs hereby could correspond to not getting access to personalized services and products or reading time-intensive privacy regulations every time a website is visited (Norman et al., 2015, pp. 81–85). While this theory gives a new point of view regarding sharing decisions based on rudimentary psychological processes, the next theory assesses a social perspective.

The legitimacy theory explores how organizations try to maintain their social legitimacy by complying to societal standards and expectations through demonstrating that their actions are aligned with social values, norms, and beliefs (Suchman, 1995, pp. 573–600). Thereby, organizations try to gain acceptance, approval and support from various stakeholders (Suchman, 1995, pp. 573–600). There are four legitimation strategies, namely the general, pragmatic, moral and cognitive one (Suchman, 1995, p. 600). These can manifest for instance in seeking certifications or locate friendly audiences to in turn receive stakeholder support, access to resources, and thereby long-term success (Suchman, 1995, p. 600).

For the following example it is assumed that in a society individuals evaluate whether companies' data handling practices align with societal norms, expectations, and legal regulations regarding privacy and data protection (Suchman, 1995, pp. 573–600). In response, companies make use of one of the above mentioned legitimation strategies to demonstrate their legitimacy by being transparent about data collection practices, providing clear and comprehensible privacy policies, implementing effective security measures or seeking consent from individuals (Suchman, 1995, p. 600). By adhering to these principles, they may enhance legitimacy, increase individuals' level of trust and ultimately increase their willingness to share personal data.

The last sharing theory presented combines several elements introduced in the previous theories. The so-called communication privacy management theory deals with the communication and management of private information and assumes individuals need for controlling private information to maintain privacy (Petronio, 2010, p. 178). While the legitimacy theory suggests four legitimation strategies to gain acceptance and approval, the communication privacy management theory in contrast consists of five principles that explain how to manage private information (Petronio, 2010, pp. 178–179). Principle one is the private information ownership, stating that individuals consider private information as personal property, feel ownership and consequently have the right

to decide how it is shared with others (Petronio, 2010, pp. 178–179). The second principle is called private information control, suggesting individuals want to keep control over private information (Petronio, 2010, p. 179). Thereby having the right to decide to whom, and what information specifically is disclosed (Petronio, 2010, p. 179). The third principle, private information rules, states that people develop rules and norms to regulate the use of private information (Petronio, 2010, pp. 179–180). Such rules can be cultural, social or individual and determine how information is stored, disclosed or kept private (Petronio, 2010, pp. 179–180). Principle four is the so-called private information co-ownership and guardianship (Petronio, 2010, pp. 180–181). According to this principle, sharing private information with other individuals makes them a co-owner, giving them a common feeling of ownership and interest in protection of that information (Petronio, 2010, pp. 180–181). A typical example are close friends or family members who share and protect each other's private information (Petronio, 2010, pp. 180–181). The last principle is private information boundary turbulence, implying that tensions and conflicts may arise if boundaries of private information are not clearly defined or violated, e. g. when information is disclosed against a person's will or if control over private information is lost (Petronio, 2010, p. 182).

These five principles applied to the context of data sharing explain communication and management of personal data on the level of interpersonal relations. According to the private information ownership, the individual considers personal data to be his or her personal property and therefore has the right to decide how it is shared (Petronio, 2010, pp. 178–179). Private information control implies that the individual wants to retain control over his or her personal data by deciding which organizations or individuals are allowed to have access and to what extent that data may be used, e. g. by excluding unauthorized third-parties (Petronio, 2010, p. 179). The private information rules hereby are specifying that sensitive information may only be shared with trusted parties or that data may only be used for specific purposes and can be deleted at any time (Petronio, 2010, pp. 179–180). Consequently, a private information boundary turbulence occurs if personal information is shared unauthorizedly or misused against the individual's will, leading to a loss of trust and a negative impact on the relationship (Petronio, 2010, p. 182). Thus, the theory suggests that individuals should be in full control of their data and make conscious decisions about how and with whom they

share personal information (Petronio, 2010, pp. 178–182). This ultimately avoids private information boundary turbulences and establishes a trustful relationship with the data collecting party (Petronio, 2010, pp. 178–182).

While most of the theories consistently emphasize individuals' perceived necessity for privacy protection, this section is closing with the so-called privacy paradox, questioning this supposedly consistent behaviour propagated by individuals (Kokolakis, 2017, p. 128). The privacy paradox describes the situation of individuals claiming high values on data protection and privacy on the one hand, yet willingly share personal data online on the other (Kokolakis, 2017, p. 128). Despite the desire for privacy and the concerns over data protection, lots of individuals behave contradictorily for instance by disclosing private information on social media for relatively low benefits of self-expression (Kokolakis, 2017, p. 128).

The privacy paradox may be explained by several factors. Convenience plays a major role in deviating from claimed or intended behaviour, as many online services offer for instance personalized experiences, services and products when sharing personal data (Chellappa and Sin, 2005, pp. 197–198; Kokolakis, 2017, p. 128). In addition, social norms and the desire for social recognition, especially in social networks, might convince users to disclose private information to belong, or to receive positive reactions from other members (Kokolakis, 2017, p. 128). Another factor is the limited awareness of privacy risks, as many individuals do not fully understand how data is collected, stored and used (Kokolakis, 2017, p. 129). The privacy paradox is further discussed in the second research question as it is highly debated in the field and needs to be further explored in state-of-the-art technology.

Summarizing, the selection of introduced theories should illustrate possible explanations for sharing and disclosure decisions ranging from profoundly psychological to predominantly calculative frameworks. The decision which theory better meets the research objectives therefore cannot be generalized and must be carefully evaluated. This evaluation is done for the underlying research objectives in the subsequent sections of the paper.

2.6 Determinants of Individuals' Willingness to Share Personal Data

The following part answers research question one by conducting a systematic literature review on the determinants of individuals' willingness to share personal data and finally provides a comprehensive overview.

The proceeding of the systematic literature review is depicted in the Prisma flow diagram in figure 2-5 (Prisma, 2023). The literature review thereby prioritised papers from the so-called Basket of Eight and tried to capture relevant research results across diverse contexts, industries and domains (AIS, 2023). The first database consulted was the Association for Information Systems (AIS, 2023). The search queries used to filter the papers were: willingness to share personal data, determinants of personal data sharing, data sharing decision, data sharing behaviour, willingness to share personal data and data sharing determinants. The search provided a total of 225 records identified. Additionally, through the Web of Science database, 10 relevant papers were identified with results equally limited to the Basket of Eight (AIS, 2023). A Google Scholar search additionally identified 38 relevant records.

After removing duplicates and screening the titles and abstracts of the total of 273 papers, 92 irrelevant records were excluded. The remaining 89 records were full-text assessed for eligibility and subsequently 36 full-text articles excluded. The main reasons for exclusions in both steps were either irrelevant types of investigated sharing behaviour, like sharing economies or knowledge sharing, as well as irrelevant research scopes not investigating sharing behaviour itself. Finally, 53 studies were included in the qualitative synthesis.

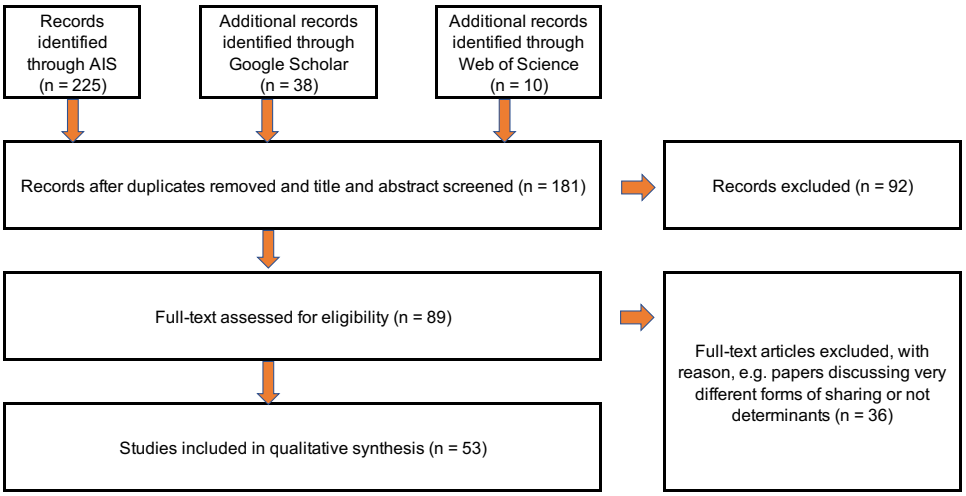


Fig. 2-5: Prisma flow diagram (Prisma, 2023)

An overview of the 53 studies can be found in A-1. The studies' contexts and investigated technologies are ranging from inter alia open data, social media, commercial, blockchain, healthcare, to the Internet of Things.

The proceeding of the analysis was the following. First, the 53 included studies were scanned for determinants, constructs or descriptions of relevant factors determining data sharing decisions. Second, repeating determinants or constructs, as well as semantically similar ones were summarized, depicted in A-2. Before diving into the determinants some remarks are given.

First, these are no explicit definitions for the determinants presented in the following part. Especially, since the mentioned constructs and determinants are partly overlapping among different studies, due to a variety of applied models, psychological theories, frameworks, and contexts. A detailed assessment of each study's definitions would thus excess the scope of this literature review and would not lead to an improved result. Therefore, in the following the determinants are defined and explained in a generalist approach. Additionally, for giving a comprehensive overview, less often mentioned determinants are depicted in the matrix in figure 2-7 at the end of this part.

Second, for better understanding the determinants' confounding influences on the willingness to share personal data, privacy calculus theory is applied where applicable. This approach is chosen since the privacy calculus theory also is the underlying theory in the research model of the second research question. Therefore, the application for each determinant clarifies the essential logic for better understanding the later parts of the .

Third, the determinants are presented in descending order, starting with determinants mentioned more often across the included studies. However, this does not necessarily imply a higher relevance but can be result of the overlapping and partly ambiguous definitions. Figure 2-6 presents an overview of the ten most common and relevant determinants in data sharing decisions.

Determinants of Data Sharing Decisions	Privacy and data security concerns
	Benefits and incentives
	Control
	Trust
	Severity
	Purpose
	Education or experience
	Access
	Awareness
	Transparency

Fig. 2-6: Overview of common determinants of data sharing decisions

Privacy and data-security concerns are mentioned in nearly all the relevant papers. This determinant describes individuals' negative perceptions or attitudes considering various areas of concerns, such as collection, usage, access or errors and thereby includes already other determinants as explanatory items (Ackermann et al., 2022, p. 376; Cichy et al., 2021, p. 1870; Smith et al., 1996, p. 172). In the context of connected cars, an example for the determinant would be a taxi driver's raised privacy and data-security concerns considering his future employment, in case the connected car is collecting private driving data and subsequently makes erroneous accident predictions, forbidding future employment (Cichy et al., 2021, p. 1870). Another example would be consumers' raised privacy concerns through online behavioural advertising and the perception of being spied on, finally causing discomfort (Ur et al., 2012, p. 1). In the privacy calculus, privacy and data-security concerns should increase privacy costs and ultimately decrease the willingness to disclose data (Keith et al., 2013, p. 1165).

Benefits and incentives are an extensively investigated key determinant of individuals' willingness to share personal data. Benefits or incentives can be very different across industries. They can monetary as well as non-monetary, like discounts, personalized products and advertisements or convenience in using products and services (Cichy et al., 2021, p. 1876; Ioannou et al., 2020, sec. 2.2; Song and Kim, 2021, pp. 2–4). Benefits and incentives given to individuals are positively weighted in the privacy calculus

and can incentivize an individual to share data to get for instance access to a service or product in return (Keith et al., 2013, p. 1165; Zhang et al., 2018, p. 484).

Control refers to individuals' perceived level of being able to exercise power over shared data, including for instance to delete it, opt-out or stay anonymous (Malhotra et al., 2004, p. 339; Pavlou, 2011, p. 978; Wang et al., 2016, sec. 2.1). Thereby the individual is given the conscious choice of accepting or rejecting the collector's purposes or intentions (Malhotra et al., 2004, p. 339; Pavlou, 2011, p. 978; Wang et al., 2016, sec. 2.1). If individuals perceive higher levels of control over their personal data, they have less privacy concerns and subsequently weigh costs of disclosing personal data less, ultimately increasing the willingness to share personal data (Ioannou et al., 2020, sec. 2.3).

Another frequently mentioned determinant is trust. While trust can be defined as a construct with lots of confounding factors, in the investigated literature it mainly describes individuals' belief that the data collecting party uses shared information responsibly and competently in the intended way (Ioannou et al., 2020, sec. 2.3; Song and Kim, 2021, p. 4). Trust is reducing the negative weight of the earlier mentioned privacy and data-security concerns in the privacy calculus (Smith et al., 2011, p. 1000). It therefore can increase individuals' willingness to engage in information disclosure by reducing privacy costs (Smith et al., 2011, p. 1000).

Severity refers to the question of how severe a loss of shared information would be (Ackermann et al., 2022, p. 378; Ioannou et al., 2020, sec. 2.4; Malhotra et al., 2004, p. 339). Thus, the type of shared data, such as location, biometric, financial or health data, the amount and finally the assessment of individuals' perceived level of sensitivity is crucial (Ackermann et al., 2022, p. 378; Ioannou et al., 2020, sec. 2.4; Malhotra et al., 2004, p. 339). Individuals are for instance more likely to share data types which align intuitively to a company's industry, such as an online bank asking for financial information not for health- or social media information (Ackermann et al., 2022, p. 384). Consequently, privacy calculus implies a positive relationship between costs of data sharing, data sensitivity and requested amount (Ackermann et al., 2022, p. 378; Ioannou et al., 2020, sec. 2.4; Malhotra et al., 2004, p. 339; Smith et al., 2011, pp. 997–1002). If highly sensitive data is requested, even generous benefits offered in return may not reduce the privacy costs, ultimately lowering the willingness to share personal data (Ackermann et al., 2022, p. 384).

Purpose is describing the advertised, portrayed or intended usage of collected data, such as research and development, innovation, personalization or marketing (Ackermann et al., 2022, p. 378; Anderson and Agarwal, 2011, p. 473; Cichy et al., 2021, p. 1870; Keith et al., 2013, p. 1172; Wang et al., 2016, sec. 1). Besides understanding and supporting the intended use, purpose may consist of personal meaning, goal-directedness and beyond-the-self orientation, in sense of behaviour trying to make a real change in the broader context, not just to oneself (Bronk et al., 2018, pp. 1–2). If individuals are given credible purpose, such as sharing personal health and training data via wearables with healthcare providers to improve disease prevention, either may lower the privacy costs or increase overall benefits, ultimately increasing the willingness to share data (Ackermann et al., 2022, p. 378; Anderson and Agarwal, 2011, p. 473; Bronk et al., 2018, pp. 1–2; Keith et al., 2013, p. 1172; Wang et al., 2016, sec. 1).

The determinant education or experience is describing either the education or experience regarding technologies and their functioning, such as smartphones, apps, internet applications, state-of-the-art technologies or data collection practices (Bélanger et al., 2021, pp. 5–8; Malhotra et al., 2004, p. 348). Hereby, experiences are interpreted mostly as personally- but partly also verbally-shared experiences made by friends or relatives (Bélanger et al., 2021, p. 5). It must be noted that education therefore refers to the above-mentioned understanding of technologies and not to the control variable education, which usually refers to accumulated degrees or school years. In privacy calculus, education or experience can both, increase or lower variables such as severity or privacy concerns, depending on the specific context, and accordingly lead to lower or higher privacy concerns (Malhotra et al., 2004, p. 348; Smith et al., 2011, pp. 997–1002).

Another important determinant, namely access, overlaps partly with control, privacy concerns and data security, as well as transparency. In various studies the determinant is even considered solely an item of the construct perceived control (Agahari and de Reuver, 2022, p. 6). However, access mainly describes individuals' concerns about the question of who accesses shared data either authorized or unauthorized (Agahari and de Reuver, 2022, pp. 3–6; Hoadley et al., 2010, p. 8; Pavlou, 2011, pp. 979–984; Smith et al., 2011, p. 1001). Fear that the company sells the collected data to third-parties on data marketplaces or to unethical firms is thereby prevalent (Wang et al.,

2016, sec. 2.2; Xu et al., 2011, p. 44). In privacy calculus, the perception of limited third-party access or high levels of control on who accesses the shared data is therefore reducing privacy concerns and ultimately privacy costs (Agahari and de Reuver, 2022, pp. 3–6; Hoadley et al., 2010, p. 8; Kokolakis, 2017, pp. 129–130; Pavlou, 2011, pp. 979–984; Smith et al., 2011, p. 1001).

Awareness refers to the level of understanding potential risks of data sharing and the degree to which the individual is informed about privacy regulations by the collecting party (Becker et al., 2021, sec. 4.3; Lowry et al., 2011, pp. 163–164; Smith et al., 2011, p. 998). Prior negative experiences may result in higher levels of awareness (Xu et al., 2011, p. 45). A lower level of awareness on the other hand can be interpreted the following way. Individuals who are not aware of lacking privacy and legal frameworks on a company's website, should consequently have less privacy concerns resulting in lower privacy costs (Smith et al., 2011, p. 998).

Transparency overlaps inter alia, with trust or privacy and data security concerns. It describes the degree to which the data collecting party is open in communicating purpose, type, usage, and accessors of shared data and whether details regarding data handling practices are tried to be covered (Becker et al., 2021, sec. 4.3.3.). Therefore, higher transparency can lower privacy costs and ultimately improve individuals' willingness to share personal data (Becker et al., 2021, sec. 4.3.3.; Smith et al., 2011, pp. 1001–1002).

Last, the willingness to share personal data basically represents the result of the privacy calculus or decision process. Referring to privacy calculus theory, the willingness to share personal data increases with its benefits and decreases with its costs (Keith et al., 2013, pp. 1165–1168; Kokolakis, 2017, pp. 129–130). Consequently, the decision to share data is positive if benefits outweigh costs (Keith et al., 2013, pp. 1165–1168; Kokolakis, 2017, pp. 129–130).

The presented set of determinants could be expanded by further including determinants that are mentioned less frequently, or which are basically manifestations of the previously mentioned constructs or determinants. It must be stressed that the definition and variation of the determinants is highly context dependant. Figure 2-7 depicts an overview of several additional determinants.

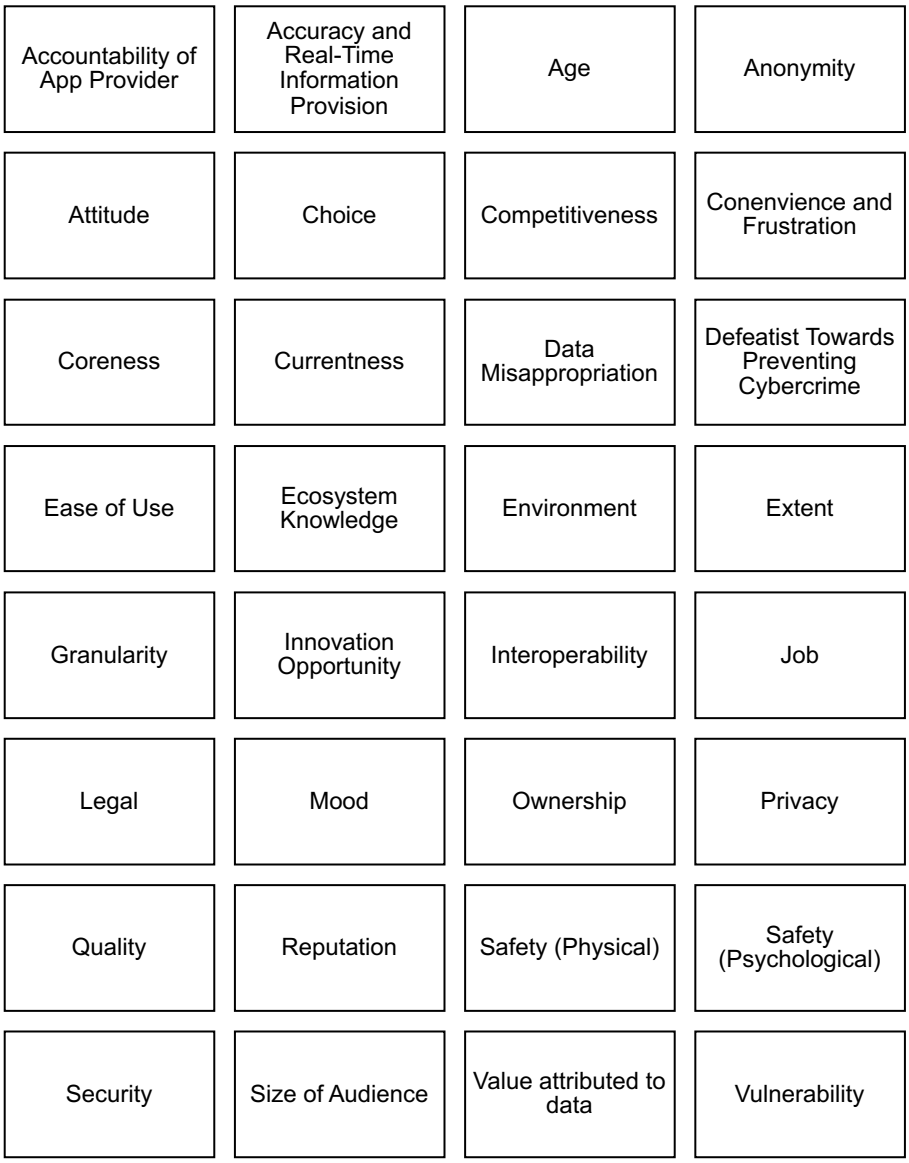


Fig. 2-7: Determinants matrix (Becker et al., 2021; Enders et al., 2020)

A-3 gives comprehensive definitions to these determinants. The matrix is expandable or reducible by e. g. bundling various determinants into one construct or the other way round, as they are often overlapping. While they could be categorised as own determinants themselves, they could be also simply assigned to the ten main determinants presented previously. For example, the determinant choice was defined as an individual having a real choice to decide whether he or she wants to share data or not (Becker et al., 2021, tbl. 3). However, it could also be considered as a sub-category of the superordinate determinant control. Namely, by giving individuals control through e. g. opt-out, they have a real choice regarding sharing decisions. This demonstrates the difficulties of defining such determinants definitely and further shows the arising ambiguities, further discussed in the limitations section.

3 Research Model and Hypotheses Development

The following sections give an overview based on a thematic map, and then derive the research model from theory. Last, the hypotheses regarding the research model are introduced and extensively prepared for the subsequent implementation.

3.1 Research Model

Based on the theoretical foundations and the understanding of the previously introduced framework on determinants, the second research question asks: *What determinants are more relevant in data sharing decisions and how are they interrelated?*

The thematic map in figure 3-1 summarizes the previous findings by showing the discussed sharing theories, investigated domains, contexts, and technologies, as well as the main determinants resulting from research question one. The coloured boxes depict the foci of the underlying paper. From the eight introduced sharing theories, the research model integrates the privacy calculus theory and further extends the understanding of the privacy paradox. Out of the framework of ten main determinants, privacy and data security concerns, benefits and incentives and trust are further investigated. To enhance the explorative character, the relatively recent introduction of ChatGPT in the field of LLM and artificial intelligence sets the technological focus. A quick Web of Science search (see for details WebOfScience, 2023) at the time of writing the paper showed a number of only 12 publications investigating determinants of data sharing in the field of LLM so far, demonstrating a notable research gap.

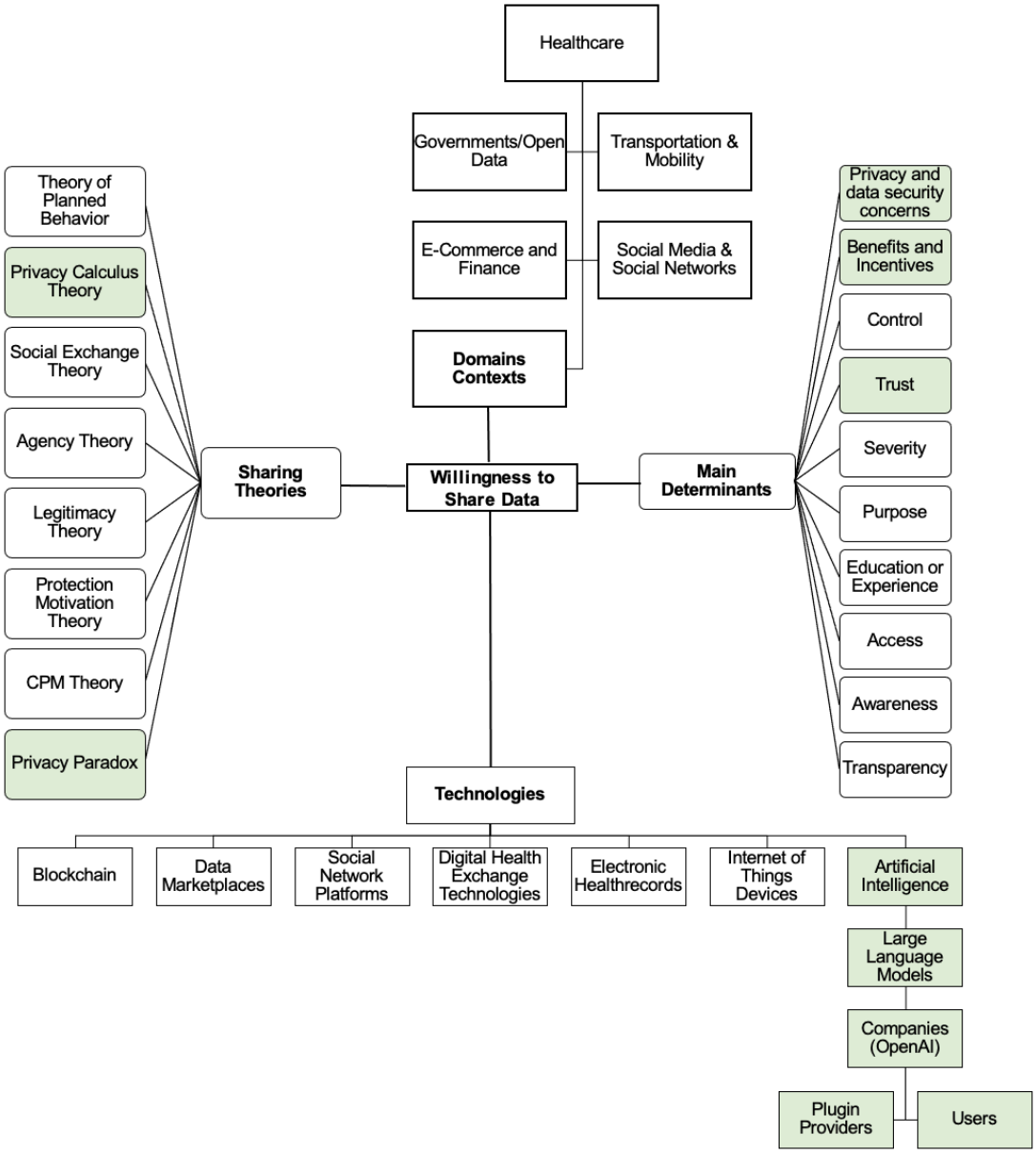


Fig. 3-1: Thematic map (note: green boxes represent the paper' foci)

By utilizing privacy calculus theory, a holistic understanding of individuals' disclosure decisions can be provided, as it allows for the integration of various contextual privacy risk factors, such as data sensitivity and data security, psychological factors from social exchange theory, and communication privacy management theory (Andrade et al., 2002, pp. 350–353; Cichy et al., 2021, pp. 1865–1877; Li, 2012, pp. 471–473). Privacy calculus theory complements elements of these other theories by considering the calculation of costs and benefits in the decision-making process and aligns with traditional approaches, used in studies on individuals' disclosure decisions (Andrade et al., 2002, pp. 350–353; Cichy et al., 2021, pp. 1865–1877; Li, 2012, pp. 471–473). It is consistent with theories such as utility maximization and expected value theory, which have been

widely applied in the field (Ajzen, 1991; Cichy et al., 2021, p. 1865; Li, 2012, pp. 471–473). Furthermore, by employing privacy calculus theory the study can build upon the broad, established knowledge and contribute to a more comprehensive understanding of the determinants of data sharing decisions in general as well as in the context of ChatGPT. While it would be also feasible to simply investigate a set of determinants without any underlying structure given by a theory, the privacy calculus theory hereby instead suggests an established, dual path approach trying to explain individuals' intentions of information disclosure decisions calculative and rationally (Wang et al., 2016, sec. 2). The framework hereby is applied in the same manner as described in the previous parts, namely by depicting a positively weighted path, perceived benefits, and a negatively weighted path, perceived risks (Wang et al., 2016, sec. 2). This constitutes a balanced view, dual-path model, trading off both antecedents simultaneously and their effects on individual's willingness to share personal data in ChatGPT (Wang et al., 2016, sec. 1).

As specifically ChatGPT and its plugins provide highly personalized services, the so-called personalization privacy paradox, as an extended version of the privacy paradox, seems predestined for application in this context. It suggests that users indicate high aversion towards data sharing, while on the other hand they are willingly disclosing their data for often relatively low benefits of personalization (Awad and Krishnan, 2006, pp. 14–20; Chellappa and Sin, 2005, pp. 197–198; Xu et al., 2011, pp. 12–13). However, predominantly LLM collect lots of data in order to train the model or personalize services for individuals (Awad and Krishnan, 2006, pp. 14–20; Xu et al., 2011, pp. 12–13). Thus, individuals' intentions trying to minimize privacy intrusion consequently causes a dilemma (Awad and Krishnan, 2006, pp. 14–20; Xu et al., 2011, pp. 12–13). As mentioned earlier, this trend in collecting and exploiting data is even increasing by the release of plugins, accessing additionally external data. While this paradox has proven to be dominant in lots of digital technologies and contexts it is important to investigate whether it is also eminent for ChatGPT, considering this is a still relatively recent and underexplored technology (Kasneci et al., 2023, p. 1; Tamkin et al., 2021, pp. 2–3).

The initial research model depicted in figure 3-2 contains six reflective constructs, each reflected by five items, as well as additionally eight control variables. The model's de-

tailed version showing all items and control variables can be found in A-4. The constructs and items were mainly adapted from the established literature to avoid low construct- or item-reliability, that could be caused by potentially implementing new, untested, and thus imprecise measures. While the first draft of the model contained additionally perceived severity and compensation for shared data, the final model only investigates privacy concerns, prior disclosure behaviour, trust, perceived risks, and benefits. The main variable of interest is willingness to share personal data in ChatGPT, or short, the willingness to share. The reasons for excluding the two constructs from the model were first, a too extensive survey regarding the scope of the paper. Second, a lower level of the constructs' importance, based on the results in the literature review. Third, overlapping constructs and items with already included constructs would increase the extent of the survey and reduce the participation rate. Further, they could highly correlate due to the conceptual similarity, lowering the overall data quality (Hair Jr et al., 2021, pp. 78–79). All included items, except indicated differently, are measured on a seven-point Likert scale reaching from strongly disagree to strongly agree. For a detailed overview of all constructs and items see A-5.

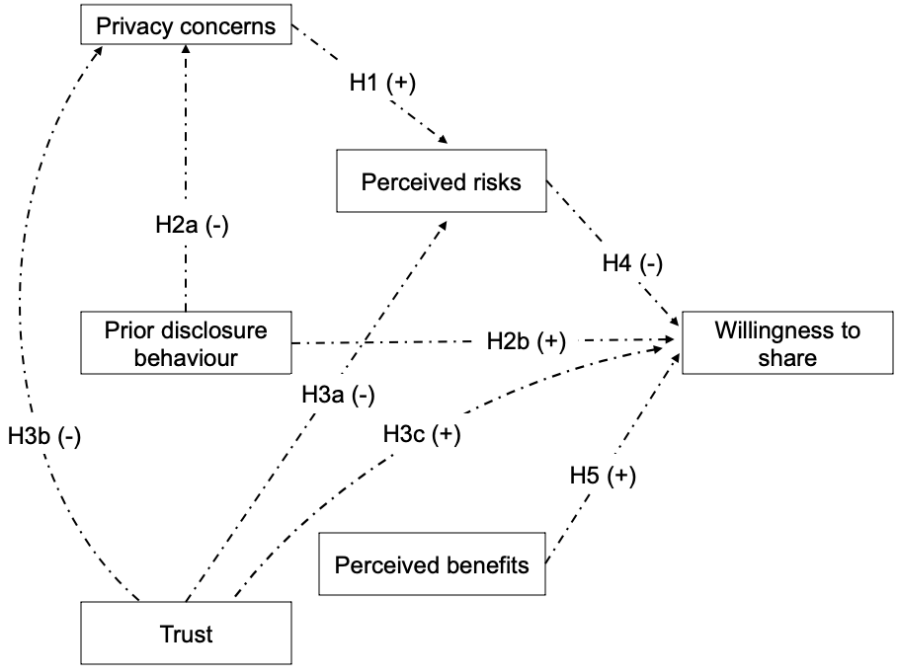


Fig. 3-2: Initial research model

Referring to A-2, privacy concerns are one of the most frequently mentioned and investigated constructs in the examined literature. Individuals with high levels of privacy concerns should be less willing to share their personal data (Ackermann et al., 2022,

p. 376). By adapting the construct to the context of ChatGPT, the items therefore should be reflecting individuals' concerns of misuse of private information submitted through ChatGPT or their availability on the internet (Dinev and Hart, 2006, p. 77). Further concerns to be measured are unforeseen data abuse by third-parties, feeling of online privacy invasion, or loss of control as a result of marketing by ChatGPT (Dinev and Hart, 2006, p. 77). Trust is mainly summarized by the individual's beliefs, that the company behind the service, here ChatGPT, does "not engage in opportunistic behaviour" (Dinev and Hart, 2006, p. 66). These beliefs imply that ChatGPT handles the submitted personal data responsibly, safely and competently and is an important factor when it comes to disclosure decisions (Dinev and Hart, 2006, p. 66). High levels of trust in ChatGPT should reflect in less caution during use, as well as being able to rely on its promises, such as abiding data privacy rules (Cichy et al., 2021, pp. 1890–1891). As ChatGPT, mainly introduced by the plugins, also enables users to engage in business transactions, users' high levels of trust could also reflect indirectly in the perception of ChatGPT as being a reliable environment to conduct business transactions (Dinev and Hart, 2006, p. 77). Two of these five items are coded in reverse (TRUST_2r, TRUST_3r), while all of them are measured on the same seven-point Likert scale ranging from strongly disagree to strongly agree.

Introduced to measure the privacy paradox, the construct prior disclosure behaviour indicates participants' past data sharing behaviour (Motiwalla and Li, 2016, sec. 3.4). The items are measuring the users' tendency to share their own, friends', colleagues' or employers' personal information, to in turn benefit from efficiency gains or other advantages (Motiwalla and Li, 2016, sec. 3.4). A specific example would be a user who copies and pastes internal company documents and information into ChatGPT to efficiently create a company report for a supervisor. Similarly, a user who enters detailed personal information on one of her friends to receive personalized gift recommendations for a birthday party would be another possible case. According to literature, users who have a conservative prior disclosure behaviour, meaning that they have been less willing to disclose personal data in the past, should also show higher levels of perceived risks of sharing data (Motiwalla and Li, 2016, sec. 3.4; Xu et al., 2011, pp. 48–50). Paradoxical would be if users indicate previous high willingness to share personal data but on the other hand indicate high levels of perceived risks in the survey (Xu et al., 2011, pp. 48–50). However, these arguments are followed-up in more detail in the

hypotheses section. While the mentioned items are also measured on a seven-point Likert scale, they are adapted to ranging from never to always. The construct perceived benefits represents the positive main path in the dual path model and is the positive weight in the privacy calculus. Reduced searching time to find relevant information, convenience, increased efficiency and a resulting improved work-life balance are the main benefits suggested in the items measuring this construct (Mohamed and Ahmad, 2012, pp. 2373–2374; Xu et al., 2011, p. 50). However, while the list of potential benefits could be extended, the selection of the items was mainly influenced by adapting established items from literature to the specifics of the context, noted in A-5.

Contrary to perceived benefits, perceived risks represent the privacy costs, the negatively weighted main path in the dual path model, determining the outcome of the privacy calculus. The construct's items reflecting the levels of users' perceived risk resulting from providing personal data through ChatGPT (Xu et al., 2011, p. 50). One of the items is reversely coded (PRISK_4r).

The main construct of interest in the research model is the individual's willingness to share personal data in ChatGPT. It represents the result of the privacy calculus, dual path model, after weighing perceived risks against perceived benefits (Keith et al., 2013, p. 1165; Kokolakis, 2017, pp. 129–130). The items are reflecting the willingness and likelihood of disclosing personal information in ChatGPT (Dinev and Hart, 2006, p. 77; Xu et al., 2011, p. 50). The willingness to share highly sensitive information, such as password-protected information or credit card information in ChatGPT are reflected in two items. Additionally, one of the items (WTS_3r) is coded reverse.

Next to the six introduced constructs the model contains eight control variables to rule out alternative explanations of the constructs of interest and improve the likelihood of finding causal relationships (Klarmann and Feurer, 2018, pp. 26–28). Likewise, the control variables were chosen mainly based on established literature in the field. Typical controls such as gender, age, education, employment and income were included (Wang et al., 2016, sec. 3.1.). Regarding potential influences of prior experience with ChatGPT, its according usage intensity as well as the daily hours spent on information technology were added (Wang et al., 2016, sec. 3.1.).

3.2 Hypotheses

The following part develops the hypotheses. Table 3-1 depicts an overview of all hypotheses, the expected effects, and related references.

H(i)	Hypotheses	Expected effect	References
H1	The stronger the user's privacy concerns, the higher the perceived risk of sharing personal information in ChatGPT.	(+)	(Cichy et al., 2021; Malhotra et al., 2004; Smith et al., 1996)
H2a	Prior disclosure behaviour is negatively related to privacy concerns.	(-)	(Carrascal et al., 2013; Egelman et al., 2013; Kokolakis, 2017)
H2b	Prior disclosure behaviour is positively related to willingness to share personal data in ChatGPT.	(+)	(Keith et al., 2013; Kokolakis, 2017; Motiwalla and Li, 2016)
H3a	Trust is negatively related to perceived risks.	(-)	(Dinev and Hart, 2006; Jarvenpaa et al., 2000; Pavlou and Gefen, 2004; Treiblmaier and Chong, 2011)
H3b	Trust is negatively related to privacy concerns.	(-)	(Chellappa and Sin, 2005; Pavlou, 2011)
H3c	Trust is positively related to willingness to share personal data in ChatGPT.	(+)	(Chellappa and Sin, 2005; Dinev and Hart, 2006; Jarvenpaa et al., 2000; Pavlou and Gefen, 2004)
H4	Perceived risks are negatively related to the willingness to share personal data in ChatGPT.	(-)	(Dinev et al., 2006; Dinev and Hart, 2006; Keith et al., 2013; Pavlou and Gefen, 2004; Treiblmaier and Chong, 2011; Wang et al., 2016; Xu et al., 2011)
H5	Perceived benefits are positively related to the willingness to share personal data in ChatGPT.	(+)	(Chellappa and Sin, 2005; Wang et al., 2016; Xu et al., 2011)

Tab. 3-1: Hypotheses

Privacy concerns are often described as the “central construct in empirical studies on privacy-related behaviour” (Cichy et al., 2021, p. 1865). They are said to be highly dependent on the context, thus it is necessary to examine the construct in the specifics of ChatGPT to close the research gap and extend the body of knowledge, as no other studies have investigated this concept in the very same context yet (Cichy et al., 2021, p. 1865). They further vary across industries, technologies and individuals and especially over time with the rise of new technologies such as ChatGPT as those bring new risks and therefore require adapted evaluations of privacy concerns (Malhotra et al., 2004, p. 338; Smith et al., 1996, pp. 190–191). Referring to the previous parts, ChatGPT’s data collection methods were discussed, showing possible considerations of user’s privacy risk evaluations. For instance, a user who submitted personal information through ChatGPT and is afraid of this information being misused, unknowingly published or exploited for marketing, should perceive high levels of privacy concerns (Dinev and Hart, 2006, p. 77). As perceived risks are the manifestation of users’ risk-evaluation of potential data losses via ChatGPT, the two determinants are expected to interrelate (Malhotra et al., 2004, p. 341). Users with high levels of privacy concerns are also expected to strongly perceive risks (Malhotra et al., 2004, p. 341). Thus, the following is hypothesized:

H1: The stronger the user’s privacy concerns, the higher the perceived risk of sharing personal information in ChatGPT.

The following hypotheses H2a and H2b are both relating to the examination of the privacy paradox and are of explorative character. While the privacy paradox has been studied from various perspectives, the concept whether it is evident in ChatGPT or not, has not yet been investigated (Kokolakis, 2017; Motiwalla and Li, 2016, p. 1). For this purpose, the previously introduced construct prior disclosure behaviour was included in the research model. In general, the privacy paradox in the context of ChatGPT would be evident if the participants’ prior disclosure decisions are inconsistent with their willingness to share or their privacy concerns (Keith et al., 2013; Kokolakis, 2017; Motiwalla and Li, 2016). Thus, if one of the following hypotheses is rejected it could be an indicator for the occurrence of the privacy paradox. However, the disparity between privacy concerns and disclosure behaviour is highly debated in privacy paradox studies (see for an overview Kokolakis, 2017). While individuals state high levels of privacy concerns regarding data collection and personal information disclosure, a lot of studies

found irrational behaviour, showing them willingly disclosing this information (see for an overview Kokolakis, 2017). For instance, in some experiments individuals were ready to sell their personal data for the price of a Big Mac meal, while in other studies the participants even paid extra money to protect their privacy (Carrascal et al., 2013; Egelman et al., 2013; Kokolakis, 2017). Consequently, rational behaviour in the context of ChatGPT would mean, individuals who show high levels of privacy concerns are also showing low levels of prior disclosure behaviour. Individuals who willingly and frequently disclosed personal information in the past are thus, expected to be less concerned about their privacy. Therefore, the following hypothesis is suggested:

H2a: Prior disclosure behaviour is negatively related to privacy concerns.

Likewise, previous studies found contrasting results when examining the privacy paradox regarding intended and actual disclosing behaviour (Keith et al., 2013, p. 1170; Kokolakis, 2017; Motiwalla and Li, 2016, p. 5). Some found individuals indicating to be not willing to disclose their information, whilst paradoxically they disclosed them actually, others found the opposite behaviour (Keith et al., 2013, p. 1170; Kokolakis, 2017; Motiwalla and Li, 2016, p. 5). As in these studies mostly experiments were conducted to test first, disclosure intentions and subsequently actual disclosure behaviour in simulations, this procedure could not be replicated in the underlying paper due to limitations in the scope of the study (see for an overview Kokolakis, 2017). Therefore, the construct prior disclosure behaviour simulates this required step of measuring individuals' disclosure intentions, by quantifying their previous disclosure behaviour, while the construct willingness to share, basically simulates the actual, specific disclosure behaviour in ChatGPT (Keith et al., 2013, p. 1170; Kokolakis, 2017; Motiwalla and Li, 2016, p. 5). However, the hypothesis tested in this model is still straight-forward. If user's are indicating a pronounced prior disclosure behaviour and showing high levels of willingness to share personal data in ChatGPT, a paradoxical behaviour would not be demonstrable (see for an overview Kokolakis, 2017). On the other hand, first, a pronounced prior disclosure behaviour and low levels of willingness to share personal data in ChatGPT. Or contrary, a reserved prior disclosure behaviour and high levels of willingness to share personal data in ChatGPT, would be indicating irrational behaviour. Thus, the following is hypothesized:

H2b: Prior disclosure behaviour is positively related to willingness to share personal data in ChatGPT.

The following three hypotheses H3a, H3b and H3c are all related to the trust construct. As one of the major constructs in the related literature, it has been intensively investigated in various contexts. For instance, negative correlations between perceived privacy risks and trust were found in the context of e-commerce transactions, as well as negative correlations between the risk perception and trust in online-stores (Dinev and Hart, 2006, pp. 71–72; Jarvenpaa et al., 2000, p. 60). Likewise, in online marketplaces, trust in the community of sellers, showed significantly negative effects on the perceived risks from the community of sellers (Pavlou and Gefen, 2004, p. 50). In an extensive survey across Austria, Australia and Hong Kong, examining trust into the internet and into the online vendor, confirmed the overall negative relationship between trust and the perceived risks of personal information disclosure, while including insignificantly cross-cultural differences (Treiblmaier and Chong, 2011, p. 86). Thus, applying these insights from previous studies the following hypothesis is suggested:

H3a: Trust is negatively related to perceived risks.

A relationship between trust and privacy concerns seems to be an obvious assumption. As one of the early studies examining the effect of privacy concerns on online personalization use, Chellappa and Sin (2005, pp. 181-190) found a significant negative correlation between trust building factors and concerns for privacy, ultimately determining the likelihood of using personalization services requiring personal information disclosure. They proposed online vendors to engage in trust building measures to reduce privacy concerns (Chellappa and Sin, 2005, pp. 197–198). This ultimately would increase individuals' usage of personalization services, here thought of being representative for personal information disclosure (Chellappa and Sin, 2005, pp. 197–198). However, while the high correlation between privacy concerns and trust seems to be a consensus in literature, the "exact (causal) directionality of the relationship (…) is still a debated issue" (Pavlou, 2011, p. 981). To shed more light on this issue and to extend the debate to a recent technology and context, the following hypothesis is proposed:

H3b: Trust is negatively related to privacy concerns.

In the context of individuals' trust in online vendors and auction websites, positive relationships between trust, individuals' willingness to use and transact on those websites, mainly via reducing perceived risk, were found (Dinev and Hart, 2006, p. 66;

Jarvenpaa et al., 2000, pp. 60–64; Pavlou and Gefen, 2004, p. 50). Trust building factors, such as improvement of the brand image or companies' relationships to trusted parties, can in turn improve individuals' trust (Chellappa and Sin, 2005, pp. 196–198). Ultimately these measures further reduce privacy concerns, as well as increase the likelihood of using personalization services, which are assumed to be representative to the disclosure of personal information (Chellappa and Sin, 2005, pp. 196–198). As this relationship has not yet been tested in the context of ChatGPT, it is hypothesized:

H3c: Trust is positively related to willingness to share personal data in ChatGPT.

Perceived benefits and perceived risks represent the underlying dual path model's two major forces in the privacy calculus when it comes to individuals' disclosure decisions (Wang et al., 2016, sec. 4.1). Perceived risks of information disclosure are the counterpart of perceived benefits and reduce the overall value of the disclosure decision in the privacy calculus and thus the willingness to disclose information (Xu et al., 2011, pp. 47–50). Individuals basically trade perceived risks against hedonic or utilitarian benefits, such as personalized services or self-presentation (Wang et al., 2016, sec. 5.2.). The negative relationship between perceived risks and the willingness to share personal data has been confirmed broadly in the literature (Dinev and Hart, 2006, p. 71; Keith et al., 2013, p. 1170; Pavlou and Gefen, 2004, p. 50; Treiblmaier and Chong, 2011, p. 86). The measurement of risk perception however is difficult and thus not consistent across studies (Dinev and Hart, 2006, p. 64). Next to challenges in measurement, risk perception also varies for instance among different countries and between collectivistic societies and individualistic societies, making the construct even harder to grasp (Dinev et al., 2006, pp. 391–397; Treiblmaier and Chong, 2011, p. 87). Whether perceived benefits or risks weigh higher in the final disclosure decision is also under debate (Keith et al., 2013, p. 1170; Wang et al., 2016, sec. 5.2.). Therefore, next to testing the expected negative relationship of individuals' perceived risks and disclosure decisions in the context of ChatGPT, the following hypothesis is also trying to contribute to the debate on whether the perceived benefits or risks weigh higher in disclosure decisions. Thus, the according hypothesis is proposed:

H4: Perceived risks are negatively related to the willingness to share personal data in ChatGPT.

Following up on the previous hypothesis, the perceived benefits represent the other driving force in the dual path model. Several potential benefits of data sharing for individuals were already discussed in the beginning of thepaper. With antecedents such as personalized services or self-presentation, perceived benefits are found in literature to have a positive relationship with the likelihood of disclosing personal information or using services that require such disclosure (Chellappa and Sin, 2005, p. 196; Wang et al., 2016, sec. 4.1.; Xu et al., 2011, pp. 47–49). In the context of ChatGPT, the benefits are including mainly intangible benefits, such as reduced searching time, increased efficiency, convenience, or improved work-life balance. However, relating to the previous hypothesis, the respective weight of perceived benefits in the privacy calculus, as well as the confirmation of an expected, positive relationship of perceived benefits on the willingness to share personal data in ChatGPT, remains open for debate (Keith et al., 2013, p. 1170; Wang et al., 2016, sec. 5.2.). The following hypothesis is therefore suggested:

H5: Perceived benefits are positively related to the willingness to share personal data in ChatGPT.

4 Methodology

The methodology part starts by deriving and explaining the study design and the subsequent data collection. After that, several modifications regarding the dataset are performed and clarified. Last, the resulting sample is described.

4.1 Study Design

To measure the previously introduced constructs within the context of data sharing in ChatGPT and its plugins, a cross-sectional survey design is chosen (Connelly, 2016, pp. 369–370; Wooldridge, 2015, p. 5). This allows to capture a snapshot of participants' attitudes, perceptions, and behaviours in a single instance and provides a holistic view (Connelly, 2016, pp. 369–370; Wooldridge, 2015, p. 5). Further, it is time- and cost-efficient as it does not require follow-up data, such as a panel dataset (Connelly, 2016, p. 369). The research approach tends more towards a deductive approach based on the following reasons (Newman, 2000, p. 3). It begins with a framework, consisting of established sharing theories, constructs, and the dual-path model (Newman, 2000, p.

3). Subsequently, general hypotheses are conceptualized from related literature, leading ultimately to the specific hypotheses presented in the previous parts (Newman, 2000, p. 3). Data collection and analysis through the underlying survey generate results to confirm or revise hypotheses and draw conclusions on importance and interrelations of the determinants (Newman, 2000, p. 3).

The introductory and background sections in the survey set clear expectations, illustrate the context, and guide participants to respond within the established frameworks (Newman, 2000, pp. 3–10). The survey design is methodically structured to cover different dimensions of participants' perceptions and behaviours and aligns with the research objectives presented above. Each block contains statements that participants rate on a Likert scale, based on their agreement, providing quantitative data for analysis (Newman, 2000, pp. 3–10). Additionally, including control variables like usage frequency, demographics, and device usage reduces potential biases by considering variables that might affect the relationships among the main constructs (Wooldridge, 2015, pp. 21–24).

The following section refers to A-6, depicting the full survey from Qualtrics in its obtainable print format. Likewise, the exact constructs and items mentioned are to be found in A-5, showing the references. The survey starts with a general introduction thanking participants for taking part in the survey and explaining purpose, guidelines, and affiliations. Data protection and anonymity in the survey is emphasized and it is pointed out to refer to personal experiences made with ChatGPT, if applicable. Further, participants are informed on the potential to take part in a raffle for three 20 Euro Amazon vouchers. Ultimately, an approximate conduction time of five to eight minutes is proposed, further explained in the next sections.

The next survey block provides background information on ChatGPT and its plugins. The functioning of ChatGPT, its role as an intermediary between users and plugins, and the types of data collected by both ChatGPT and plugins are explained to align all users’ knowledge. Further, examples of collected data by ChatGPT and plugins are provided. Finally, the section highlights potential uses of collected data, such as improving services, personalized communication, or marketing.

Before progressing to the question blocks a control question was implemented following the background information to test whether the participants attentively read them.

This control question should reduce the rate of participants who would rush through the survey without reading instructions and subsequent questions properly (Meade and Craig, 2012, p. 7). As the survey collected a large part of its total participants from SurveyCircle this is especially important, since the point-based reward system on the platform could motivate users to rush time-effectively through other surveys to faster collect points.

After filtering out potential disruptive factors, the survey then progresses to the main blocks of questions.

First the prior disclosure behaviour block lets the participants rate their willingness to share several types of personal information through ChatGPT, such as health-related data, behavioural data, and other sensitive information.

Second, perceived benefits are measured by letting the participants rate the extent to which they agree with statements regarding various benefits of using ChatGPT, such as improving efficiency, work-life balance, and easy access to information.

In the third block, perceived risks, participants rate their perceptions of risks associated with sharing personal data through ChatGPT, including potential problems, disclosure risks, and uncertainties.

Similarly, the fourth block measuring privacy concerns displays statements expressing concerns about potential misuse of submitted information, privacy invasion, and unforeseen usage.

In the fifth block, participants rate their level of trust in ChatGPT regarding reliability, promises, handling of personal data, and providing a secure environment for transactions.

Last, the main construct willingness to share personal data in ChatGPT is measured by letting participants rate their likelihood and tendency to disclose personal data through ChatGPT for various purposes, such as conducting sales transactions or sharing highly personal information for access to services.

As mentioned earlier, the eight control variables collect information about participants' usage of ChatGPT, frequency of use, device usage, gender, age, education, employment status, and finally income. Further, having the option to submit feedback, participants are encouraged to give comments and suggestions regarding the survey in the

last steps. Finally, participants can enter the raffle by submitting their e-mails on a second, detached survey to not compromise their anonymity.

Summarizing the survey design follows a logical flow, starting with introduction, background information, and then diving into different aspects of participants' attitudes and behaviours regarding data sharing through ChatGPT. The question blocks cover the earlier introduced constructs, including perceived benefits, risks, privacy concerns, trust, and willingness to share personal data. The control variables section captures demographic information that could influence participants' responses.

4.2 Data Collection and Sample

The survey was conducted fully online via Qualtrics in the period from 17.07.2023 until 09.08.2023 and collected a total of 357 recorded responses. Before its release, comprehensibility, conciseness and conclusiveness of the questions and instructions were ensured by pre-testing the survey among the circle of acquaintances.

The population of interest in this study is the set of all individuals at the time of the data collection and is not further characterised by any specific requirement, such as a certain age group, profession or similar. The only characteristic that differentiates the underlying population is the requirement of speaking English, as the survey was conducted fully in English. The analysed sample therefore should represent a cross section of the society, however, due to the language barriers, it only can represent a cross section of English-speaking individuals.

Based on the description of the data collection methods, the according sampling strategies could therefore be best described as convenience sampling and self-selection sampling (Henry, 1990, pp. 3–4; Sharma, 2017, p. 752). Convenience sampling is mainly selecting participants “based on their availability for the study” (Henry, 1990, p. 4). Further, individuals’ voluntary participation is necessary for being included in the sample. With platforms, such as SurveyCircle or PollPool, reached individuals are self-selecting into studies they consider either interesting or attractive in terms of credit points achievable and therefore fulfil several important characteristics of self-selection sampling (PollPool, 2023; Sharma, 2017, p. 752; SurveyCircle, 2023). The sampling strategies were mainly chosen due to the constraint cost budget and a tight schedule (Sharma, 2017, pp. 749–752). However while the recruitment still includes a certain randomness, since the study was accessible to everyone, distributed across several

sources with a wide dispersion, such as Instagram, LinkedIn and flyers, it does not satisfy the random sampling condition that “each member of the study population has an equal probability of being selected” (Henry, 1990, p. 15). Several problems and disadvantages that are accompanying the chosen sampling strategies are discussed in the limitations section.

Before the dataset is analysed, several steps are conducted to ensure a valid final sample. The number of total responses in Qualtrics before filtering was 357. Before downloading the dataset, following filters in Qualtrics were applied. First, 73 unfinished responses are excluded, subsequently, 53 of the participants answered the control question wrong and therefore are excluded. Thus, a total of 231 answers remains in the downloaded sample.

The next steps were then performed in Microsoft Excel. An analysis regarding the participants’ time spent on the survey is conducted to filter for respondents who either likely rushed through the survey, did likely not finish the survey attentive, or in one session, and therefore exceed a certain duration threshold (Meade and Craig, 2012, pp. 2–7). Since the determination of this threshold is highly context dependant and literature gives no definite answer, the following assumptions were made (Meade and Craig, 2012, pp. 2–7). The pre-tests conducted before publishing the survey showed an expected duration of approximately five to eight minutes, which also justified the recommendation in the survey introduction on Qualtrics. After getting feedback and taking into account characteristics such as fast and slow readers, or interested and uninterested participants, the expected duration time for the data analysis therefore was afterwards assumed to range approximately between three to ten minutes (Meade and Craig, 2012, pp. 2–7). Subsequently, from a quantitative perspective, quantiles were analysed and revealed that the 10 %- and 90 %-quantiles correspond to a minimum duration of 3.1 minutes and a maximum duration of 10.8 minutes (Meade and Craig, 2012, pp. 2–7). Removing the outlying responses, 185 answers remained.

The next step tried to eliminate careless respondents, so-called straight-liners or long-string answers, who are showing identical answer patterns (Meade and Craig, 2012, p. 7). For example, a participant answering in the construct trust for all five items with a seven on the according Likert-scale. This behaviour could indicate for instance inattentive, careless reading or not reading the questions at all and could ultimately distort

the results (Meade and Craig, 2012, p. 7). This could be specifically the case for participants who come from SurveyCircle or PollPool since those individuals may trade-off platform credit points against time invested. However, several individuals remarked in pre-tests as well as in the survey feedback the perceived similarity of response options. While this was intentionally applied to measure consistent constructs, the participants might felt repetitive or redundant character of questions or answers and therefore responded in a repetitive pattern (Meade and Craig, 2012, p. 7). To not accidentally remove these attentive respondents, only straight-liners who showed suspicious behaviour across more than three constructs were removed. This led to excluding only one respondent who straight-lined across four constructs.

After cleaning the dataset, the following steps corrected problematic variable codes to ensure an error-free processing in smartPLS (2022).

First, the reversed items were recoded by deducting the measured values from eight. This is equivalent to replacing manually the opposite value with, e. g., an IF-function in Excel. For labelling those inverted items, their suffixes were changed from *r* to *inv*.

Second, some control variables were recoded. Usage was recoded to a dummy-variable indicating one for used, and zero for not used. Intensity and activity did not need any adaptions, as the two variables are already measured in a contextual ascending logic, with higher numbers representing higher usage intensities or higher levels of activity. Analysing the control variable gender, the answers revealed only three individuals in the categories Non-Binary/Third Gender and three in Prefer not to say. For simplifying the data analysis these six values were treated as missing values and the remaining categories male and female were coded as a dummy, one for male and zero for female. Age and income were already measured in an ascending logic and therefore needed no further adaptions. However, one individual in the control variable age and 16 in the control variable income chose the option Prefer not to say, which were subsequently treated as missing values, too. The control variable employment was entirely coded new as the initial coding from Qualtrics would lead to non-interpretable results. As it is difficult to determine a logic or qualitative order between different employment choices, a subjective, ascending order in terms of working level and intensity was applied. Thus, "Unemployed" becomes 1, "School Student" becomes 2, "Part-Time Employed" becomes 3, "Apprenticeship" becomes 4, "University Student" becomes 5, "Self-Employed" becomes 6, "Full-Time Employed" becomes 7, and "Prefer

not to say" becomes a missing value. As mentioned, this definition is subjective and can be argued and is done solely for the purpose of simplifying data interpretation and analysis.

The following section presents the descriptive statistics derived from the survey responses provided by the participants based on the above introduced, cleaned dataset. To simplify and consistently report the following descriptive statistics, the above introduced missing values are herein summarized in the category *other*. The according graphics are accessible in A-7 to A-14.

Regarding the utilization of ChatGPT, respondents were asked about their familiarity with the technology. Most participants, 90 %, indicated that they have used ChatGPT at some point before. However, 16 % of the respondents reported using ChatGPT less than once a month and 21 % indicated a monthly usage pattern. In contrast, 20 % reported using ChatGPT at least once a week, 22 % several times a week and lastly, 10 % of participants reported using ChatGPT every day. Thus, the majority uses ChatGPT quite regularly.

Concerning technological devices usage, participants were questioned about their daily use of devices such as smartphones, computers, and tablets. The analysis found 5 % of respondents using such devices for less than an hour each day. A slightly higher proportion, namely 7 %, reported a usage duration of one to two hours. Further, 10 % of participants reported two to three hours daily and 18 % indicated a usage duration of three to four hours daily. While 13 % reported using these devices for four to five hours each day, impressively, nearly half of the respondents, namely 47 %, reported using these devices for more than five hours daily.

Examining the distribution of gender among the participants, the data revealed 40 % male respondents and a significant majority of 57 % female. 3 % of the participants were grouped in the category other.

The age distribution of participants showed diverse age groups. Only 1 % fell within the age group below 16 years and 2 % comprised individuals between 17-20 years. The majority, namely 75 % participants, were aged between 21-30 years. Additionally, 14 % were aged 31-40 years, while a smaller share of 3 % fell within the age groups of 41-50 and 51-60 years. A likewise small percentage of 1 % represented individuals over 60 years, while another 1 % preferred not to disclose the age and therefore was

categorized as other. Thus, the sample is relatively young, making the following statistic regarding employment status plausible.

3 % of the participants are school students, while the great majority, 60 %, consists of university students. Notably, no respondent indicated being in an apprenticeship. More than a quarter, 26 %, were employed on a full-time basis, and a smaller proportion, 6 %, reported being part-time employed. A minor proportion of 3 % reported being self-employed, and 2 % indicated unemployment. Similarly, 1 % preferred not to disclose their employment status and were categorized as other.

Concerning the participants' highest level of education, the analysis indicated that a significant proportion, 47 %, held a bachelor's degree. Additionally, 23 % reported having a master's degree, diploma, or state examination. Furthermore, 15 % held qualifications enabling admission to universities of applied sciences and a smaller percentage, 5 %, completed vocational training programs. Equally, 5 % held doctoral degrees, while 2 % held advanced technical college entrance qualifications. Even smaller proportions of 2 % and 1 % held only intermediate and secondary school leaving certificates.

Lastly, participants were asked to indicate their approximate monthly income after tax. The analysis showed that 17 % had a monthly income below 500 €, while 21 % reported an income between 501 € and 1000 €. Moreover, 27 % reported an income ranging from 1001 € to 2000 €, and 16 % had an income from 2001 € to 3000 €. Notably, 10 % had an income exceeding 3001 €. Additionally, 9% preferred not to disclose their monthly income and were categorized as other.

In conclusion, the descriptive statistics provided, offer a better understanding of the demographics and usage patterns surrounding ChatGPT among the participant group. Considering the major share of students among the participants the data seems plausible, consistent and well-balanced.

The smartPLS (2022) analyses of indicators and correlations, depicted in appendices A-15 to A-16b, reveal standard deviations varying from as low as 0.304, for the control variable usage, to as high as 1.861, for item 4 of prior disclosure behaviour. This range of values suggests that some variables have relatively low dispersion, indicating that the data points are closer to the mean (Wooldridge, 2015, pp. 656–658). Variables with higher standard deviations, for instance income or employment, have more spread

values, implying greater variability (Wooldridge, 2015, pp. 656–658). Standard deviations close to zero suggest that the values are clustered around the mean and could therefore indicate consistency within those items and variables, however this is not the case in the underlying dataset, considering 0.304 as the smallest observable standard deviation (Wooldridge, 2015, pp. 656–658). Further, for the majority of items and variables negative excess kurtosis values are observed, indicating flatter and more spread distributions compared to a normal distribution (DeCarlo, 1997, pp. 292–294; Wooldridge, 2015, p. 658). This suggests lighter tails and more evenly spread data points (DeCarlo, 1997, pp. 292–294; Wooldridge, 2015, p. 658). With a range of excess kurtosis values varying from -1.885 to 6.188 the data's distribution is quite diverse across the items and variables (DeCarlo, 1997, pp. 292–294; Wooldridge, 2015, p. 658).

5 Results

This part covers the results based on the introduced sample. It starts with justifying the methodological approach, continues by an assessment of the measurement model and finishes with the structural model analysis.

For the study, a partial least squares structural equation model via smartPLS (2022) was chosen. SmartPLS is a so-called second-generation statistical method with several advantages over covariance-based structural equation models, usually conducted with LISREL or AMOS (Jahn, 2007, pp. 11–16; SmartPLS, 2022; Wang et al., 2016, sec. 3.3).

PLS-SEM ensures even with smaller sample sizes high statistical power, while larger sample sizes increase consistency and precision (Hair Jr et al., 2021, pp. 15–16). However, the exact definition of a small or large sample is still under debate, thus an relatively inaccurate rule of thumb can at least give a clue (Hair Jr et al., 2021, p. 16). This recommended sample size would correspond in the underlying model to around 130 participants, assuming six constructs and eight control variables (Hair Jr et al., 2021, p. 16). Thus, the underlying sample of 184 fulfils this rule of thumb. In comparison LISREL or AMOS would need bigger samples for similar conditions (Jahn, 2007, p. 16).

Another advantage of smartPLS is the application without distributional assumptions while still giving robust results, at least if the share of missing values is below five percent (Hair Jr et al., 2021, p. 12; Jahn, 2007, p. 15). As above-mentioned, only 26

values were coded and thus treated as missing values in a total of 6992 data points. The share of missing values therefore corresponds to only 0.37%.

In general, PLS-SEM further allows for multi-item measures, reflective measurement models, complex relationships and therefore fits the requirements introduced by research question two (Hair Jr et al., 2021, p. 12).

In conclusion, “PLS-SEM-based model estimation and assessment follow a causal-predictive paradigm, in which the objective is to test the predictive power of a model, derived from theory and logic” (Hair Jr et al., 2021, p. 14). For analyses of the measurement model and structural model two different calculation methods are used in smartPLS (Thürmel et al., 2021, p. 4256).

First, the PLS-SEM calculation is used for the reflective measurement model, focussing on the assessment of reliability and validity (Thürmel et al., 2021, p. 4256). Indicator reliability, internal consistency reliability, convergent validity and divergent validity are investigated (Hair Jr et al., 2021, pp. 76–80).

Second, the bootstrapping calculation method is used for the structural model (Hair Jr et al., 2021, pp. 116–123; Thürmel et al., 2021, p. 4256). The structural model focuses on understanding the results and relationships between the constructs, assessing their effects on each other, estimate path coefficients, assess model fit, and test hypotheses (Hair Jr et al., 2021, pp. 116–123). Thus, first the measurement model should confirm its validity which then allows to move on to the bootstrapping calculation method for the structural model (Hair Jr et al., 2021, p. 116).

5.1 Measurement Model Analysis

A-17 shows the development of the path model in smartPLS before running calculations. Each of the six constructs is depicted as a grey square with a blue frame, while each reflective construct’s items are depicted as light grey rectangles with black frames. Further, the eight control variables are depicted as white circles with a dark blue frame. The arrows between the constructs are representing the previously introduced effects and hypotheses.

The PLS-SEM algorithm is run on a path weighting scheme, standardized type of results, the default setting for the initial weight and mean replacement for missing values as this procedure is applicable considering reasonable limits (Hair Jr et al., 2021, p.

20). Further, the 26 missing values are only affecting the control variables in the model and none of the constructs.

First, the outer loadings, or reflective indicator loadings, are checked to ensure indicator reliability (Hair Jr et al., 2021, p. 77). According to literature, a threshold of above 0.708 confirms that the construct is explaining at least 50 % of the indicator's variance (Hair Jr et al., 2021, p. 77). The following table 5-1 shows the results.

Constructs/items	**Outer loadings**
PBENEFITS_1 <- Perceived benefits	0.898
PBENEFITS_2 <- Perceived benefits	0.860
PBENEFITS_3 <- Perceived benefits	0.885
PBENEFITS_4 <- Perceived benefits	0.808
PBENEFITS_5 <- Perceived benefits	0.885
PC_1 <- Privacy concerns	0.891
PC_2 <- Privacy concerns	0.850
PC_3 <- Privacy concerns	0.899
PC_4 <- Privacy concerns	0.914
PC_5 <- Privacy concerns	0.818
PDB_1 <- Prior disclosure behaviour	0.602
PDB_2 <- Prior disclosure behaviour	0.810
PDB_3 <- Prior disclosure behaviour	0.769
PDB_4 <- Prior disclosure behaviour	0.775
PDB_5 <- Prior disclosure behaviour	0.804
PRISK_1 <- Perceived risk	0.689
PRISK_2 <- Perceived risk	0.810
PRISK_3 <- Perceived risk	0.774
PRISK_4inv <- Perceived risk	0.675
PRISK_5 <- Perceived risk	0.735
TRUST_1 <- Trust	0.788
TRUST_2inv <- Trust	0.710
TRUST_3inv <- Trust	0.601
TRUST_4 <- Trust	0.583
TRUST_5 <- Trust	0.718
WTS_1 <- Willingness to share	0.877
WTS_2 <- Willingness to share	0.911
WTS_3inv <- Willingness to share	0.475
WTS_4 <- Willingness to share	0.809
WTS_5 <- Willingness to share	0.788

Tab. 5-1: Outer loadings

The great majority of indicator loadings is above the threshold, however PDB_1, PRISK_1, PRISK_4inv, TRUST_3inv, TRUST_4 and WTS_3inv are below the threshold. For loadings in the range of 0.40 and 0.70 it is recommended to investigate whether removal of those below-the-threshold indicators affects reliability or validity

measures, done in the next steps, while only loadings of below 0.40 should be removed directly (Hair Jr et al., 2021, p. 77). However, this model does not contain indicators with such a weak loading. Thus, internal consistency reliability and validity measures are assessed in the next step, depicted in table 5-2, to further investigate the values in the above-mentioned range closer.

Constructs	**Cronbach's alpha**	**Composite reliability (rho_a)**	**Composite reliability (rho_c)**	**Average variance extracted (AVE)**
Perceived benefits	0.918	0.918	0.938	0.753
Perceived risk	0.791	0.794	0.856	0.545
Prior disclosure be-haviour	0.809	0.817	0.868	0.572
Privacy concerns	0.923	0.931	0.942	0.766
Trust	0.715	0.738	0.813	0.468
Willingness to share	0.834	0.877	0.887	0.620

Tab. 5-2: Construct reliability and validity - Overview

The four measures depicted are assessing “the extent to which indicators measuring the same construct are associated with each other” (Hair Jr et al., 2021, p. 77). Cronbach’s alpha, composite reliability rho_a as well as rho_c fulfil the associated thresholds of each measure (Hair Jr et al., 2021, p. 77). However, the assessment of the convergent validity, measured by the average variance extracted, shows a value of less than 0.50 for the construct trust (Hair Jr et al., 2021, pp. 77–78). The recom-mended approach is to delete the construct’s indicators with the lowest outer loadings, one by one, and subsequently assess the above-mentioned values again (Hair Jr et al., 2021, pp. 77–78). After removing the low-loaded indicators PDB_1, PRISK_1, PRISK_4inv and WTS_3inv the analysis reveals for nearly all constructs’ reliability and validity measures higher values, depicted in table 5-3.

Constructs	**Cronbach's alpha**	**Composite reliability (rho_a)**	**Composite reliability (rho_c)**	**Average variance extracted (AVE)**

Perceived benefits	0.918	0.918	0.938	0.753
Perceived risk	0.758	0.757	0.861	0.675
Prior disclosure behaviour	0.821	0.823	0.882	0.651
Privacy concerns	0.923	0.928	0.942	0.766
Trust	0.688	0.709	0.809	0.516
Willingness to share	0.880	0.886	0.918	0.736

Tab. 5-3: Construct reliability and validity – Overview after removal of low-loaded indicators

The average variance extracted for the construct trust is after removal of indicator TRUST_4 higher and exceeds the threshold of 0.50 (Hair Jr et al., 2021, pp. 77–78). However, Cronbach's alpha for the construct trust is now below the threshold. Thus, the decision whether to delete TRUST_4 or not corresponds to deciding between a valid average variance extracted or a valid Cronbach's alpha regarding the construct trust. Nevertheless, Cronbach's alpha is described in literature as a conservative measure for internal consistency reliability and in case of ambiguity the reliability coefficient rho_a can be considered an "acceptable compromise" (Hair Jr et al., 2021, pp. 77–78). Further, in the underlying analysis Cronbach's alpha only decreases slightly from 0.715 to 0.688 (delta -0.027), while the average variance extracted increases simultaneously from 0.468 to 0.516 (delta +0.048) after removal of TRUST_4. A comparative analysis of equally removing low-loaded indicators TRUST_3inv or TRUST_5 revealed even worse results and is therefore not further considered. Thus, trying to balance the trade-off, only the above-mentioned indicators are removed from the model (Hair Jr et al., 2021, pp. 77–78).

In the last step of the evaluation of the measurement model the discriminant validity is assessed to analyse the distinction of a construct from other constructs in the model (Hair Jr et al., 2021, pp. 78–79). While the Fornell-Larcker criteria is often chosen for this purpose, it is not recommended in literature, especially if the outer loadings are only slightly differing between indicators (Hair Jr et al., 2021, pp. 78–79). This is quite the case in the underlying model, especially considering the list of outer loadings after removing the low-loaded indicators. Therefore, the heterotrait-monotrait ratio is assessed, depicted in the bar chart in A-18 (Hair Jr et al., 2021, pp. 78–79).

Nearly all heterotrait-monotrait ratios are considerably below the two thresholds recommended, namely 0.90 for conceptually similar constructs, which most constructs could be identified as in this study, and 0.85 for conceptually different constructs (Hair Jr et al., 2021, pp. 79–80). Only privacy concerns with perceived risks reveal a threshold-exceeding value, namely 0.856 (Hair Jr et al., 2021, pp. 79–80). While it could be debated whether the constructs privacy concerns and perceived risk are conceptually similar or not, the more conservative threshold of 0.85 is at least exceeded by 0.006 (Hair Jr et al., 2021, pp. 79–80). Therefore, the above-mentioned Fornell-Larcker criteria, comparing each construct's average variance extracted with the inter-construct correlations, is assessed additionally in table 5-4 (Hair Jr et al., 2021, pp. 78–79).

Constructs	Perceived benefits	Perceived risk	Prior disclosure behaviour	Privacy concerns	Trust	Willingness to share
Perceived benefits	0.8678					
Perceived risk	-0.3040	0.8216				
Prior disclosure behaviour	0.4810	-0.1320	0.8068			
Privacy concerns	-0.3060	0.7190	-0.1240	0.8752		
Trust	0.6260	-0.5170	0.4580	-0.5900	0.7183	
Willingness to share	0.4040	-0.3110	0.5740	-0.2090	0.4520	0.8579

Tab. 5-4: Fornell-Larcker analysis in Microsoft Excel

To fulfil the Fornell-Larcker criteria, the assessed values on the diagonal have to be higher than all values of the same row and column in the matrix (Hair Jr et al., 2021, pp. 78–79). This condition is true for all constructs and thus confirms the constructs' discriminant validities from a second perspective (Hair Jr et al., 2021, pp. 78–79).

Summarizing, the measurement model analysis assessed reflective indicator loadings, internal consistency reliability, convergent and discriminant validity. After several adjustments nearly all recommended metrics and thresholds are fulfilled and minor issues remaining are underlined but tolerated (Hair Jr et al., 2021, p. 80). Consequently it can be noted that "the measurement of constructs is reliable and valid" (Hair Jr et al., 2021,

p. 116). Therefore, it is proceeded to the evaluation of the structural model in the next section.

5.2 Structural Model Analysis

First, the structural model is assessed for collinearity issues by checking the VIF, depicted in table 5-5 (Hair Jr et al., 2021, p. 117).

Constructs/Items	VIF
CV_ACTIVITY -> Willingness to share	1.051
CV_AGE -> Willingness to share	1.240
CV_EDUCATION -> Willingness to share	1.307
CV_EMPLOYMENT_recoded -> Willingness to share	1.326
CV_GENDER_recoded -> Willingness to share	1.161
CV_INCOME -> Willingness to share	1.415
CV_INTENSITY -> Willingness to share	1.915
CV_USAGE -> Willingness to share	1.497
Perceived benefits -> Willingness to share	2.428
Perceived risk -> Willingness to share	1.473
Prior disclosure behaviour -> Privacy concerns	1.265
Prior disclosure behaviour -> Willingness to share	1.474
Privacy concerns -> Perceived risk	1.533
Trust -> Perceived risk	1.533
Trust -> Privacy concerns	1.265
Trust -> Willingness to share	2.381

Tab. 5-5: VIF – Inner model

All VIF values for the inner model are far below the critical level of five and therefore indicate no collinearity problems (Hair Jr et al., 2021, p. 117). Likewise, the VIF values for the outer model are all below five, illustrated in A-19. Further, A-20 shows the measurement model with outer loadings after removal of low-loaded indicators, the constructs' R-squared values and the inner model showing the path coefficients and p values, for the non-bootstrapping calculation method.

However, for evaluating the significance and relevance of path coefficients in the structural model, the bootstrapping technique is applied (Hair Jr et al., 2021, pp. 116–123). Bootstrapping refers to a resampling technique that provides more robust estimates of parameter distributions (Hair Jr et al., 2021, pp. 116–123). It allows to calculate confidence intervals and significance tests by taking the underlying sample and re-sample it for the number of indicated subsamples (Hair Jr et al., 2021, pp. 116–123). By using

this technique, more accurate p-values and confidence intervals for the path coefficients and other model parameters can be achieved (Hair Jr et al., 2021, pp. 116–123). For the following calculation 5000 subsamples, bias-corrected and accelerated bootstrap, two-tailed test type, path weighting scheme, case-wise deletion and a significance level of 0.05 were applied (Thürmel et al., 2021, p. 4256). The random number generator was set on fixed seed and parallel processing as well as generation of results per sample were allowed. The most important results are visualized in figure 5-1, the detailed version can be found in A-21.

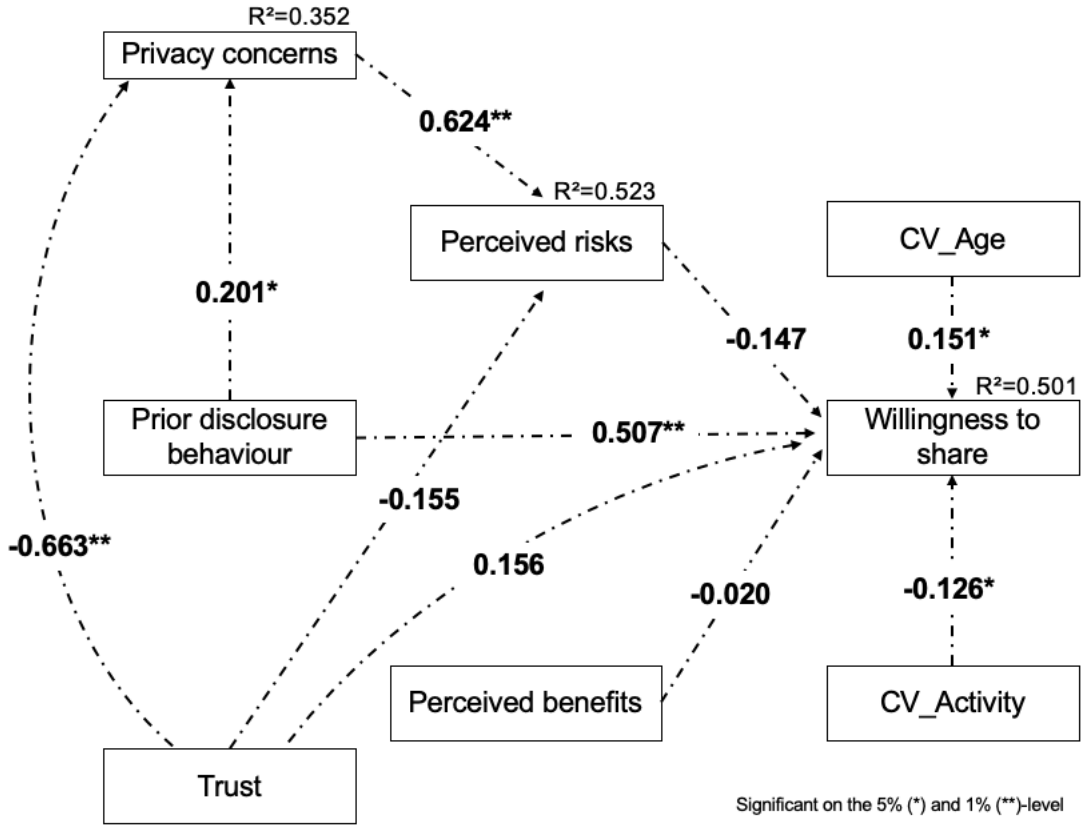


Fig. 5-1: PLS-SEM bootstrapping results (n=184; subsamples=5000; *p<0.05; **p<0.01)

The figure shows the constructs' R-squared values and path coefficients with their significance levels indicated by asterisks. The explanatory power of the model, here measured by R-squared of the endogenous constructs, shows following results (Hair Jr et al., 2021, p. 118). The model explains 35.2 % of the variation in privacy concerns, 52.3 % of the variation in perceived risk and 50.1 % in the variation of willingness to share. Thus, the explanatory power of perceived risk and willingness to share is considered to be moderate, while privacy concerns' explanatory power is considered weak (Hair Jr et al., 2021, p. 118). However, R-squared has to be interpreted in the study's

context (Hair Jr et al., 2021, p. 118). Very high R-squared values in models predicting individuals' perceptions or attitudes, as the underlying model does, would likely imply so-called model overfit (Hair Jr et al., 2021, pp. 118–119). Therefore the measured R-squared values can be considered appropriate (Hair Jr et al., 2021, pp. 118–119).

Next to figure 5-1, A-22 gives the detailed path coefficient table with means, standard deviations, T- and p-values tested on the 0.05 significance level. The effect of privacy concerns on perceived risk is positive and statistically significant on the 1 %-level (p=.000). Hypothesis H1 is therefore supported.

For hypothesis H2a the expected negative effect of prior disclosure behaviour on privacy concerns is on the contrary positive and significant on the 5 %-level (p=.025). Individuals who showed a tendency to willingly disclose information in the past however also show high levels of privacy concerns. The hypothesis is therefore not supported. Prior disclosure behaviour is further significantly positively related to willingness to share personal data in ChatGPT on the 1 %-level, supporting H2b (p=.000).

Regarding the construct trust however, only one of the three hypotheses made, is supported by the results. While the effect of trust on perceived risks is indeed negative, it is not significant on the 5 %-level (p=.053). Thus, H3a is not supported. With its p value close to the 5 %-threshold, tested on the 10 %-level the same hypothesis however would be supported. H3b, stating that trust is negatively related to privacy concerns, is supported by a strong and significant negative effect of trust on privacy concerns on the 1 %-level (p=.000). Nevertheless, H3c hypothesizing a positive relationship of trust to willingness to share personal data in ChatGPT is not supported (p=.214).

For hypothesis H4 a negative effect of perceived risks on willingness to share was expected and found, although, the effect is not significant and H4 is not supported (p=.130).

H5 hypothesized a positive relationship between perceived benefits and the willingness to share. Unexpectedly, the results reveal a negative relationship of perceived benefits on willingness to share, though, a not significant one (p=.830). Therefore, H5 is not supported.

Besides activity and age, all other control variables' effects on willingness to share are insignificant. Though, the controls reveal that individuals who spent more time per day using smartphones, laptops and other devices are significantly less likely to disclose

their personal information in ChatGPT (p=.033). Further, a higher age is associated with a higher willingness to share personal information in ChatGPT (p=.041).

Table 5-6 shows a summary of the hypotheses, expected and measured effects, and whether the hypothesis was supported on the according significance levels or not.

H(i)	**Hypotheses**	**Expected effect**	**Measured effect**	**Hypothesis supported (*p<0.05; **p<0.01)**
H1	The stronger the user's *privacy concerns*, the higher the *perceived risk* of sharing personal information in ChatGPT.	(+)	(+)	Yes** (p=.000)
H2a	Prior disclosure behaviour is negatively related to privacy concerns.	(-)	(+)	No* (p=.025)
H2b	Prior disclosure behaviour is positively related to willingness to share personal data in ChatGPT.	(+)	(+)	Yes** (p=.000)
H3a	Trust is negatively related to perceived risks.	(-)	(-)	No (p=.053)
H3b	Trust is negatively related to privacy concerns.	(-)	(-)	Yes** (p=.000)
H3c	Trust is positively related to willingness to share personal data in ChatGPT.	(+)	(+)	No (p=.214)
H4	Perceived risks are negatively related to the willingness to share personal data in ChatGPT.	(-)	(-)	No (p=.130)
H5	Perceived benefits are positively related to the willingness to share personal data in ChatGPT.	(+)	(-)	No (p=.830)

Tab. 5-6: Summary of hypotheses

Additional to the previous analyses, referring to the posed research questions, a post-hoc analysis as an add-on is conducted to investigate further interesting relationships, provoke thoughts for future research and fully exploit the data. The analysis' results are depicted in A-23, using the same bootstrapping settings as the previously explained model in the main analysis. The effects and insights are further discussed at the end of the following discussion in part six.

6 Discussion

The following part discusses the introduced results, conducts a post-hoc analysis, and gives several implications for practitioners as well as theorists. Finally, limitations of the underlying approach are debated, and future research recommendations are given.

This study investigated what determinants are more relevant in individuals' data sharing decisions and how they are interrelated. It built on the established literature and applied the dual-path research model on a novel technology. As little is known of privacy and disclosure decisions in the context of ChatGPT, one of the currently most prominent and promising LLM, the study context was set accordingly. Privacy calculus' two driving forces, represented by perceived benefits and perceived risks, intended to explain individuals' disclosure decisions in ChatGPT (Wang et al., 2016, sec. 4). Besides this, trust and privacy concerns as central constructs in the privacy literature were included in the model to assess their relevance and interrelations (Cichy et al., 2021, p. 1870; Ioannou et al., 2020, sec. 2.3; Smith et al., 2011, p. 172; Song and Kim, 2021, p. 4). Further, prior disclosure behaviour as a construct was introduced to finally measure whether the privacy paradox is also relevant in the context of ChatGPT (Motiwalla and Li, 2016, sec. 3.4).

Several expectations were confirmed by the study. The highly context-dependent privacy concerns were expected to increase with the level of perceived risk of sharing personal information in ChatGPT (Malhotra et al., 2004, p. 341). Consistently, the results indicated a strong positive relationship of privacy concerns on perceived risks. With strongly varying levels of privacy concerns across industries, contexts and technologies the positive relationship between the two constructs in ChatGPT is thereby confirmed (Malhotra et al., 2004, p. 338; Smith et al., 1996, pp. 190–191). This finding

is consistent with similar studies in different contexts, that also found positive relationships between these two constructs (Dinev et al., 2006, p. 396; Featherman and Pavlou, 2003, p. 467; Fortes et al., 2017, p. 322). The results indicate that individuals who are highly concerned about submitting personal information in ChatGPT, due to fear of potential misuse, unauthorized publishing, or privacy invasion because of marketing, are also experiencing high levels of perceived risks. In the dual path model, with perceived risks as direct antecedent of the willingness to share personal data in ChatGPT, privacy concerns therefore could decrease ultimately the willingness to share via significantly increasing perceived risks. However, here it must be noted that the negative effect of perceived risk on the willingness to share has not been proven significant in this study.

As the privacy paradox had not yet been investigated in the context of ChatGPT before, this study examined the relation between prior disclosure behaviour and privacy concerns, as well as the willingness to share (Kokolakis, 2017; Motiwalla and Li, 2016, p. 1). In the previous sections it was argued that either the rejection of hypothesis H2a or H2b could indicate the presence of the privacy paradox concerning ChatGPT. Interestingly, H2b was not rejected and the positive relationship between prior disclosure behaviour and willingness to share personal data in ChatGPT was confirmed. The expected positive effect, implying no irrational or paradoxical behaviour, was found. This indicates that individuals who were willingly disclosing personal information in the past online also show high willingness to share personal data in ChatGPT (Kokolakis, 2017). However, on the contrary, H2a was rejected, revealing the opposite of the expected effect of prior disclosure behaviour on privacy concerns. In the hypotheses section the case was made, that rational behaviour would be evident if individuals who have a conservative prior disclosure behaviour online, should also show high levels of privacy concerns (Carrascal et al., 2013; Egelman et al., 2013; Kokolakis, 2017). Paradoxically in this study, individuals who showed high levels of prior disclosure behaviour, equivalent to willingly disclosing personal information online in the past, on the other hand indicated also high levels of privacy concerns. Therefore these ambiguous insights add to the contradicting results in previous studies (Keith et al., 2013, p. 1170; Kokolakis, 2017; Motiwalla and Li, 2016, p. 5). While in this study H2b indicates no privacy paradox, H2a on the other hand gives evidence for the privacy paradox. It could be debated, whether one hypothesis overweighs the other, however in the hypotheses

section it was argued that one rejected hypothesis is considered to be sufficient for verifying the privacy paradox in this study. Further, looking at the interrelations in the path model, the case could be made that the positive effect of prior disclosure behaviour on willingness to share, is indirectly diminished via the positive effect of prior disclosure behaviour on privacy concerns, which ultimately reduces the willingness to share via the path, perceived risks. Nevertheless, the exact effect is hard to quantify and due to the insignificance of perceived risk on willingness to share anyway questionable. Summarizing, based on the study, the privacy paradox is evident in the context of ChatGPT and could be explained by the still limited understanding of ChatGPT's privacy risks and data collection practices due to the complexity and novelty of LLM in general (Kokolakis, 2017, p. 129).

Further, the effects of trust on privacy concerns, perceived risks, and willingness to share personal data in ChatGPT were examined.

The negative effect of trust on perceived risks found in the literature was also evident in the underlying study, however insignificant (Dinev and Hart, 2006, pp. 71–72; Jarvenpaa et al., 2000, p. 60). Likewise trust had an insignificant, positive effect on the willingness to share personal data in ChatGPT. Previous studies found trust to reduce perceived risks, ultimately leading to a higher willingness to transact e. g. on online websites (Dinev and Hart, 2006, p. 66; Jarvenpaa et al., 2000, pp. 60–64; Pavlou and Gefen, 2004, p. 50). The fact that both findings are insignificant in the underlying study, while pointing in the same direction, seems at least consistent (Dinev and Hart, 2006, p. 66; Jarvenpaa et al., 2000, pp. 60–64; Pavlou and Gefen, 2004, p. 50). On the other hand, a strong negative relationship between trust and privacy concerns was confirmed in this study, in accordance with the consensus in literature (Pavlou, 2011, p. 981). Therefore, the presented results shed more light on the debated directionality of the relationship by establishing a strong negative effect of trust on privacy concerns (Pavlou, 2011, p. 981).

Perceived benefits and perceived risks as the two driving forces in the dual-path model, were intended to explain individuals' disclosure decisions in ChatGPT (Wang et al., 2016). The results should have supported optimally the balanced framework of the privacy calculus, introduced earlier. In previous studies, perceived benefits were mostly measured to be having a positive effect on disclosure decisions while privacy-related costs, or perceived risks, had mostly a negative effect (Keith et al., 2013, p.

1169; Wang et al., 2016, sec. 4; Xu et al., 2011, p. 47). Consistent with these prior studies, perceived risks also revealed a negative relationship on the willingness to share personal data in ChatGPT, nevertheless the effect was found to be insignificant. Based on the indicators of the construct perceived risks, it can therefore be said that providing personal information in ChatGPT is, however insignificantly, associated with unexpected problems, risks, potential loss, and uncertainty. This can ultimately lead to an insignificantly lower willingness to share personal data in ChatGPT. Having confirmed the reliability and validity of the model in previous parts, it is therefore assumed that indeed the very different context of ChatGPT, compared to the previous studies, caused this result and not measurement errors. Potentially, due to the novelty of the technology, individuals cannot yet evaluate risks and their effect on disclosure decisions coherently, as it is the case for studies examining familiar contexts such as e-commerce transactions or mobile apps (Keith et al., 2013; Li et al., 2010; Wang et al., 2016; Xu et al., 2011).

Even more surprising is the result for the second driving force in the dual-path model, perceived benefits. While it was expected to make a simple case by hypothesizing that perceived benefits would have a positive effect on the willingness to share personal data in ChatGPT, the opposite was true. Previous studies consistently confirmed the strong positive effect of this construct in privacy calculus models (Keith et al., 2013, p. 1169; Li et al., 2010, p. 10; Wang et al., 2016, sec. 4; Xu et al., 2011, p. 47). Nevertheless, in this study, the effect of perceived benefits on willingness to share personal data was found to be negative, however, insignificant. Interpreting this effect would indicate, that individuals who perceive the benefits of using ChatGPT very strong are on the other hand less willing to share their personal information. This would in turn basically constrain them from further usage and is therefore to be considered irrational. Likewise, the result is deviating from the consensus in literature and needs an explanation. As the analysed literature contained no similar case, the following explanations can only be considered a speculation. Again, due to the novelty of the technology, participants might have difficulties in coherently assessing all facets of ChatGPT. They could perceive the capabilities of ChatGPT too good to be true, or overly advantageous, ultimately raising suspicions and thereby reducing the willingness to share personal data.

Alternatively, framing effects, established in literature, can lead with even slight changes in the wording of scenarios to highly differing outcomes, especially in contexts of risks and benefits (Kahneman and Tversky, 1979; Malenka et al., 1993). Thus, the overly positive framing of the perceived benefits items could have led to raising suspicions and making the participants more aware of drawbacks associated with ChatGPT's use. Looking at the correlations-matrix, participants who are perceiving ChatGPT's benefits as strong were at the same time heavier users. Possibly, the framing of the willingness to share construct could have led them to thinking that sharing information is fully optional and does not relate to potential constraints in ChatGPT-usage. If the same construct would have been framed more towards a restricted access or even exclusion from further use of ChatGPT in case of unwillingness to share personal data, the measured effect could have been different.

Evaluating the post-hoc analysis, several insights and conclusions can be drawn. First, the direct relationship between privacy concerns and the willingness to share personal data in ChatGPT is investigated. As discussed in the previous section, perceived benefits' negative effect on willingness to share personal data in ChatGPT has already been unexpected. This post-hoc analysis shows the same contradictory direction for privacy concerns on willingness to share, if deviating from the dual path model's positive path. The positive effect of 0.160 is insignificant would however indicate, that individuals with higher privacy concerns are also more willing to share personal data in ChatGPT, which could be another indication for the privacy paradox (Kokolakis, 2017, p. 128). As mentioned, the unexpected effect of perceived benefits in the main model, motivated one more investigation of the construct. Thereby the effect of perceived benefits on privacy concerns is measured and found to be positive, though insignificant.

Besides this, especially the control variable age caught further attention, as this control variable should be expected to have a major impact on several constructs other than willingness to share (see for an extensive overview Marwick et al., 2010). Thus, three additional relationships were tested. Namely, the effects of first, age on trust, second, age on privacy concerns and last, age on perceived risks. The results show negative effects for all three tested relations, which could be caused by the conceptually similar constructs, as mentioned earlier. Even though the effects are found to be insignificant, they would indicate that younger individuals are being more concerned about their privacy, perceive higher levels of risks and have less trust in ChatGPT. This could be due

to a more sophisticated digital literacy and thus more pronounced ability to evaluate digital dangers (see for an extensive overview Marwick et al., 2010).

6.1 Theoretical Contributions

Thepaper offers various theoretical contributions. The first part of thepaper answered research question one by creating a comprehensive overview of the determinants of data sharing. A striking observation made was that the literature on data sharing decisions is rarely describing or defining determinants, constructs, or items neither in detail nor consistently. This ambiguity made it hard to develop a definite framework. However, even despite these hurdles, the above presented framework tried to capture the most important determinants in the according field and aimed at delivering a generalizable overview that adds to the body of literature. This framework helps to close the gap in literature regarding a missing comprehensive framework on determinants of data sharing decisions. It enables future research to easily choose relevant determinants from established literature and conduct further studies investigating importance or interrelations in other contexts and research approaches.

The privacy calculus dual-path model, introduced in the second research question, adds to the wide range of studies applying similar theoretical approaches (Keith et al., 2013; Li et al., 2010; Wang et al., 2016; Xu et al., 2011). Applying this established approach to the relatively novel field of LLM and ChatGPT, contributes to the understanding of the privacy calculus framework and its applicability in diverse contexts. The two fundamental paths, privacy costs and benefits, were found to be neither significant nor consistent with literature in the underlying study. Thus, the outcome should stimulate a debate on potential limitations of this approach or specifics regarding novel technologies.

The study further contributes to the debate on the directionality of privacy concerns and trust, revealing a strong and significant negative effect of trust on privacy concerns (Pavlou, 2011, p. 981).

Likewise, ChatGPT is added to the extensive body of literature regarding the privacy paradox. As mentioned earlier, several previous studies thereby examined disclosure intentions and actual disclosure behaviour of individuals (see for an overview Kokolakis, 2017). This study however applied a rather unusual approach by implementing the

construct prior disclosure behaviour due to the limited scope and timeframe of thepaper. Thus, this approach sheds light on an alternative way of investigating and measuring the privacy paradox and validates, based on the introduced assumptions, the privacy paradox in ChatGPT.

6.2 Implications for Practitioners

Several insights from the study offer implications for practitioners. These implications are specifically suited for businesses working with LLM or ChatGPT. Based on the results of the first research question, practitioners of different technologies in diverse contexts can now choose from a set of determinants, provided in the comprehensive overview. As the overview contains studies from a broad range of technologies and contexts the given determinants help managers and practitioners to understand and take into account different determinants of data sharing when making decisions. While determinants such as privacy concerns or control likely were considered before, other important determinants like severity or purpose might have not been in focus. Thus, managers can adapt their data handling practices and the way they are communicated with consumers. For instance, providing a senseful purpose and communicating transparently why the business collects personal data can ultimately help to increase individuals' willingness to share personal data and improve business operations (Ackermann et al., 2022, p. 378; Anderson and Agarwal, 2011, p. 473; Bronk et al., 2018, pp. 1–2; Keith et al., 2013, p. 1172; Wang et al., 2016, sec. 1).

The study, to answer the second research question, also brings several implications especially for LLM providers, ChatGPT, or businesses who run services based on such models.

Regarding the positive effect of privacy concerns on perceived risks, managers and practitioners should focus on educating consumers on data handling practices in ChatGPT, such as improved and strengthened data regulations introduced in the beginning of thepaper. Communicating control, data security, and transparency, ultimately reduce privacy concerns and helps to mitigate the perceived risks of data sharing in these environments.

Concerning evidence on the privacy paradox in ChatGPT, associated businesses can help to reduce this irrational consumer behaviour. This could be done by implementing a feedback mechanism within LLM by helping users who shared personal data with

the model, to understand the potential implications and risks of these actions. For instance, users who indicate high privacy concerns in a feedback session but on the other hand showed a high rate of personal data sharing should be made aware of their behaviour. This could ultimately help to reduce privacy concerns, improve transparency and trust into the business to support individuals' sound decision making. Alternatively, the privacy paradox could be leveraged by introducing incentives or rewards, such as discounts, to users who show high privacy concerns but still actively engage with ChatGPT. This approach could help to acknowledge and value cautious yet active consumers and create a positive cycle that supports data sharing.

The negative effect of trust on privacy concerns indicates that businesses should in general invest in trust building strategies, such as showing transparency, ethical data handling practices and trust, by displaying certifications or user testimonials (Chellappa and Sin, 2005, pp. 196–198). Further, specifically in ChatGPT, comprehensible information on how the model is trained and what security protocols are implemented should be considered to build trust. These measures are partly already in place, however there is still confusion judging by the number of debates in governments and among data protection officers, discussed earlier in thepaper. One solution, based on the results of the literature review as well as the study, would be to implement granular privacy controls. These privacy controls could for instance, allow users of ChatGPT to customize data sharing preferences, such as specifying types of data they are willing to share, certain types of information they refuse to share or purposes of data sharing they support.

Finally, these approaches should empower users, reduce their privacy concerns, and enhance their trust in the business. Thereby, ChatGPT and associated businesses could not only address privacy concerns but also encourage responsible data sharing and improve the relationship between users, ChatGPT and businesses adopting LLM-associated services.

6.3 Limitations

Although the study and literature review were conducted under greatest care, several limitations must be noted. Thus, in the following, limitations of the first part, referring to research question one and subsequently the limitations of the second part, referring to research question two, are discussed.

The systematic literature review focused on six major domains and seven technologies. This coverage brings a limitation in the external validity of the framework, as it is possible that there are deviations in uncovered technologies or domains. However, based on the assessed studies these deviations should be minor, as the presented main determinants are covering a large share of described constructs and determinants as well as a notable share of the available, relevant literature in the field. Additionally, to increase the replicability and transparency of the review methods used, the procedure, as depicted in the Prisma diagram, and the standards for search and inclusion of studies were explained. While the possibility of a selection bias of studies cannot be ruled out with certainty, the likelihood of such was minimized by using multiple databases, different domains, and search methods, like reassuring comprehensiveness through differing search queries. Furthermore, the synthesis tried to enrich the qualitative results with quantitative elements, by manually counting the frequencies of determinants across studies to enhance the understanding of each determinant's relevance. The additionally 32 mentioned determinants further showed that the descriptive granularity of the framework can be increased or decreased, indicating a certain flexibility of the suggested framework and nomenclature.

In the second part an online survey was conducted, and a partial least squares structural equation model implemented, introducing several limitations. By applying convenience and self-selection sampling, typically the respondents may not be representative of the entire population, diminishing the external validity of the study (Henry, 1990, pp. 4–7; King and He, 2005, pp. 889–891).

Likewise, surveys can introduce a self-reporting bias, such as a social desirability bias (Steenkamp et al., 2010, pp. 1–2). For instance, reporting low levels of prior disclosure behaviour regarding friends' or colleagues' personal information or indicating unwillingness to share personal data, solely because it is socially agreed upon (Brenner and DeLamater, 2016, pp. 346–350; Steenkamp et al., 2010). This in turn could distort the measured effects. Depending on the direction of the social desirability bias, for instance, a generally higher willingness to share personal information, or less conservative prior disclosure behaviour might be measured.

As the survey was conducted entirely in English, this automatically excludes individuals who are not proficient in English. Additionally, it may decrease the quality of the data if individuals who are less proficient in English, misunderstood questions and therefore

submit inaccurate answers. Thereby, external validity could be compromised by limiting the results' generalizability at least to the English-speaking population.

As mentioned before, reversely coded items were partly implemented to avoid agreement bias, rushing respondents and to improve scale validity (Weijters and Baumgartner, 2012, p. 1). Besides those advantages however, poor reliability and model fit, and how the measurement model showed, often small factor loadings are induced (Weijters and Baumgartner, 2012, p. 1). Consequently, several of the reverse coded items were removed in the process. It would be interesting to analyse whether the results significantly change if reversed items would be fully replaced by non-reversed items.

Further, general limitations introduced using the PLS-SEM approach, such as non-established goodness-of-fit measures could be pointed out (Hair Jr et al., 2021, pp. 13–14). However, based on the arguments pointed out in the methods section, the chosen approach was considered the best-fit regarding research objectives and available data.

6.4 Future Research

The study gives several implications and conclusion for future research. One striking observation made is that the literature on data sharing decision is rarely describing or defining determinants, constructs or items in detail (see for example Becker et al., 2021). To reveal further important determinants that have not been investigated yet, qualitative research approaches such as interviews or case studies could be conducted. It would be of importance to conduct these studies either in so far non-investigated domains, technologies, or try to fully detach the study from any context or technology to generate general and encompassing outcomes.

Based on the weak results regarding trust on willingness to share personal data in ChatGPT in this study, effects and validity of trust building factors by companies, already studied in other contexts, could be also investigated further in the context of LLM in future (Chellappa and Sin, 2005, pp. 196–198).

One of the major limitations of this study is the convenience sample and its potential implications on external validity. Reassessing the study based on a bigger and fully

randomized sample could thus help to establish external validity and try to confirm the findings.

Likewise, future studies could try to validate the empirical findings by investigating effects of different cultures, geographical settings or different beliefs on disclosure decisions (see for example Dinev et al., 2006). The novelty of LLM is thereby stressing the necessity for such further studies. The unexpected, measured effects of benefits on the willingness to share personal data in ChatGPT in this study could be either resulting from a problem in the data or potentially being unique to this new technology. One suggestion to test this, is the reassessment and redefinition of the benefits, as well as the privacy costs, here privacy risks, in the dual-path model, since both driving forces were not confirmed.

Further, the privacy paradox in this study was tested rather unusually. Future research could therefore implement a more established approach by conducting randomized experiments examining sharing intentions and actual decisions in the context of LLM. Besides that, the post-hoc analysis introduced several other aspects that could be investigated further, such as the vague relationships between the control variables and constructs, as well as measuring direct effects between the constructs detached from privacy calculus.

7 Conclusion

This paper systematically investigated the determinants of individuals' data sharing decisions, their relevance, and interrelations. To answer research question one, a systematic literature review that examined 53 studies across fields and technologies was conducted, providing a comprehensive overview of the determinants of data sharing decisions. Despite the difficulties of finding consistent definitions, the framework successfully captures the most important determinants. Acknowledging potential limitations of external validity due to the specific fields and technologies investigated, selection bias was minimized by collecting a broad sample. Qualitative insights, supported by quantitative information gave a deeper understanding of the determinants' importance. Thus, the results regarding the first research question advance the understanding of determinants of data sharing, close the literature gap constituted by the missing comprehensive framework, and give implications for theorists and practitioners.

Building up on these results, research question two examined the relationships and relevance of several determinants of data sharing decisions, focusing on ChatGPT. However, sampling methods were based on convenience and selection, driven mainly by financial and time constraints. Although not random, the high accessibility and diffusion of the study reduced the extent of this limitation.

A partial least squares structural equation model in smartPLS (2022) examined the complex interactions of privacy concerns, perceived risks, perceived benefits, trust, prior disclosure behaviour and the willingness to share personal data in ChatGPT. The results revealed trust being identified as a significant determinant negatively affecting privacy concerns, consistent with the established literature, while its effects on the willingness to share personal data, and perceived risks, were not statistically significant. While perceived risks showed an expected negative relation with willingness to share personal data in ChatGPT, perceived benefits revealed an unexpected negative effect. However, both were found to be insignificant. It was argued that these unexpected results could be attributed to the unique and novel environment of ChatGPT, affecting individuals' perceptions in decision-making processes. Further, framing effects introduced by the survey design may have contributed to these results, highlighting the sensitivity of the wording used in such studies. Next to this, the privacy paradox was investigated and found to be evident in the context of ChatGPT, based on the assumptions made in thepaper.

Finally, this study contributes particularly to our understanding of the determinants of data sharing and disclosure decisions in emerging technologies. The comprehensive framework, insights into the privacy paradox, and what these imply for businesses using ChatGPT, LLM or similar technologies, provide a valuable contribution to the field. Acknowledging the limitations mentioned, it is hoped that thispaper helps practitioners and theorists and opens the door for future research.

References

Ackermann, K.A., Burkhalter, L., Mildenberger, T., Frey, M., and Bearth, A. 2022. "Willingness to Share Data: Contextual Determinants of Consumers' Decisions to Share Private Data with Companies," *Journal of Consumer Behaviour* (21:2), pp. 375-386.

Acquisti, A. 2010. "The Economics of Personal Data and the Economics of Privacy," *Carnegie Mellon University Research Showcase*).

Adams, J.S., and Freedman, S. 1976. "Equity Theory Revisited: Comments and Annotated Bibliography," *Advances in experimental social psychology* (9), pp. 43-90.

Agahari, W., and de Reuver, M. 2022. "Rethinking Consumers' Data Sharing Decisions with the Emergence of Multi-Party Computation: An Experimental Design for Evaluation," *ECIS*.

AIS. 2023. "Senior Scholars' List of Premier Journals." Retrieved 05/10, 2023, from https://aisnet.org/page/SeniorScholarListofPremierJournals.

Ajzen, I. 1991. "The Theory of Planned Behavior," *Organizational behavior and human decision processes* (50:2), pp. 179-211.

Alemany, J., Del Val, E., and García-Fornes, A. 2020. "Empowering Users Regarding the Sensitivity of Their Data in Social Networks through Nudge Mechanisms," *Proceedings of the 53rd Hawaii International Conference on System Sciences*, Hawaii: CORE, pp. 2539-2548.

Anderson, C.L., and Agarwal, R. 2009. "Genetic Information Altruists: How Far and to Whom Does Their Generosity Extend?," *ICIS 2009 Proceedings*, Phoenix.

Anderson, C.L., and Agarwal, R. 2011. "The Digitization of Healthcare: Boundary Risks, Emotion, and Consumer Willingness to Disclose Personal Health Information," *Information Systems Research* (22:3), pp. 469-490.

Andrade, E.B., Kaltcheva, V., and Weitz, B. 2002. "Self-Disclosure on the Web: The Impact of Privacy Policy, Reward, and Company Reputation," *ACR North American Advances Consumer Research Volume* (29), pp. 350-353.

Anshari, M., Almunawar, M.N., Lim, S.A., and Al-Mudimigh, A. 2019. "Customer Relationship Management and Big Data Enabled: Personalization & Customization of Services," *Applied Computing and Informatics* (15), pp. 94-101.

Araci, D. 2019. "Finbert: Financial Sentiment Analysis with Pre-Trained Language Models," *arXiv preprint arXiv:1908.10063*).

Armstrong, J.S. 1975. "Monetary Incentives in Mail Surveys," *The Public Opinion Quarterly* (39), pp. 111-116.

Awad, N.F., and Krishnan, M.S. 2006. "The Personalization Privacy Paradox: An Empirical Evaluation of Information Transparency and the Willingness to Be Profiled Online for Personalization," *MIS Quarterly* (30:1), pp. 13-28.

Ayyagari, R., Grover, V., and Purvis, R. 2011. "Technostress: Technological Antecedents and Implications," *MIS Quarterly* (35:4), pp. 831-858.

Banerjee, S., Hemphill, T., and Longstreet, P. 2018. "Wearable Devices and Healthcare: Data Sharing and Privacy," *The Information Society* (34:1), pp. 1-9.

Bansal, G., and Nah, F. 2020. "Measuring Privacy Concerns with Government Surveillance and Right-to-Be-Forgotten in Nomological Net of Trust and Willingness-to-Share," *AMCIS 2020 Proceedings*, Virtual Event.

Bäumler, H., and Mutius, A. 2013. *Datenschutz Als Wettbewerbsvorteil: Privacy Sells: Mit Modernen Datenschutzkomponenten Erfolg Beim Kunden*. Springer-Verlag.

Becker, I., Posner, R., Islam, T., Ekblom, P., Borrion, H., McGuire, M., and Li, S. 2021. "Privacy in Transport? Exploring Perceptions of Location Privacy through User Segmentation," *Proceedings of 54th Hawaii International Conference on System Sciences*, Hawaii.

Bélanger, F., Resor, J., Crossler, R.E., Finch, T.A., and Allen, K.R. 2021. "Smart Home Speakers and Family Information Disclosure Decisions," *Twenty-Seventh Americas Conference on Information Systems*, Montreal.

Benndorf, V., and Normann, H.T. 2014. "The Willingness to Sell Personal Data," *DICE Discussion Paper*:143), pp. 1260-1278.

Bin Sulaiman, R., Schetinin, V., and Sant, P. 2022. "Review of Machine Learning Approach on Credit Card Fraud Detection," *Human-Centric Intelligent Systems* (2:1-2), pp. 55-68.

Both, L.E. 2021. "Willingness to Share Personal Information," *Psychological Applications and Trends*).

Braun, D., and Guston, D.H. 2003. "Principal-Agent Theory and Research Policy: An Introduction," *Science and public policy* (30:5), pp. 302-308.

Brenner, P.S., and DeLamater, J. 2016. "Lies, Damned Lies, and Survey Self-Reports? Identity as a Cause of Measurement Bias," *Social psychology quarterly* (79:4), pp. 333-354.

Bronk, K.C., Riches, B.R., and Mangan, S.A. 2018. "Claremont Purpose Scale: A Measure That Assesses the Three Dimensions of Purpose among Adolescents," *Research in Human Development* (15:2), pp. 101-117.

Brown, B., Chui, M., and Manyika, J. 2011. "Are You Ready for the Era of 'Big Data'," *McKinsey Quarterly* (4:1), pp. 24-35.

Brown, H., Lee, K., Mireshghallah, F., Shokri, R., and Tramèr, F. 2022. "What Does It Mean for a Language Model to Preserve Privacy?," *2022 ACM Conference on Fairness, Accountability, and Transparency*, Seoul, pp. 2280-2292.

Brown, S.M., Bell, S.K., Roche, S.D., Dente, E., Mueller, A., Kim, T.-E., O'Reilly, K., Lee, B.S., Sands, K., and Talmor, D. 2016. "Preferences of Current and Potential Patients and Family Members Regarding Implementation of Electronic Communication Portals in Intensive Care Units," *Annals of the American Thoracic Society* (13:3), pp. 391-400.

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J.D., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., and Askell, A. 2020. "Language Models Are Few-Shot Learners," in: *34th Conference on Neural Information Processing Systems* Vancouver: pp. 1877-1901.

Caine, K., and Hanania, R. 2013. "Patients Want Granular Privacy Control over Health Information in Electronic Medical Records," *Journal of the American Medical Informatics Association* (20:1), pp. 7-15.

Canhoto, A.I., and Arp, S. 2017. "Exploring the Factors That Support Adoption and Sustained Use of Health and Fitness Wearables," *Journal of Marketing Management* (33:1-2), pp. 32-60.

Cannell, C., and Henson, R. 1974. "Incentives, Motives, and Response Bias," in *Annals of Economic and Social Measurement,* S.V. Berg (ed.). NBER, pp. 307-317.

Carlini, N., Tramer, F., Wallace, E., Jagielski, M., Herbert-Voss, A., Lee, K., Roberts, A., Brown, T., Song, D., and Erlingsson, U. 2021. "Extracting Training Data from Large Language Models," *30th USENIX Security Symposium* Virtual Event, pp. 2633-2650.

Carrascal, J.P., Riederer, C., Erramilli, V., Cherubini, M., and De Oliveira, R. 2013. "Your Browsing Behavior for a Big Mac: Economics of Personal Information Online," *Proceedings of the 22nd international conference on World Wide Web*, Rio de Janeiro, pp. 189-200.

Cascella, M., Montomoli, J., Bellini, V., and Bignami, E. 2023. "Evaluating the Feasibility of Chatgpt in Healthcare: An Analysis of Multiple Clinical and Research Scenarios," *Journal of Medical Systems* (47:1), p. 33.

Cate, F.H., and Mayer-Schönberger, V. 2013. "Notice and Consent in a World of Big Data," *International Data Privacy Law* (3:2), pp. 67-73.

Chellappa, R.K., and Sin, R.G. 2005. "Personalization Versus Privacy: An Empirical Examination of the Online Consumer's Dilemma," *Information Technology and Management* (6), pp. 181-202.

Cichy, P., Salge, T.O., and Kohli, R. 2021. "Privacy Concerns and Data Sharing in the Internet of Things: Mixed Methods Evidence from Connected Cars," *MIS Quarterly* (45:4).

Citron, D.K. 2006. "Reservoirs of Danger: The Evolution of Public and Private Law at the Dawn of the Information Age," *Southern California Law Review* (80), p. 241.

Cocosila, M., and Archer, N. 2014. "Perceptions of Chronically Ill and Healthy Consumers About Electronic Personal Health Records: A Comparative Empirical Investigation," *BMJ open* (4:7), p. e005304.

Coiera, E.W., Verspoor, K., and Hansen, D.P. 2023. "We Need to Chat About Artificial Intelligence," *Medical Journal of Australia*:219 (3), pp. 98-100.

Connelly, L.M. 2016. "Cross-Sectional Survey Research," *Medsurg nursing* (25:5), p. 369.

Cropanzano, R., and Mitchell, M.S. 2005. "Social Exchange Theory: An Interdisciplinary Review," *Journal of management* (31:6), pp. 874-900.

Daigle, B., and Khan, M. 2020. "The Eu General Data Protection Regulation: An Analysis of Enforcement Trends by Eu Data Protection Authorities," *Journal of International Commerce and Economics*), pp. 1-38.

De Nijs, R. 2017. "Behavior-Based Price Discrimination and Customer Information Sharing," *International Journal of Industrial Organization* (50), pp. 319-334.

De Schaepdrijver, L., Baecke, P., and Tackx, K. 2022. "What Makes Consumers Willing to Share Their Data in Addressable Tv Advertising?: The Influence of Personal and Situational Factors on Consumer Willingness to Disclose Information," *Journal of Advertising Research* (62:2), pp. 131-147.

DeCarlo, L.T. 1997. "On the Meaning and Use of Kurtosis," *Psychological methods* (2:3), pp. 292-307.

Deuker, A., Rosenkranz, C., and Albers, A. 2012. "The Usage of Individual Privacy Settings on Social Networking Sites-Drawing Desired Digital Images of Oneself," *ECIS 2012 Proceedings*, Barcelona.

Dilmegani, C. 2023. "The Future of Large Language Models." Retrieved 05/09, 2023, from https://research.aimultiple.com/future-of-large-language-models/.

Dinev, T., Bellotto, M., Hart, P., Russo, V., Serra, I., and Colautti, C. 2006. "Privacy Calculus Model in E-Commerce–a Study of Italy and the United States," *European Journal of Information Systems* (15:4), pp. 389-402.

Dinev, T., and Hart, P. 2006. "An Extended Privacy Calculus Model for E-Commerce Transactions," *Information systems research* (17:1), pp. 61-80.

Donovan-Kicken, E., Mackert, M., Guinn, T.D., Tollison, A.C., Breckinridge, B., and Pont, S.J. 2012. "Health Literacy, Self-Efficacy, and Patients' Assessment of Medical Disclosure and Consent Documentation," *Health Communication* (27:6), pp. 581-590.

Durneva, P. 2020. "To Share or Not to Share: Optimal Value of Insurance Rewards for Sharing Data Generated from Wearable Devices for Hypertensive Patients," *AMCIS 2020 TREOs*, 14.

Eberendu, A.C. 2016. "Unstructured Data: An Overview of the Data of Big Data," *International Journal of Computer Trends and Technology* (38:1), pp. 46-50.

Egelman, S., Felt, A.P., and Wagner, D. 2013. "Choice Architecture and Smartphone Privacy: There'sa Price for That," in *The Economics of Information Security and Privacy*. Berlin: Springer, pp. 211-236.

Eggmann, F., Weiger, R., Zitzmann, N.U., and Blatz, M.B. 2023. "Implications of Large Language Models Such as Chatgpt for Dental Medicine," *Journal of Esthetic and Restorative Dentistry*).

Enders, T., Wolff, C., and Satzger, G. 2020. "Knowing What to Share: Selective Revealing in Open Data," *Proceedings of 28th European Conference on Information Systems*, Marrakesh, Morocco.

Esmaeilzadeh, P. 2019. "The Process of Building Patient Trust in Health Information Exchange (Hie): The Impacts of Perceived Benefits, Perceived Transparency of

Privacy Policy, and Familiarity," *Communications of the Association for Information Systems* (45), pp. 364-396.

Esmaeilzadeh, P. 2020. "The Effect of the Privacy Policy of Health Information Exchange (Hie) on Patients' Information Disclosure Intention," *Computers & Security* (95), p. 101819.

Fama, E.F. 1970. "Efficient Capital Markets: A Review of Theory and Empirical Work," *The journal of Finance* (25:2), pp. 383-417.

Farrelly, R., and Chew, E. 2016. "Who's in to Win?: Participation Rate in a Primary Personal Information Market," in: *Australasian Conference on Information Systems*. Wollongong.

Featherman, M.S., and Pavlou, P.A. 2003. "Predicting E-Services Adoption: A Perceived Risk Facets Perspective," *International journal of human-computer studies* (59:4), pp. 451-474.

Fischer, M. 1993. *Die Principal-Agent-Theorie*. Springer-Verlag.

Fortes, N., Rita, P., and Pagani, M. 2017. "The Effects of Privacy Concerns, Perceived Risk and Trust on Online Purchasing Behaviour," *International Journal of Internet Marketing and Advertising* (11:4), pp. 307-329.

Frey, R., Wörner, D., and Ilic, A. 2016. "Collaborative Filtering on the Blockchain: A Secure Recommender System for E-Commerce," in: *Twenty-second Americas Conference on Information Systems*. San Diego: p. 5.

Friedman, M., and Savage, L.J. 1952. "The Expected-Utility Hypothesis and the Measurability of Utility," *Journal of Political Economy* (60:6), pp. 463-474.

Gaylin, D.S., Moiduddin, A., Mohamoud, S., Lundeen, K., and Kelly, J.A. 2011. "Public Attitudes About Health Information Technology, and Its Relationship to Health Care Quality, Costs, and Privacy," *Health services research* (46:3), pp. 920-938.

Goddard, M. 2017. "The Eu General Data Protection Regulation (Gdpr): European Regulation That Has a Global Impact," *International Journal of Market Research* (59:6), pp. 703-705.

Greshake, K., Abdelnabi, S., Mishra, S., Endres, C., Holz, T., and Fritz, M. 2023. "More Than You've Asked For: A Comprehensive Analysis of Novel Prompt Injection Threats to Application-Integrated Large Language Models," *arXiv preprint arXiv:2302.12173*).

Gumbus, A., and Grodzinsky, F. 2015. "Era of Big Data: Danger of Descrimination," *ACM SIGCAS Computers and Society* (45:3), pp. 118-125.

Hair Jr, J.F., Hult, G.T.M., Ringle, C.M., Sarstedt, M., Danks, N.P., and Ray, S. 2021. *Partial Least Squares Structural Equation Modeling (Pls-Sem) Using R: A Workbook*, (1 ed.). Switzerland: Springer Cham.

Hann, I.-H., Hui, K.-L., Lee, S.-Y.T., and Png, I.P. 2007. "Overcoming Online Information Privacy Concerns: An Information-Processing Theory Approach," *Journal of management information systems* (24:2), pp. 13-42.

Hasnain-Wynia, R., Taylor-Clark, K., and Anise, A. 2011. "Collecting Race, Ethnicity, and Language Data to Identify and Reduce Health Disparities: Perceptions of Health Plan Enrollees," *Medical Care Research and Review* (68:3), pp. 367-381.

Henry, G.T. 1990. *Practical Sampling*. Thousand Oaks: Sage.

Hoadley, C.M., Xu, H., Lee, J.J., and Rosson, M.B. 2010. "Privacy as Information Access and Illusory Control: The Case of the Facebook News Feed Privacy Outcry," *Electronic commerce research and applications* (9:1), pp. 50-60.

Inmon, W.H., and Nesavich, A. 2007. *Tapping into Unstructured Data: Integrating Unstructured Data and Textual Analytics into Business Intelligence*. Pearson Education.

Instacart. 2023. "Privacy Policy." Retrieved 06/23, 2023, from https://www.instacart.com/privacy.

Ioannou, A., Tussyadiah, I., and Lu, Y. 2020. "Privacy Concerns and Disclosure of Biometric and Behavioral Data for Travel," *International Journal of Information Management* (54), p. 102122.

Jahn, S. 2007. "Strukturgleichungsmodellierung Mit Lisrel, Amos Und Smartpls: Eine Einführung (an Introduction to Structural Equation Modeling with Lisrel, Amos and Smartpls)." Chemnitz: Der Dekan der

Fakultät für Wirtschaftswissenschaften an der Technischen Universität Chemnitz.

Janoschka, A., Wozniak, T., Dahinden, L., and Albisser, M. 2022. "How Consumers Progress to More Advanced Levels of Data-Based Products and Services: A Scenario-Based Approach," *Bled 2022 Proceedings*, Bled.

Jarvenpaa, S.L., Tractinsky, N., and Vitale, M. 2000. "Consumer Trust in an Internet Store," *Information Technology and Management* (1), pp. 45-71.

Kahneman, D., and Tversky, A. 1979. "Prospect Theory: An Analysis of Decision under Risk," *Econometrica* (47:2), pp. 263-291.

Karampela, M., Ouhbi, S., and Isomursu, M. 2019. "Connected Health User Willingness to Share Personal Health Data: Questionnaire Study," *Journal of Medical Internet Research* (21:11), p. e14537.

Kasneci, E., Seßler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., and Hüllermeier, E. 2023. "Chatgpt for Good? On Opportunities and Challenges of Large Language Models for Education," *Learning and Individual Differences* (103).

Keith, M.J., Thompson, S.C., Hale, J., Lowry, P.B., and Greer, C. 2013. "Information Disclosure on Mobile Devices: Re-Examining Privacy Calculus with Actual User Behavior," *International journal of human-computer studies* (71:12), pp. 1163-1173.

Kerath, S.M., Klein, G., Kern, M., Shapira, I., Witthuhn, J., Norohna, N., Kline, M., Baksh, F., Gregersen, P., and Taioli, E. 2013. "Beliefs and Attitudes Towards Participating in Genetic Research - a Population Based Cross-Sectional Study," *BMC Public Health* (13), p. 114.

Kim, D., Park, K., Park, Y., and Ahn, J.-H. 2019. "Willingness to Provide Personal Information: Perspective of Privacy Calculus in Iot Services," *Computers in Human Behavior* (92), pp. 273-281.

Kim, K.K., Sankar, P., Wilson, M.D., and Haynes, S.C. 2017. "Factors Affecting Willingness to Share Electronic Health Data among California Consumers," *BMC medical ethics* (18:25), pp. 1-10.

Kim, M.S., and Kim, S. 2018. "Factors Influencing Willingness to Provide Personal Information for Personalized Recommendations," *Computers in Human Behavior* (88), pp. 143-152.

King, W.R., and He, J. 2005. "External Validity in Is Survey Research," *Communications of the Association for Information Systems* (16:1), pp. 880-894.

Klarmann, M., and Feurer, S. 2018. "Control Variables in Marketing Research," *Marketing ZFP – Journal of Research and Management* (40:2), pp. 26-40.

Kokolakis, S. 2017. "Privacy Attitudes and Privacy Behaviour: A Review of Current Research on the Privacy Paradox Phenomenon," *Computers & security* (64), pp. 122-134.

Krishnamurthy, B., Naryshkin, K., and Wills, C. 2011. "Privacy Leakage Vs. Protection Measures: The Growing Disconnect," *Proceedings of the Web*, pp. 1-10.

Lavie, T., Sela, M., Oppenheim, I., Inbar, O., and Meyer, J. 2010. "User Attitudes Towards News Content Personalization," *International journal of human-computer studies* (68:8), pp. 483-495.

Leon, P.G., Ur, B., Wang, Y., Sleeper, M., Balebako, R., Shay, R., Bauer, L., Christodorescu, M., and Cranor, L.F. 2013. "What Matters to Users? Factors That Affect Users' Willingness to Share Information with Online Advertisers," *Proceedings of the Ninth Symposium on Usable Privacy and Security*, Newcastle, UK, pp. 1-12.

Leppäniemi, M., Karjaluoto, H., and Saarijärvi, H. 2017. "Customer Perceived Value, Satisfaction, and Loyalty: The Role of Willingness to Share Information," *The International Review of Retail, Distribution and Consumer Research* (27:2), pp. 164-188.

Li, H., Sarathy, R., and Xu, H. 2010. "Understanding Situational Online Information Disclosure as a Privacy Calculus," *Journal of Computer Information Systems* (51:1), pp. 62-71.

Li, Y. 2012. "Theories in Online Information Privacy Research: A Critical Review and an Integrated Framework," *Decision Support Systems* (54:1), pp. 471-481.

Lowry, P.B., Cao, J., and Everard, A. 2011. "Privacy Concerns Versus Desire for Interpersonal Awareness in Driving the Use of Self-Disclosure Technologies: The Case of Instant Messaging in Two Cultures," *Journal of management information systems* (27:4), pp. 163-200.

Madhok, A. 2002. "Reassessing the Fundamentals and Beyond: Ronald Coase, the Transaction Cost and Resource-Based Theories of the Firm and the Institutional Structure of Production," *Strategic management journal* (23:6), pp. 535-550.

Mäki, M., Kauttonen, J., and Alamäki, A. 2023. "How Students' Information Sensitivity, Privacy Trade-Offs, and Stages of Customer Journey Affect Consent to Utilize Personal Data," *Interdisciplinary Journal of Information, Knowledge, and Management* (18), pp. 127-147.

Malenka, D.J., Baron, J.A., Johansen, S., Wahrenberger, J.W., and Ross, J.M. 1993. "The Framing Effect of Relative and Absolute Risk," *Journal of General Internal Medicine* (8:10), pp. 543-548.

Malhotra, N.K., Kim, S.S., and Agarwal, J. 2004. "Internet Users' Information Privacy Concerns (Iuipc): The Construct, the Scale, and a Causal Model," *Information systems research* (15:4), pp. 336-355.

Martin, K. 2015. "Ethical Issues in the Big Data Industry," *MIS Quarterly Executive* (14), p. 2.

Marwick, A.E., Murgia-Diaz, D., and Palfrey, J.G. 2010. "Youth, Privacy and Reputation," *Public Law & Legal Theory Working Paper Series* (Harvard Public Law Working Paper No. 10-29).

Matz, S.C., and Netzer, O. 2017. "Using Big Data as a Window into Consumers' Psychology," *Current opinion in behavioral sciences* (18), pp. 7-12.

Meade, A.W., and Craig, S.B. 2012. "Identifying Careless Responses in Survey Data," *Psychological Methods* (17:3), pp. 437-455.

Medford-Davis, L.N., Chang, L., and Rhodes, K.V. 2017. "Health Information Exchange: What Do Patients Want?," *Health informatics journal* (23:4), pp. 268-278.

Mertens, P., Bodendorf, F., König, W., Picot, A., Schumann, M., and Hess, T. 2005. *Grundzüge Der Wirtschaftsinformatik*, (12 ed.). Berlin: Springer Gabler.

Minch, R.P. 2015. "Location Privacy in the Era of the Internet of Things and Big Data Analytics," *HICSS*, Hawaii: IEEE, pp. 1521-1530.

Mohamed, N., and Ahmad, I.H. 2012. "Information Privacy Concerns, Antecedents and Privacy Measure Use in Social Networking Sites: Evidence from Malaysia," *Computers in Human Behavior* (28:6), pp. 2366-2375.

Moreira, J., Carvalho, A., and Horvath, T. 2018. *A General Introduction to Data Analytics*. USA: John Wiley & Sons.

Morey, T., Forbath, T., and Schoop, A. 2015. "Customer Data: Designing for Transparency and Trust," *Harvard Business Review* (93:5), pp. 96-105.

Motiwalla, L.F., and Li, X.-B. 2016. "Unveiling Consumers' Privacy Paradox Behaviour in an Economic Exchange," *International Journal of Business Information Systems* (23:3), pp. 307-329.

Moura, J., and Serrão, C. 2015. "Security and Privacy Issues of Big Data," in *Handbook of Research on Trends and Future Directions in Big Data and Web Intelligence*. n.p.: IGI Global, pp. 20-52.

Newman, I. 2000. "A Conceptualization of Mixed Methods: A Need for Inductive/Deductive Approach to Conducting Research," in: *Annual Meeting of the American Educational Research Asociation*. New Orleans: ERiC, pp. 2-13.

Norman, P., Boer, H., Seydel, E.R., and Mullan, B. 2015. "Protection Motivation Theory," in *Predicting Health Behaviour*. Open University Press McGraw Hill Education, pp. 70-106.

OpenAi. 2023a. "Chatgpt Plugins." Retrieved 27/05, 2023, from https://openai.com/blog/chatgpt-plugins.

OpenAi. 2023b. "Gpt-4 System Card." Retrieved 27/05, 2023, from https://cdn.openai.com/papers/gpt-4-system-card.pdf.

OpenAI. 2023c. "Privacy Policies." Retrieved 27/05, 2023, from https://openai.com/policies/privacy-policy.

Patel, V.N., Dhopeshwarkar, R.V., Edwards, A., Barron, Y., Likourezos, A., Burd, L., Olshansky, D., and Kaushal, R. 2011. "Low-Income, Ethnically Diverse Consumers' Perspective on Health Information Exchange and Personal Health Records," *Informatics for Health and Social Care* (36:4), pp. 233-252.

Pavlou, P.A. 2011. "State of the Information Privacy Literature: Where Are We Now and Where Should We Go?," *MIS Quarterly* (4:35), pp. 977-988.

Pavlou, P.A., and Gefen, D. 2004. "Building Effective Online Marketplaces with Institution-Based Trust," *Information Systems Research* (15:1), pp. 37-59.

Petridis, S., Diakopoulos, N., Crowston, K., Hansen, M., Henderson, K., Jastrzebski, S., Nickerson, J.V., and Chilton, L.B. 2023. "Anglekindling: Supporting Journalistic Angle Ideation with Large Language Models," *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, Hamburg, pp. 1-16.

Petronio, S. 2010. "Communication Privacy Management Theory: What Do We Know About Family Privacy Regulation?," *Journal of family theory & review* (2:3), pp. 175-196.

Phelps, J., Nowak, G., and Ferrell, E. 2000. "Privacy Concerns and Consumer Willingness to Provide Personal Information," *Journal of public policy & marketing* (19:1), pp. 27-41.

Piccoli, G., and Watson, R.T. 2008. "Profit from Customer Data by Identifying Strategic Opportunities and Adopting the" Born Digital" Approach," *MIS Quarterly Executive* (7:3).

Pickard, K.T., and Swan, M. 2014. "Big Desire to Share Big Health Data: A Shift in Consumer Attitudes toward Personal Health Information," *Big Data Becomes Personal: Knowledge into Meaning: Papers from the AAAI Spring Symposium*, California.

PollPool. 2023. "Pollpool - Find Participants Now." Retrieved 07/17, 2023, from https://www.poll-pool.com/?locale=en.

Pongratz, E. 2023. "Daten- Und Jugendschutz: Italien Sperrt Chatgpt." Retrieved 03/31, 2023, from https://www.tagesschau.de/ausland/europa/italien-chatgpt-ki-101.html.

Prisma. 2023. "Transparent Reporting of Systematic Reviews and Meta-Analyses." Retrieved 05/09, 2023, from http://www.prisma-statement.org.

Punj, G.N. 2019. "Understanding Individuals' Intentions to Limit Online Personal Information Disclosures to Protect Their Privacy: Implications for Organizations and Public Policy," *Information Technology and Management* (20), pp. 139-151.

Redmond, M. 2015. "Social Exchange Theory." Iowa State University.

Romanov, E., and Copplestone, P. 2023. "Building Chatgpt Plugins." Retrieved 07/10, 2023, from https://supabase.com/blog/chatgpt-plugins-support-postgres.

Sagiroglu, S., and Sinanc, D. 2013. "Big Data: A Review," *2013 International Conference on Collaboration Technologies and Systems*, San Diego: IEEE, pp. 42-47.

Sallam, M. 2023. "The Utility of Chatgpt as an Example of Large Language Models in Healthcare Education, Research and Practice: Systematic Review on the Future Perspectives and Potential Limitations," *medRxiv*), pp. 2-34.

Sarathy, R., and Muralidhar, K. 2006. "Secure and Useful Data Sharing," *Decision Support Systems* (42:1), pp. 204-220.

Schudy, S., and Utikal, V. 2017. "'You Must Not Know About Me'—on the Willingness to Share Personal Data," *Journal of Economic Behavior & Organization* (141), pp. 1-13.

Sharma, G. 2017. "Pros and Cons of Different Sampling Techniques," *International journal of applied research* (3:7), pp. 749-752.

Shiller, B.R. 2014. *First Degree Price Discrimination Using Big Data*. Brandeis University, Department of Economics.

Shin, J. 2017. "Patient Privacy Decision Making in the Health Big Data Era," *ICIS 2017 Proceedings. 17*, South Korea.

Skatova, A., McDonald, R., Ma, S., and Maple, C. 2023. "Unpacking Privacy: Valuation of Personal Data Protection," *PLoS One* (18:5), p. e0284581.

SmartPLS. 2022. "Smartpls 4." SmartPLS GmbH.

Smith, H.J., Dinev, T., and Xu, H. 2011. "Information Privacy Research: An Interdisciplinary Review," *MIS Quarterly* (35:4), pp. 989-1015.

Smith, H.J., Milberg, S.J., and Burke, S.J. 1996. "Information Privacy: Measuring Individuals' Concerns About Organizational Practices," *MIS Quarterly* (20:No. 02), pp. 167-196.

Solino-Fernandez, D., Ding, A., Bayro-Kaiser, E., and Ding, E.L. 2019. "Willingness to Adopt Wearable Devices with Behavioral and Economic Incentives by Health Insurance Wellness Programs: Results of a Us Cross-Sectional Survey with Multiple Consumer Health Vignettes," *BMC Public Health* (19:1), p. 1649.

Song, C.S., and Kim, Y.-K. 2021. "Predictors of Consumers' Willingness to Share Personal Information with Fashion Sales Robots," *Journal of Retailing and Consumer Services* (63), p. 102727.

Steenkamp, J.-B.E.M., De Jong, M.G., and Baumgartner, H. 2010. "Socially Desirable Response Tendencies in Survey Research," *Journal of Marketing Research* (47:2), pp. 199-214.

Suchman, M.C. 1995. "Managing Legitimacy: Strategic and Institutional Approaches," *Academy of Management Review* (20:3), pp. 571-610.

Supantha, M., Weir, K., Sharma, A., and W. Stevenson, S. 2023. "Mit Altersprüfung: Italien Genehmigt Rückkehr Von Chatgpt." Retrieved 04/28, 2023, from https://www.tagesspiegel.de/wirtschaft/mit-altersprufung-italien-genehmigt-ruckkehr-von-chatgpt-9741452.html.

SurveyCircle. 2023. "Studienteilnehmer Finden. Mit Surveycircle." Retrieved 07/17, 2023, from https://www.surveycircle.com/de/.

Syed, A., Gillela, K., and Venugopal, C. 2013. "The Future Revolution on Big Data," *International Journal of Advanced Research in Computer and Communication Engineering* (2:6), pp. 2446-2451.

Tamkin, A., Brundage, M., Clark, J., and Ganguli, D. 2021. "Understanding the Capabilities, Limitations, and Societal Impact of Large Language Models," *arXiv preprint arXiv:2102.02503*).

Teubner, T., Flath, C.M., Weinhardt, C., van der Aalst, W., and Hinz, O. 2023. "Welcome to the Era of Chatgpt Et Al. The Prospects of Large Language Models," *Business & Information Systems Engineering* (65(2)), pp. 95–101

Thibaut, J.W., and Kelley, H. 1959. *The Social Psychology of Groups*. New York: Wiley.

Thürmel, V., Berger, B., and Hess, T. 2021. "Look What I'm Interested In! Toward a Better Understanding of How Personalization and Self-Reference Drive News Sharing," *Proceedings of the 54th Hawaii International Conference on System Sciences*, Hawaii, pp. 4250-4259.

Trabucchi, D., and Buganza, T. 2019. "Data-Driven Innovation: Switching the Perspective on Big Data," *European Journal of Innovation Management* (22:1), pp. 23-40.

Treiblmaier, H. 2019. "Privacy Revisited: The Impact of Blockchain Technology on the Disclosure of Personal Data," *DIGIT 2019 Proceedings*, Munich.

Treiblmaier, H., and Chong, S. 2011. "Trust and Perceived Risk of Personal Information as Antecedents of Online Information Disclosure: Results from Three Countries," *Journal of Global Information Management (JGIM)* (19:4), pp. 76-94.

Treiblmaier, H., and Pollach, I. 2007. "Users' Perceptions of Benefits and Costs of Personalization," *ICIS 2007 Proceedings*, Montreal: AISeL, p. 141.

Ur, B., Leon, P.G., Cranor, L.F., Shay, R., and Wang, Y. 2012. "Smart, Useful, Scary, Creepy: Perceptions of Online Behavioral Advertising," *SOUPS '12: Proceedings of the Eighth Symposium on Usable Privacy and Security*, Washington D.C.: ACM Press, pp. 1-15.

Vassakis, K., Petrakis, E., and Kopanakis, I. 2018. "Big Data Analytics: Applications, Prospects and Challenges," in *Mobile Big Data. Lecture Notes on Data Engineering and Communications Technologies,* G. Skourletopoulos, Mastorakis, G., Mavromoustakis, C., Dobre, C., Pallis, E. (ed.). Switzerland: Springer Cham, pp. 3-20.

Vesselkov, A., Hämmäinen, H., and Töyli, J. 2019. "Design and Governance of Mhealth Data Sharing," *Communications of the Association for Information Systems* (45:1), pp. 299-321.

Von Neumann, J., and Morgenstern, O. 2007. "Theory of Games and Economic Behavior," in *Theory of Games and Economic Behavior*. Princeton university press.

Wakefield, R. 2013. "The Influence of User Affect in Online Information Disclosure," *The Journal of Strategic Information Systems* (22:2), pp. 157-174.

Wang, T., Duong, T.D., and Chen, C.C. 2016. "Intention to Disclose Personal Information Via Mobile Applications: A Privacy Calculus Perspective," *International Journal of Information Management* (36:4), pp. 531-542.

WebOfScience. 2023. "Results from Web of Science Core Collection." Retrieved 09/27, 2023, from

https://www.webofscience.com/wos/woscc/summary/b4ff8860-eddf-4934-972b-4e86904f4199-a6ee74c8/relevance/1.

Weijters, B., and Baumgartner, H. 2012. "Misresponse to Reversed and Negated Items in Surveys: A Review," *Journal of Marketing Research* (49:5), pp. 737-747.

Weitzman, E.R., Kelemen, S., Kaci, L., and Mandl, K.D. 2012. "Willingness to Share Personal Health Record Data for Care Improvement and Public Health: A Survey of Experienced Personal Health Record Users," *BMC Medical Informatics and Decision Making* (12:1), p. 39.

Wooldridge, J.M. 2015. *Introductory Econometrics: A Modern Approach*, (6 ed.). Boston, MA, USA: Cengage learning.

Wottrich, V.M., van Reijmersdal, E.A., and Smit, E.G. 2018. "The Privacy Trade-Off for Mobile App Downloads: The Roles of App Value, Intrusiveness, and Privacy Concerns," *Decision Support Systems* (106), pp. 44-52.

Xiao, M. 2022. "Supervision Strategy Analysis on Price Discrimination of E-Commerce Company in the Context of Big Data Based on Four-Party Evolutionary Game," *Computational Intelligence and Neuroscience* (2022), p. 2900286.

Xu, H., Luo, X.R., Carroll, J.M., and Rosson, M.B. 2011. "The Personalization Privacy Paradox: An Exploratory Study of Decision Making Process for Location-Aware Marketing," *Decision Support Systems* (51:1), pp. 42-52.

Youn, S. 2009. "Determinants of Online Privacy Concern and Its Influence on Privacy Protection Behaviors among Young Adolescents," *Journal of Consumer affairs* (43:3), pp. 389-418.

Young, J.M., O'Halloran, A., McAulay, C., Pirotta, M., Forsdike, K., Stacey, I., and Currow, D. 2015. "Unconditional and Conditional Incentives Differentially Improved General Practitioners' Participation in an Online Survey: Randomized Controlled Trial," *Journal of Clinical Epidemiology* (68:6), pp. 693-697.

Zhang, X., Liu, S., Chen, X., Wang, L., Gao, B., and Zhu, Q. 2018. "Health Information Privacy Concerns, Antecedents, and Information Disclosure Intention in Online Health Communities," *Information & Management* (55:4), pp. 482-493.

Zhao, W.X., Zhou, K., Li, J., Tang, T., Wang, X., Hou, Y., Min, Y., Zhang, B., Zhang, J., and Dong, Z. 2023. "A Survey of Large Language Models," *arXiv preprint arXiv:2303.18223*).

Zhou, J. 2018. "Factors Influencing People's Personal Information Disclosure Behaviors in Online Health Communities: A Pilot Study," *Asia Pac Journal of Public Health* (30:3), pp. 286-295.

Zhuang, J.J. 2023. "Introducing the Instacart Plugin for Chatgpt." Retrieved 03/23, 2023, from https://www.instacart.com/company/updates/instacart-chatgpt/.

Ziefle, M., Halbey, J., and Kowalewski, S. 2016. "Users' Willingness to Share Data on the Internet: Perceived Benefits and Caveats," *Proceedings of the International Conference on Internet of Things and Big Data*, Stuttgart: Science and Technology Publications, pp. 255-265.

Appendix

References	**Context**
(Ackermann et al., 2022)	Experiment on data sharing with companies of different contexts.

(Agahari and de Reuver, 2022)	Experiment on data sharing in data marketplaces.
(Alemany et al., 2020)	Experiment on sharing behaviour in social network platforms.
(Anderson and Agarwal, 2009)	Experiment on digital health exchange of information.
(Anderson and Agarwal, 2011)	Experiment on sharing behaviour of personal health information.
(Bansal and Nah, 2020)	Experiment on data sharing in the internet, with online businesses and governments.
(Becker et al., 2021)	Experiment on location privacy and information disclosure in transportation services.
(Bélanger et al., 2021)	Experiment on privacy behaviour with Internet of Things devices, namely smart home speakers.
(Benndorf and Normann, 2014)	Experiment on the willingness to sell data in various contexts.
(Both)	Experiment on the willingness to share personal information online in social media, online stores and general public.
(Brown et al., 2016)	Experiment on the sharing behaviour in electronic communication portals in healthcare.
(Caine and Hanania, 2013)	Investigates patients‘ sharing behaviour regarding health information in electronic medical record systems.
(Cichy et al., 2021)	Investigates the sharing and privacy behaviour in the context of Internet of Things, namely connected cars.
(Cocosila and Archer, 2014)	Investigates individual perception of using electronic personal health records.
(Deuker et al., 2012)	Investigates individuals‘ privacy protection behaviour and rationales to use privacy settings on social networking sites.
(Donovan-Kicken et al., 2012)	Investigates individuals‘ behaviour of informing and giving consent in healthcare decisions.
(Durneva, 2020)	Examines patients‘ data sharing behaviour via wearable devices in the healthcare context.
(Esmaeilzadeh, 2019)	Investigates patients‘ behaviour with health information exchange.
(Esmaeilzadeh, 2020)	Investigates patients‘ information disclosure intention in health information exchange applications.
(Farrelly and Chew, 2016)	Investigates individuals‘ willingness to share their social media data in exchange for monetary compensation.
(Frey et al., 2016)	Examines users‘ sharing behaviour and privacy-related behaviour if companies use safer blockchain-technology.

(Gaylin et al., 2011)	Investigates individuals‘ attitudes regarding health information technologies, also regarding data sharing.
(Hann et al., 2007)	Investigates data sharing behaviour in different settings, by altering incentives and benefits in financial, health care and travel websites.
(Hasnain-Wynia et al., 2011)	Investigates individuals‘ data sharing behaviour and attitudes in health care.
(Ioannou et al., 2020)	Examines travellers’ willingness to share data and arising privacy concerns interacting with travel providers.
(Janoschka et al., 2022)	Studies individuals‘ sharing behaviour and concerns when utilizing data-based products and services.
(Karampela et al., 2019)	Investigates users‘ attitudes and willingness to share personal health data.
(Kerath et al., 2013)	Investigates individuals’ attitudes and willingness to share personal genetic information in genetic research.
(Kim and Kim, 2018)	Studies factors influencing the willingness to share personal information for personalized services in media.
(Kim et al., 2017)	Studies individuals‘ attitudes and behaviour in sharing personal health data for healthcare and research via electronic data sharing technology.
(Kim et al., 2019)	Investigates willingness to share personal information in Internet of Things services in healthcare, smart home and transportation.
(Li et al., 2010)	Investigates individuals‘ decision-making regarding information disclosure in the setting of a commercial website.
(Mäki et al., 2023)	Investigates individuals‘ attitudes and data sharing behaviour in various contexts and settings.
(Malhotra et al., 2004)	Examines users‘ privacy concerns and behaviour in various online commercial and financial settings.
(Medford-Davis et al., 2017)	Investigates patients‘ willingness to share their health information through Health Information Exchange technology.
(Patel et al., 2011)	Investigates individuals‘ willingness to participate in electronic health information exchange.
(Pavlou, 2011)	Summarizes relevant findings across various studies in the information privacy literature.
(Phelps et al., 2000)	Investigates individuals‘ personal information exchange behaviour in online shopping.
(Pickard and Swan, 2014)	Investigates individuals’ willingness to share personal health data for research purposes or recommendations.

(Punj, 2019)	Studies individuals‘ willingness to limit shared data by anonymizing their digital identity without any specific context.
(Schudy and Utikal, 2017)	Investigates individuals’ behaviour regarding sharing personal data with unidentified recipients in differently incentivized settings.
(Shin, 2017)	Investigates individuals’ sharing behaviour regarding the exchange of personal health information.
(Skatova et al., 2023)	Investigates individuals‘ behaviour in sharing personal data in a variety of environments.
(Smith et al., 2011)	Interdisciplinary review on the information privacy research.
(Song and Kim, 2021)	Investigates individuals‘ willingness to share personal information with fashion sales robots.
(Treiblmaier, 2019)	Studies individuals‘ willingness to share personal information in the context of blockchain technology.
(Vesselkov et al., 2019)	Examines data sharing in mobile health applications via wearables and digital health devices.
(Wakefield, 2013)	Investigates individuals‘ online disclosure behaviour on commercial websites.
(Wang et al., 2016)	Investigates individuals‘ intentions to disclose personal data in mobile applications.
(Weitzman et al., 2012)	Studies willingness to share personally controlled health records across health topics.
(Zhang et al., 2018)	Examines individuals‘ privacy behaviour in online health communities
(Zhou, 2018)	Examines individuals‘ disclosure behaviour in online health communities
(Ziefle et al., 2016)	Investigates users’ willingness to share personal data in digital services and social networks.

A-1: Literature review synthesis

A-2: Overview on determinants of data sharing decisions across studies

References	Privacy & Data Security Concerns	Benefits & Incentives	Control	Trust	Severity	Purpose	Education or Experience	Access	Awareness	Transparency
Ackermann *et al.*, 2022	x	x	x		x	x	x			
Agahari and de Reuver,	x		x	x				x		
Alemany *et al.*, 2020	x	x	x		x		x	x		
Anderson and Agarwal,					x	x		x	x	
Anderson and Agarwal,		x								
Bansal and Nah, 2020	x	x	x	x		x		x	x	
Becker *et al.*, 2021	x	x	x	x	x		x	x	x	x
Bélanger *et al.*, 2021	x		x		x				x	
Benndorf and Normann,		x		x	x	x				
Both, 2021	x			x	x					
Brown *et al.*, 2016			x			x				
Caine and Hanania, 2013			x	x	x			x		
Cichy *et al.*, 2021	x	x	x	x		x		x	x	x
Cocosila and Archer, 2014	x	x		x		x	x			
Deuker *et al.*, 2012			x	x				x		
Donovan-Kicken *et al.*, 2012					x		x		x	
Durneva, 2020		x								
Esmaeilzadeh, 2019	x	x	x	x			x	x		x
Esmaeilzadeh, 2020				x					x	x
Farrelly and Chew, 2016		x								
Frey *et al.*, 2016		x	x							x
Gaylin *et al.*, 2011	x	x					x			
Hann *et al.*, 2007	x	x	x	x						x
Hasnain-Wynia *et al.*, 2011		x			x	x				
Ioannou *et al.*, 2020	x		x	x	x		x		x	
Janoschka *et al.*, 2022		x		x					x	
Karampela *et al.*, 2019				x		x	x			
Kerath *et al.*, 2013	x	x		x	x	x	x			
Kim and Kim, 2018	x	x	x	x	x	x	x		x	x
Kim *et al.*, 2017	x	x		x	x					
Kim *et al.*, 2019	x	x	x	x	x	x	x			
Li *et al.*, 2010	x	x								
Mäki *et al.*, 2023	x	x			x					
Malhotra *et al.*, 2004	x		x	x	x		x	x	x	x
Medford-Davis *et al.*, 2017	x	x	x		x	x				
Patel *et al.*, 2011	x	x	x		x	x				
Pavlou, 2011	x	x	x	x	x	x		x		x
Phelps *et al.*, 2000	x	x	x	x	x			x		
Pickard and Swan, 2014	x	x	x	x	x	x			x	x
Punj, 2019	x		x		x					
Schudy and Utikal, 2017	x	x	x	x	x	x		x		x
Shin, 2017	x		x	x	x		x			
Skatova *et al.*, 2023		x	x		x	x	x			
Smith *et al.*, 2011	x	x	x	x	x		x	x	x	x
Song and Kim, 2021		x		x						
Treiblmaier, 2019			x	x						
Vesselkov et al., 2019			x					x		
Wakefield, 2013	x			x						
Wang *et al.*, 2016	x	x	x		x	x		x		
Weitzman *et al.*, 2012	x		x		x	x				
Zhang *et al.*, 2018	x	x		x	x	x				x
Zhou, 2018	x									
Ziefle *et al.*, 2016	x	x	x	x	x	x	x			
	32	**32**	**31**	**29**	**29**	**20**	**16**	**16**	**13**	**12**

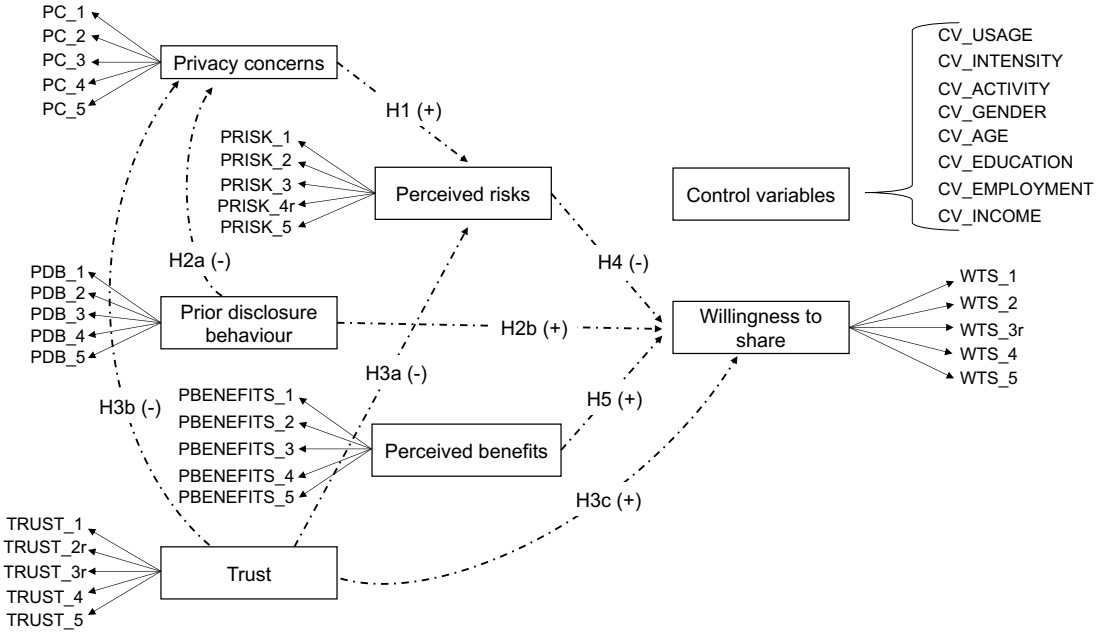
Determinant	Definition	Superordinate main variables	Reference
Accountability of the app provider	Lack of accountability in events of data breaches or regarding previous incidents	Trust, Transparency, Control	(Becker et al., 2021)
Accuracy and real-time information provision	More accurate location sharing reveals higher benefits (e. g. navigation)	Benefits and Incentives	(Becker et al., 2021)
Age	Age and generational differences	Demographics	(Becker et al., 2021)
Anonymity	If shared data is not traceable back to provider.	Privacy and data security concerns, Control, Severity	(Becker et al., 2021)
Attitude	Different a priori risk/privacy types and behaviour towards privacy	Privacy and data security concerns, Control	(Becker et al., 2021)
Choice	Having a real choice to decide about sharing data or not	Control	(Becker et al., 2021)
Competitiveness	"risk that the competitive advantage or competitiveness in general of the organisation is negatively impacted by sharing data with the general public"	Control, Access	(Enders et al., 2020, p. 5)
Convenience and frustration	Direct benefits/costs from using the (data sharing) application	Benefits and Incentives	(Becker et al., 2021)
Coreness	"proximity of the data to core of business operations"	Severity	(Enders et al., 2020, p. 5)
Currentness	"explains how recent the data is and its gradient of value decay depending on the type of data"	Severity	(Enders et al., 2020, p. 5)

Data Misappropriation	"risk that data is used for purposes other than initially intended. Drawbacks may not only impact the data provider itself but extend to a macro level"	Access, Control, Transparency	(Enders et al., 2020, p. 5)
Defeatist towards preventing cybercrime	Level of feeling helpless and unable to prevent cyber crime	Control, Trust	(Becker et al., 2021)
Ease of use	Benefits from convenience and high usability	Benefits and Incentives	(Becker et al., 2021)
Ecosystem knowledge	Knowing types and formats of shared data	Education or Experience, Transparency	(Becker et al., 2021)
Environment	Decisions influenced by environmental factors (not specified in study but can be for instance focus groups environment)	n. a.	(Becker et al., 2021)
Extent	"what percentage of a given dataset is shared"	Severity, Transparency	(Enders et al., 2020, p. 5)
Granularity	"measure of level of detail of given dataset"	Severity, Transparency	(Enders et al., 2020, p. 5)
Innovation Opportunity	"potential benefit of innovation being created in the open data ecosystem by providing data publicly. Innovation may benefit the data provider directly and/or create value on a macroeconomic level"	Purpose, Benefits and Incentives	(Enders et al., 2020, p. 5)
Interoperability	"ability of software or computer systems to exchange information in a given data ecosystem"	Benefits and Incentives	(Enders et al., 2020, p. 5)
Job	Not further specified, but likely in terms of demographic control variable	Demographics	(Becker et al., 2021)

Legal	"risk that the shared data may violate legal restrictions"	Transparency, Privacy and data security concerns	(Enders et al., 2020, p. 5)
Mood	Decisions influenced by current mood	n. a.	(Becker et al., 2021)
Ownership	Decision depends on whether the data is stored permanently or only shared temporarily	Control, Trust, Access, Transparency	(Becker et al., 2021)
Privacy	"risk that published data can be used to track information back to an individual or an organisation"	Control, Trust, Access, Transparency	(Enders et al., 2020, p. 5)
Quality	"level of completeness and accuracy of a given dataset"	Severity, Awareness, Transparency	(Enders et al., 2020, p. 5)
Reputation	Whether the app provider in the past misused data or experienced data breaches	Trust	(Becker et al., 2021)
Safety (Physical)	In case of data breaches e. g. physical attacks or burglaries can be induced by revealing personal information	Control, Trust, Severity, Access	(Becker et al., 2021)
Safety (Psychological)	Psychological feeling of improved/reduced safety through e. g. knowing where someone is through location sharing	Control, Trust, Severity, Access	(Becker et al., 2021)
Security	Security of the device and the infrastructure around the data requesting party	Trust, Access	(Becker et al., 2021)
Size of Audience	Lack of knowledge on size and specific audience of the shared data	Access, Transparency, Education or Experience	(Becker et al., 2021)

A-3: Definitions of further determinants

Value attributed to data	Whether the users value the data high or low	Severity, Purpose	(Becker et al., 2021)
Vulnerability	Vulnerability of the network itself	Trust, Transparency	(Becker et al., 2021)



A-4: Initial research model in detailed version

References	Constructs	Items	Descriptions
(Motiwalla and Li, 2016)	Prior disclosure behaviour (PDB)	PDB_1	I give consent to share my personal data.
		PDB_2	I enter my own/friend's/colleague's full name, address, or other sensitive personally identifiable information.
		PDB_3	I enter data from my employer/university/school in order to improve my efficiency.
		PDB_4	I enter my health-related information in order to get for example information on potential diagnosis.
		PDB_5	I enter information on my own/friend's/colleague's behaviour regarding for example purchase behaviour, preferences, dislikes.
(Mohamed and Ahmad, 2012; Xu et al., 2011)	Perceived benefits (PBENEFITS)	PBENEFITS_1	ChatGPT reduces my searching time to find information that I need.
		PBENEFITS_2	ChatGPT provides me with the convenience to instantly access the information that I need.
		PBENEFITS_3	ChatGPT increases my efficiency.
		PBENEFITS_4	ChatGPT improves my work-life balance.
		PBENEFITS_5	Overall, I feel that using ChatGPT is beneficial.
(Xu et al., 2011)		PRISK_1	Providing my personal information in

	Perceived risks (PRISKS		ChatGPT would involve many unexpected problems.
		PRISK_2	Disclosing my personal information to ChatGPT would be risky.
		PRISK_3	The potential for loss in disclosing my personal information to ChatGPT would be high.
		PRISK_4r	I feel safe giving my personal data to ChatGPT.
		PRISK_5	There would be too much uncertainty associated with giving my personal data to ChatGPT.
(Dinev and Hart, 2006)	Privacy concerns (PC)	PC_1	I am concerned that the information I submit through ChatGPT could be misused.
		PC_2	I am concerned that a person can find private information about me on the internet that I submitted through ChatGPT.
		PC_3	I am concerned about submitting information through ChatGPT, because of what others might do with it.
		PC_4	I am concerned about submitting information through ChatGPT, because it could be used in a way I did not foresee.
		PC_5	I believe that online privacy is invaded if control is lost or unwillingly reduced as a result of marketing by ChatGPT.

(Cichy et al., 2021; Dinev and Hart, 2006)	Trust (TRUST)	TRUST_1	I believe that ChatGPT is trustworthy.
		TRUST_2r	I would find it necessary to be cautios in dealing with ChatGPT.
		TRUST_3r	I believe ChatGPT could not be relied upon to keep its promises.
		TRUST_4	ChatGPT provides a reliable environment in which to conduct business transactions.
		TRUST_5	ChatGPT handles personal information submitted by users in a competent fashion.
(Dinev and Hart, 2006; Xu et al., 2011)	Willingness to share (WTS)	WTS_1	I am likely to disclose my personal information in ChatGPT.
		WTS_2	I am willing to disclose my personal information in ChatGPT.
		WTS_3r	Disclosing my personal information in ChatGPT for its services is unlikely for me.
		WTS_4	I am willing to disclose highly personal and password-protected information in ChatGPT.
		WTS_5	I am willing to conduct sales transactions through ChatGPT that require me to provide credit card information.
(Wang et al., 2016)	Control variables (CV)	CV_USAGE	Have you ever used ChatGPT?
		CV_INTENSITY	How often do you use ChatGPT?

		CV_ACTIVITY	How many hours per day do you approximately use smartphone, computer, tablet etc.?
		CV_GENDER	Please indicate your gender.
		CV_AGE	Please indicate your age.
		CV_EDUCATION	Please indicate your highest level of education.
		CV_EMPLOYMENT	Which statement best describes your current employment status?
		CV_INCOME	Please indicate your approximate income per month after tax (Nettoeinkommen).

A-5: Constructs and items

Data Sharing & ChatGPT

Start of Block: Introduction

Dear participant,

thank you for your willingness to participate in this survey.

It is **conducted by the Institute for Digital Management and New Media** (DMM) at the Ludwig Maximilian University of Munich and **focuses on data sharing in ChatGPT (operated by OpenAI) and its plugins** (see next page for more information)**.**

The following applies:
Chance to win **3 x 20 € Amazon vouchers** after finishing the survey (Gutscheinverlosung).
Participation in the study will take approximately **5-8 minutes**.
Your answers will be treated **anonymously** and confidentially.
No data will be passed on to third parties.
Always refer to your experiences with ChatGPT (if you have already made some).
Please **answer all questions carefully**.
If you have any questions, please do not hesitate to contact me at: julian.held@campus.lmu.de.

Let's get started!

Imprint and Privacy Statement

End of Block: Introduction

Start of Block: Briefing

Background

Please read the background information carefully:

What is ChatGPT?

one of the biggest and most famous large language models, operated by the company "OpenAI"
it works by predicting the next, most-likely word in a sentence (see examples, capabilities and limitations of the model below)

What are plugins for ChatGPT?

plugins enable users to access and use external data for improved results (plugins access e.g. private documents, payment data...)
plugins are developed by lots of external providers (see below) and need to be installed on the user's device ChatGPT thereby is intermediate between plugin and user Examples:

Exemplary data collected by ChatGPT & plugins

OpenAI and the plugins collect lots of users' data
this collected data may include:

1) **Exemplary data collected by ChatGPT:**

Account information (name, contact information, payment information, transaction history...)
User content (input, file uploads, feedbacks)
Log data (IP address, browser information, interaction behavior)
Usage data (engaged content, computer type, device type, settings...)
Device information (device name, system, browser information)
Cookies
Analytics (interaction behavior)
Communication behavior
Social media information (contact details, interaction behavior, activity)

2) **Exemplary data collected by plugins** (examplary "*Instacart*", see above):

Personal health information (past or present mental and physical health condition)
Order information (delivery address, ordered items...)
Age/identity verification Information (government ID information)
Vehicle information (vehicle license plate numbers)
Payment information (billing address, payment method)
Information posted in public spaces (uncontrolled third-party access)
Information from surveys, retail partners or interactions
Information about others (information on referred friends' or family members' addresses, names, mails)
Location

Users' collected data may be used for:

provide, manage and improve services
personalize communication and customer support
marketing and advertising activities

identify trends or predict behavior
third-party disclosure...

Page Break

Control Question **Control Question:**
Please choose the correct answer based on your understanding of the previous introduction.

ChatGPT and its plugins may collect following data:

- ○ Data on payments, identity information, location and more (1)
- ○ Only cookies (2)

Display This Question:

If Loop current: Control Question: Please choose the correct answer based on your understanding of the previous in... = Only cookies

Incorrect answer **You have answered the control question incorrectly.**
For the sake of the validity of the survey you are therefore exempt from further participation.

Thank you for your participation!

End of Block: Briefing

Start of Block: Prior Disclosure Behavior

PDB Please rate to what extent the following statements apply to you.

	Never (1)	(2)	(3)	(4)	(5)	(6)	Always (7)
I give consent to share my personal data. (PDB_1)	○	○	○	○	○	○	○
I enter my own/friend's/colleague's full name, address or other sensitive personally identifiable information. (PDB_2)	○	○	○	○	○	○	○
I enter data from my employer/university/school in order to improve my efficiency. (PDB_3)	○	○	○	○	○	○	○
I enter my health related information in order to get for example information on potential diagnosis. (PDB_4)	○	○	○	○	○	○	○
I enter information on my own/friend's/colleague's behavior regarding for example purchase behavior, preferences, dislikes. (PDB_5)	○	○	○	○	○	○	○

End of Block: Prior Disclosure Behavior

Start of Block: Perceived Benefits

PBENEFITS Please rate to what extent the following statements apply to you.

	Strongly disagree (1)	(2)	(3)	(4)	(5)	(6)	Strongly agree (7)
ChatGPT reduces my searching time to find information that I need. (PBENEFITS_1)	○	○	○	○	○	○	○
ChatGPT provides me with the convenience to instantly access the information that I need. (PBENEFITS_2)	○	○	○	○	○	○	○
ChatGPT increases my efficiency. (PBENEFITS_3)	○	○	○	○	○	○	○
ChatGPT improves my work-life balance. (PBENEFITS_4)	○	○	○	○	○	○	○
Overall I feel that using ChatGPT is beneficial. (PBENEFITS_5)	○	○	○	○	○	○	○

End of Block: Perceived Benefits

Start of Block: Perceived Risks

PRISK Please rate to what extent the following statements apply to you.

	Strongly disagree (1)	(2)	(3)	(4)	(5)	(6)	Strongly agree (7)
Providing my personal information in ChatGPT would involve many unexpected problems. (PRISK_1)	○	○	○	○	○	○	○
Disclosing my personal information to ChatGPT would be risky. (PRISK_2)	○	○	○	○	○	○	○
The potential for loss in disclosing my personal information to ChatGPT would be high. (PRISK_3)	○	○	○	○	○	○	○
I feel safe giving my personal data to ChatGPT. (PRISK_4r)	○	○	○	○	○	○	○
There would be too much uncertainty associated with giving my personal data to ChatGPT. (PRISK_5)	○	○	○	○	○	○	○

End of Block: Perceived Risks

Start of Block: Privacy Concerns

PC Please rate to what extent the following statements apply to you.

	Strongly disagree (1)	(2)	(3)	(4)	(5)	(6)	Strongly agree (7)
I am concerned that the information I submit through ChatGPT could be misused. (PC_1)	○	○	○	○	○	○	○
I am concerned that a person can find private information about me on the internet that I submitted through ChatGPT. (PC_2)	○	○	○	○	○	○	○
I am concerned about submitting information through ChatGPT, because of what others might do with it. (PC_3)	○	○	○	○	○	○	○
I am concerned about submitting information through ChatGPT, because it could be used in a way I did not foresee. (PC_4)	○	○	○	○	○	○	○
I believe that online privacy is invaded if control is lost or unwillingly reduced as a result of marketing by ChatGPT. (PC_5)	○	○	○	○	○	○	○

End of Block: Privacy Concerns

Start of Block: Trust

TRUST Please rate to what extent the following statements apply to you.

	Strongly disagree (1)	(2)	(3)	(4)	(5)	(6)	Strongly agree (7)
I believe that ChatGPT is trustworthy. (TRUST_1)	○	○	○	○	○	○	○
I would find it necessary to be cautios in dealing with ChatGPT. (TRUST_2r)	○	○	○	○	○	○	○
I believe ChatGPT could not be relied upon to keep its promises. (TRUST_3r)	○	○	○	○	○	○	○
ChatGPT provides a reliable environment in which to conduct business transactions. (TRUST_4)	○	○	○	○	○	○	○
ChatGPT handles personal information submitted by users in a competent fashion. (TRUST_5)	○	○	○	○	○	○	○

End of Block: Trust

Start of Block: Willingness to share personal data

WTS Please rate to what extent the following statements apply to you.

	Strongly disagree (1)	(2)	(3)	(4)	(5)	(6)	Strongly agree (7)
I am likely to disclose my personal information in ChatGPT. (WTS_1)	○	○	○	○	○	○	○
I am willing to disclose my personal information in ChatGPT. (WTS_2)	○	○	○	○	○	○	○
Disclosing my personal information in ChatGPT for its services is unlikely for me. (WTS_3r)	○	○	○	○	○	○	○
I am willing to disclose highly personal and password-protected information in ChatGPT. (WTS_4)	○	○	○	○	○	○	○
I am willing to conduct sales transactions through ChatGPT that require me to provide credit card information. (WTS_5)	○	○	○	○	○	○	○

End of Block: Willingness to share personal data

Start of Block: Control Variables

CV_USAGE **Have you ever used ChatGPT?**

- ○ Yes (1)
- ○ No (2)

CV_INTENSITY **How often do you use ChatGPT?**

- ○ Never (1)
- ○ Less than once a month (2)
- ○ Monthly (3)
- ○ At least once a week (4)
- ○ Several times a week (5)
- ○ Every Day (6)

CV_ACTIVITY **How many hours per day do you approximately use smartphone, computer, tablet etc.?**

- ○ < 1 hour (1)
- ○ 1-2 hours (2)
- ○ 2-3 hours (3)
- ○ 3-4 hours (4)
- ○ 4-5 hours (5)
- ○ > 5 hours (6)

Page Break

CV_GENDER **Please indicate your gender.**

- ○ Male (1)
- ○ Female (2)
- ○ Non-Binary/Third Gender (3)
- ○ Prefer not to say (4)

CV_AGE **Please indicate your age.**

- ○ < 16 (1)
- ○ 17-20 (2)
- ○ 21-30 (3)
- ○ 31-40 (4)
- ○ 41-50 (5)
- ○ 51-60 (6)
- ○ > 60 (7)
- ○ Prefer not to say (8)

CV_EDUCATION **Please indicate your highest level of education.**

- ○ No degree (1)
- ○ Secondary school leaving certificate/elementary school leaving certificate (2)
- ○ Intermediate school leaving certificate (Realschulabschluss) (3)
- ○ Advanced technical college entrance qualification (Fachabitur) (4)
- ○ General entrance qualification for universities of applied sciences (Abitur) (5)
- ○ Completed vocational training (Berufsausbildung) (6)
- ○ Bachelor's degree (7)
- ○ Master's degree/Diploma/State examination (8)
- ○ Doctoral degree (9)
- ○ Prefer not to say (10)

CV_EMPLOYMENT **Which statement best describes your current employment status?**

- ○ School Student (1)
- ○ University Student (2)
- ○ Apprenticeship (Ausbildung) (3)
- ○ Part-Time Employed (4)
- ○ Full-Time Employed (5)
- ○ Self-Employed (6)
- ○ Unemployed (7)
- ○ Prefer not to say (8)

CV_INCOME **Please indicate your approximate income per month after tax (Nettoeinkommen).**

- ○ < 500 € (1)
- ○ 501 - 1000 € (2)
- ○ 1001 - 2000 € (3)
- ○ 2001 - 3000 € (4)
- ○ > 3001 € (5)
- ○ Prefer not to say (6)

End of Block: Control Variables

Start of Block: Feedback

Feedback **Please continue to the next page to submit your answer.**

Comments, suggestions, reactions, positive/negative aspects, or problems encountered with regard to the survey are very welcome in the field below.

We appreciate your feedback very much!

_____

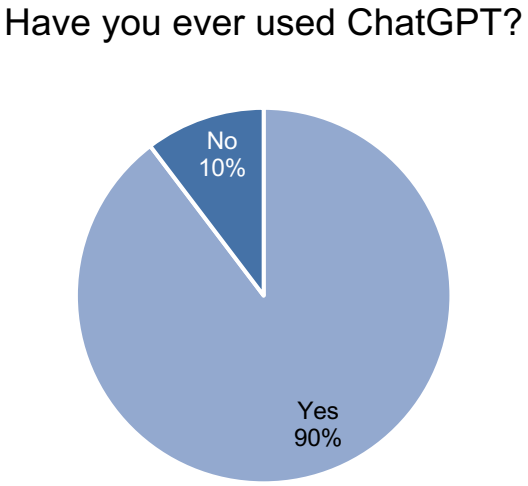
End of Block: Feedback

Start of Block: Raffle

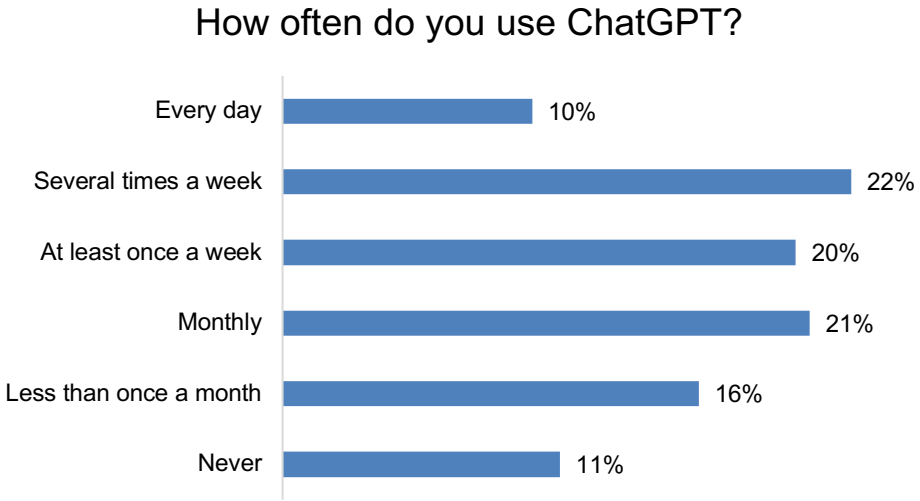
Raffle **Would you like to enter the raffle (Gutscheinverlosung) to win a prize? Your response will still remain anonymous.**

- ○ Yes (1)
- ○ No (2)

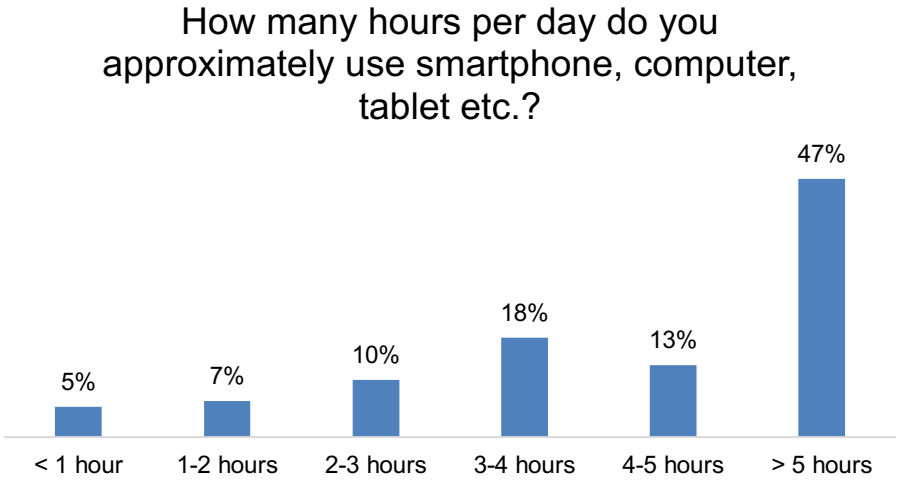
End of Block: Raffle



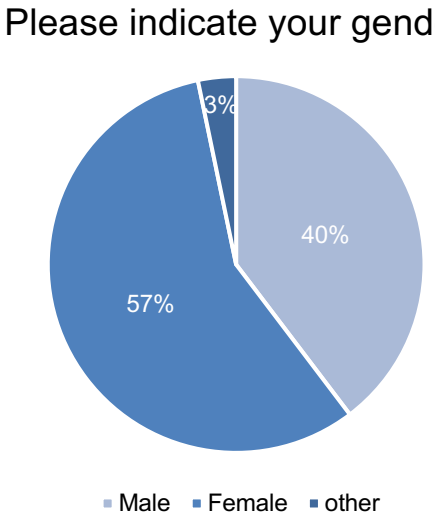
A-7: CV_Usage



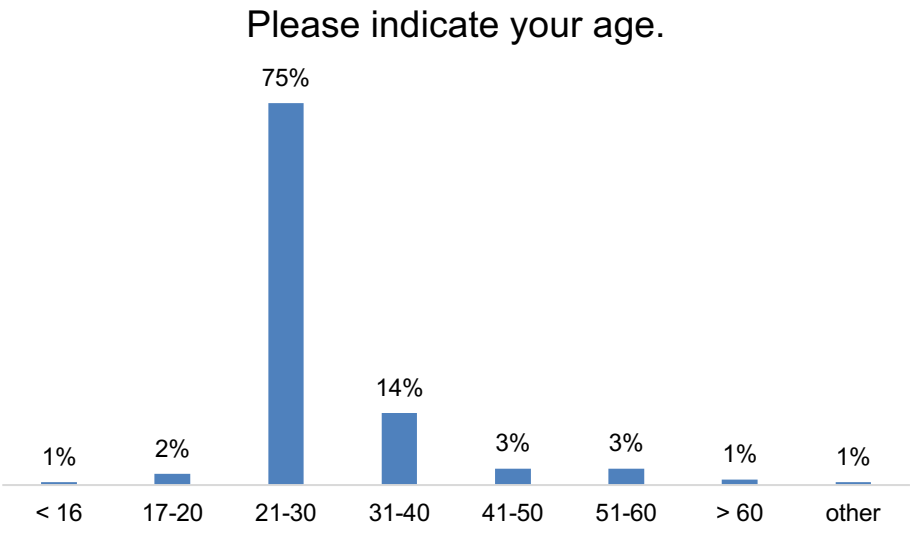
A-8: CV_Intensity



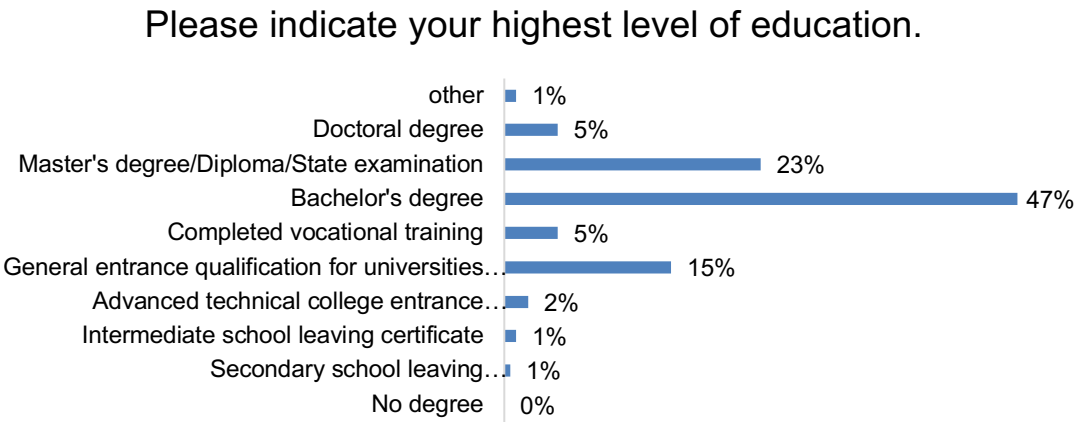
A-9: CV_Activity



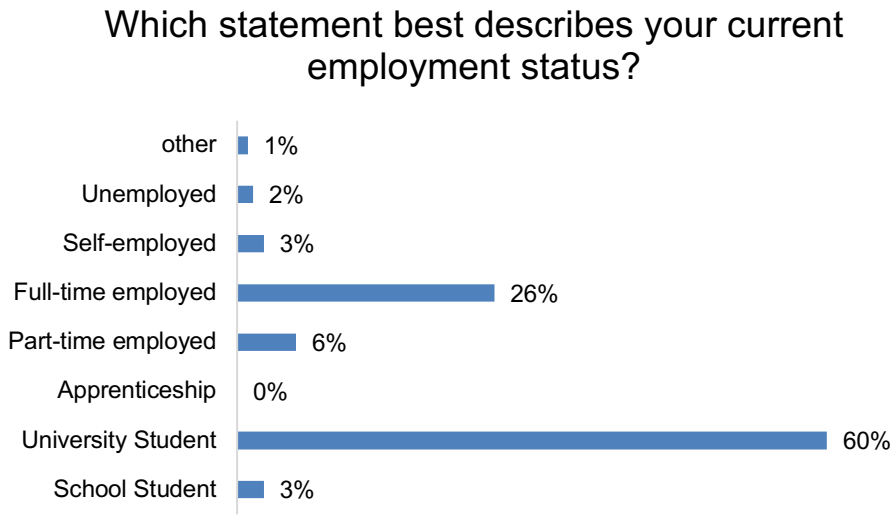
A-10: CV_Gender



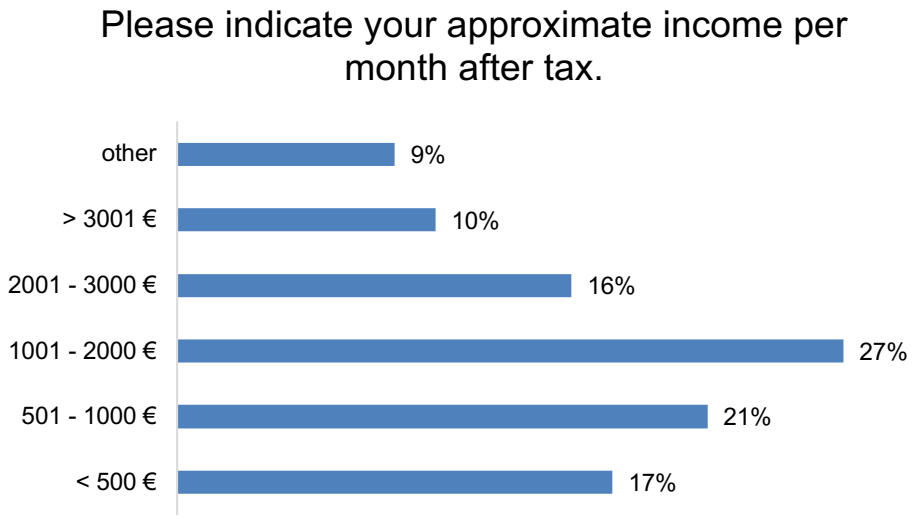
A-11: CV_Age



A-12: CV_Education



A-13: CV_Employment



A-14: CV_Income

Name	No.	Type	Missings	Mean	Median	Scale min	Scale max	Observed min	Observed max	Standard deviation	Excess kurtosis	Skewness	Cramér-von Mises p value
PDB_1	0	MET	0	3.609	3.000	1.000	7.000	1.000	7.000	1.564	-0.893	0.179	0.000
PDB_2	1	MET	0	2.886	2.000	1.000	7.000	1.000	7.000	1.698	-0.417	0.757	0.000
PDB_3	2	MET	0	3.364	3.000	1.000	7.000	1.000	7.000	1.646	-0.740	0.316	0.000
PDB_4	3	MET	0	2.908	2.000	1.000	7.000	1.000	7.000	1.861	-0.699	0.702	0.000
PDB_5	4	MET	0	2.880	2.000	1.000	7.000	1.000	7.000	1.762	-0.673	0.634	0.000
PBENEFITS_1	5	MET	0	4.929	5.000	1.000	7.000	1.000	7.000	1.710	-0.115	-0.750	0.000
PBENEFITS_2	6	MET	0	4.723	5.000	1.000	7.000	1.000	7.000	1.705	-0.400	-0.635	0.000
PBENEFITS_3	7	MET	0	5.076	5.000	1.000	7.000	1.000	7.000	1.695	0.080	-0.943	0.000
PBENEFITS_4	8	MET	0	3.712	4.000	1.000	7.000	1.000	7.000	1.769	-1.018	0.046	0.000
PBENEFITS_5	9	MET	0	5.152	6.000	1.000	7.000	1.000	7.000	1.658	0.093	-0.916	0.000
PRISK_1	10	MET	0	4.679	5.000	2.000	7.000	2.000	7.000	1.323	-0.679	-0.074	0.000
PRISK_2	11	MET	0	4.962	5.000	1.000	7.000	1.000	7.000	1.505	-0.679	-0.504	0.000
PRISK_3	12	MET	0	4.609	5.000	1.000	7.000	1.000	7.000	1.343	-0.231	-0.112	0.000
PRISK_4inv	13	MET	0	5.332	6.000	1.000	7.000	1.000	7.000	1.450	-0.436	-0.616	0.000
PRISK_5	14	MET	0	5.125	5.000	1.000	7.000	1.000	7.000	1.564	-0.375	-0.639	0.000
PC_1	15	MET	0	5.163	5.000	1.000	7.000	1.000	7.000	1.484	-0.307	-0.636	0.000
PC_2	16	MET	0	4.625	5.000	1.000	7.000	1.000	7.000	1.774	-0.983	-0.361	0.000
PC_3	17	MET	0	4.951	5.000	1.000	7.000	1.000	7.000	1.547	-0.574	-0.494	0.000
PC_4	18	MET	0	5.304	5.000	1.000	7.000	1.000	7.000	1.461	0.226	-0.815	0.000
PC_5	19	MET	0	5.109	5.000	1.000	7.000	1.000	7.000	1.414	-0.456	-0.427	0.000
TRUST_1	20	MET	0	3.674	4.000	1.000	7.000	1.000	7.000	1.490	-0.688	-0.063	0.000
TRUST_2inv	21	MET	0	2.701	3.000	1.000	7.000	1.000	7.000	1.244	0.242	0.636	0.000
TRUST_3inv	22	MET	0	3.500	4.000	1.000	6.000	1.000	6.000	1.281	-0.555	-0.063	0.000
TRUST_4	23	MET	0	2.984	3.000	1.000	7.000	1.000	7.000	1.498	-0.401	0.507	0.000
TRUST_5	24	MET	0	3.549	4.000	1.000	7.000	1.000	7.000	1.346	0.113	0.205	0.000
WTS_1	25	MET	0	2.886	3.000	1.000	7.000	1.000	7.000	1.572	-0.196	0.698	0.000
WTS_2	26	MET	0	2.674	2.000	1.000	7.000	1.000	7.000	1.526	0.454	1.008	0.000
WTS_3inv	27	MET	0	3.125	3.000	1.000	7.000	1.000	7.000	1.598	-0.707	0.502	0.000
WTS_4	28	MET	0	1.929	1.000	1.000	7.000	1.000	7.000	1.567	2.355	1.825	0.000
WTS_5	29	MET	0	1.853	1.000	1.000	7.000	1.000	7.000	1.443	3.101	1.923	0.000
CV_USAGE	30	0|1	0	0.897	1.000	0.000	1.000	0.000	1.000	0.304	4.966	-2.629	0.000
CV_INTENSITY	31	MET	0	3.560	4.000	1.000	6.000	1.000	6.000	1.502	-1.015	-0.111	0.000
CV_ACTIVITY	32	MET	0	4.668	5.000	1.000	6.000	1.000	6.000	1.551	-0.307	-0.904	0.000
CV_GENDER_recoded	33	0|1	6	0.410	0.000	0.000	1.000	0.000	1.000	0.492	-1.885	0.369	0.000
CV_AGE	34	MET	1	3.317	3.000	1.000	7.000	1.000	7.000	0.829	6.188	2.197	0.000
CV_EDUCATION	35	MET	2	6.841	7.000	2.000	9.000	2.000	9.000	1.259	1.098	-0.946	0.000
CV_EMPLOYMENT_recoded	36	MET	2	5.286	5.000	1.000	7.000	1.000	7.000	1.324	1.196	-0.725	0.000
CV_INCOME	37	MET	15	2.769	3.000	0.000	5.000	0.000	5.000	1.269	-0.882	0.144	0.000

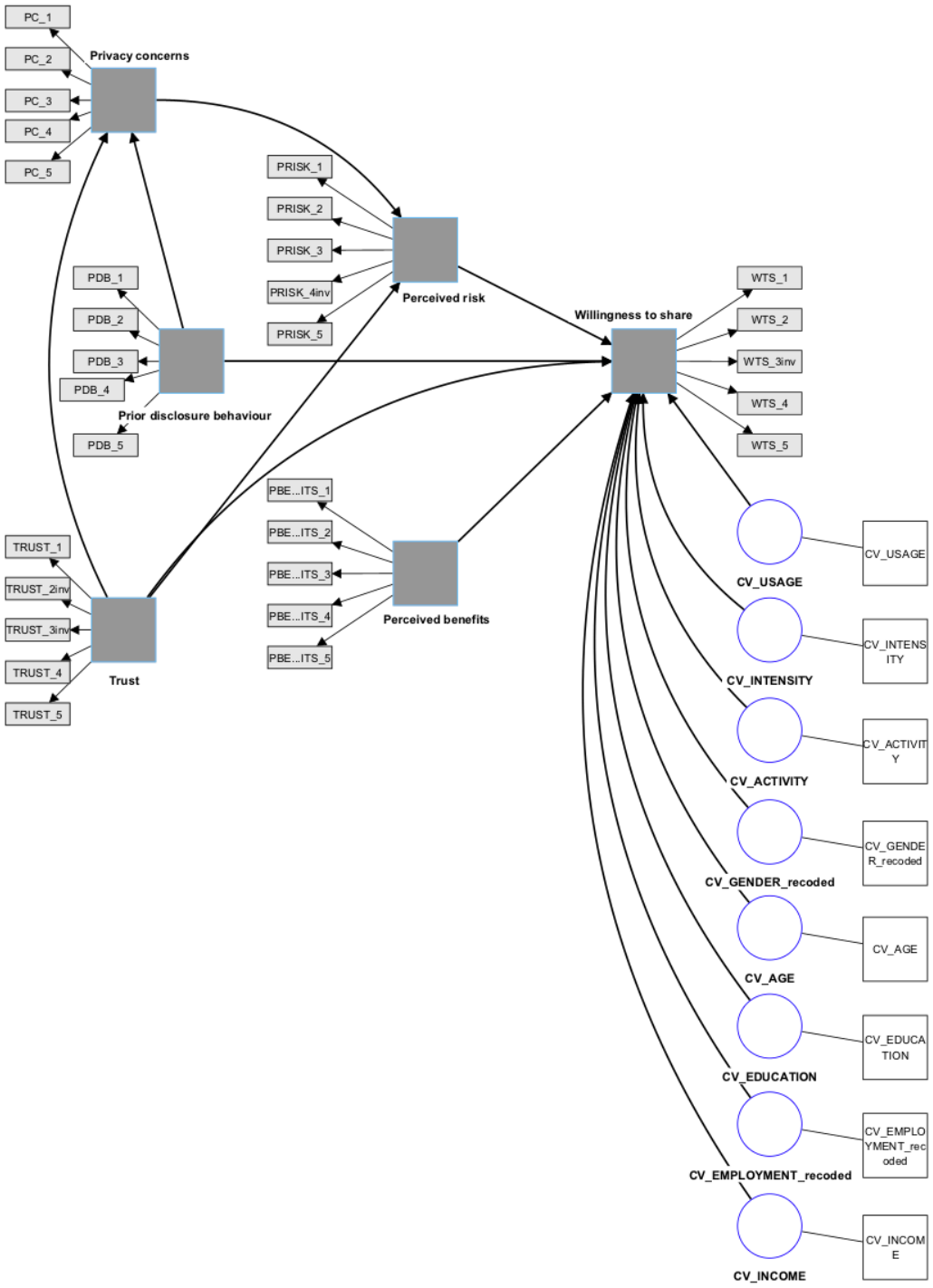
A-15: Analysis of indicators

	PDB_1	PDB_2	PDB_3	PDB_4	PDB_5	PBENEFITS_1	PBENEFITS_2	PBENEFITS_3	PBENEFITS_4	PBENEFITS_5	PRISK_1	PRISK_2	PRISK_3	PRISK_4inv	PRISK_5	PC_1	PC_2	PC_3	PC_4	PC_5
CV_ACTIVITY	-0.065	0.076	0.017	-0.003	-0.044	-0.099	-0.072	-0.048	-0.267	-0.111	0.041	0.113	0.011	0.204	0.031	0.071	0.020	0.032	0.119	0.016
CV_AGE	-0.033	-0.021	-0.128	-0.147	-0.072	-0.201	-0.173	-0.170	-0.145	-0.147	-0.033	-0.050	0.056	0.006	-0.031	-0.029	0.036	-0.002	-0.008	0.128
CV_EDUCATION	0.070	0.076	-0.073	-0.118	0.024	-0.252	-0.220	-0.221	-0.200	-0.212	0.083	-0.004	0.106	-0.011	0.077	0.036	0.018	0.062	0.084	0.077
CV_EMPLOYMENT_recoded	0.011	0.023	0.064	0.041	0.134	-0.090	-0.032	0.011	0.026	-0.030	0.099	0.064	0.149	-0.063	0.204	0.214	0.197	0.182	0.174	0.272
CV_GENDER_recoded	0.004	0.064	0.215	0.098	0.056	0.104	0.132	0.200	0.103	0.207	-0.015	-0.090	0.071	-0.181	-0.042	-0.128	-0.087	-0.116	-0.171	-0.064
CV_INCOME	0.104	-0.013	0.011	-0.014	0.043	-0.016	0.025	0.039	-0.021	0.002	0.094	-0.001	0.132	-0.039	0.019	0.080	0.033	0.087	0.028	0.173
CV_INTENSITY	0.269	0.187	0.267	0.180	0.222	0.493	0.387	0.566	0.451	0.596	-0.178	-0.214	-0.115	-0.412	-0.164	-0.243	-0.192	-0.217	-0.268	-0.169
CV_USAGE	0.178	0.082	0.151	0.108	0.068	0.362	0.332	0.394	0.308	0.462	-0.177	-0.163	-0.152	-0.243	-0.076	-0.191	-0.142	-0.172	-0.186	-0.151
PBENEFITS_1	0.232	0.321	0.364	0.464	0.354	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
PBENEFITS_2	0.224	0.327	0.321	0.394	0.315	0.812	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
PBENEFITS_3	0.278	0.296	0.317	0.402	0.309	0.744	0.690	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
PBENEFITS_4	0.177	0.201	0.260	0.401	0.359	0.610	0.583	0.631	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
PBENEFITS_5	0.247	0.242	0.358	0.380	0.332	0.746	0.657	0.792	0.641	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
PC_1	-0.214	-0.144	-0.207	-0.162	-0.196	-0.227	-0.182	-0.251	-0.297	-0.359	0.494	0.521	0.531	0.447	0.602	1.000	0.000	0.000	0.000	0.000
PC_2	-0.155	-0.021	-0.059	-0.043	-0.020	-0.140	-0.018	-0.187	-0.135	-0.276	0.472	0.457	0.543	0.260	0.440	0.686	1.000	0.000	0.000	0.000
PC_3	-0.183	-0.012	-0.008	-0.103	-0.036	-0.221	-0.156	-0.272	-0.226	-0.368	0.478	0.485	0.556	0.397	0.559	0.744	0.748	1.000	0.000	0.000
PC_4	-0.217	-0.074	-0.148	-0.118	-0.077	-0.224	-0.208	-0.279	-0.279	-0.369	0.483	0.504	0.490	0.447	0.597	0.787	0.698	0.780	1.000	0.000
PC_5	-0.126	-0.056	-0.085	-0.085	-0.069	-0.165	-0.139	-0.207	-0.213	-0.320	0.457	0.434	0.449	0.255	0.547	0.647	0.627	0.646	0.702	1.000
PDB_1	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
PDB_2	0.450	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
PDB_3	0.349	0.561	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
PDB_4	0.279	0.507	0.467	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
PDB_5	0.301	0.531	0.549	0.591	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
PRISK_1	-0.147	-0.011	-0.074	-0.010	0.037	-0.159	-0.129	-0.212	-0.197	-0.297	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
PRISK_2	-0.184	-0.168	-0.047	-0.129	-0.063	-0.195	-0.201	-0.203	-0.206	-0.281	0.485	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
PRISK_3	-0.156	-0.067	0.008	-0.058	-0.038	-0.173	-0.152	-0.156	-0.125	-0.259	0.523	0.668	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
PRISK_4inv	-0.489	-0.470	-0.415	-0.438	-0.384	-0.420	-0.392	-0.446	-0.444	-0.473	0.242	0.414	0.293	1.000	0.000	0.000	0.000	0.000	0.000	0.000
PRISK_5	-0.162	-0.093	-0.134	-0.157	-0.077	-0.261	-0.225	-0.243	-0.221	-0.311	0.434	0.422	0.442	0.389	1.000	0.000	0.000	0.000	0.000	0.000
TRUST_1	0.365	0.310	0.381	0.313	0.327	0.426	0.437	0.451	0.410	0.495	-0.276	-0.301	-0.227	-0.541	-0.309	-0.387	-0.293	-0.365	-0.431	-0.280
TRUST_2inv	0.242	0.239	0.303	0.298	0.336	0.350	0.307	0.410	0.369	0.446	-0.441	-0.421	-0.376	-0.403	-0.383	-0.539	-0.435	-0.493	-0.584	-0.492
TRUST_3inv	0.152	-0.021	0.068	0.154	0.096	0.227	0.180	0.328	0.275	0.379	-0.467	-0.280	-0.303	-0.230	-0.340	-0.363	-0.326	-0.377	-0.366	-0.321
TRUST_4	0.208	0.196	0.287	0.319	0.269	0.367	0.296	0.351	0.451	0.390	-0.057	-0.145	-0.054	-0.360	-0.257	-0.175	0.026	-0.089	-0.124	-0.030
TRUST_5	0.324	0.272	0.324	0.320	0.369	0.439	0.424	0.477	0.416	0.518	-0.167	-0.215	-0.191	-0.494	-0.247	-0.276	-0.157	-0.269	-0.245	-0.186
WTS_1	0.314	0.421	0.407	0.439	0.456	0.369	0.365	0.325	0.340	0.374	-0.057	-0.342	-0.181	-0.556	-0.222	-0.265	-0.111	-0.127	-0.193	-0.119
WTS_2	0.334	0.403	0.342	0.418	0.450	0.343	0.341	0.318	0.368	0.365	-0.111	-0.365	-0.211	-0.553	-0.229	-0.301	-0.146	-0.205	-0.238	-0.195
WTS_3inv	0.185	0.157	0.119	0.276	0.185	0.242	0.246	0.229	0.266	0.274	-0.148	-0.228	-0.248	-0.259	-0.180	-0.196	-0.177	-0.162	-0.202	-0.201
WTS_4	0.277	0.438	0.370	0.430	0.388	0.227	0.202	0.213	0.304	0.236	0.073	-0.294	-0.134	-0.446	-0.178	-0.231	-0.084	-0.147	-0.242	-0.161
WTS_5	0.244	0.350	0.302	0.341	0.356	0.196	0.260	0.240	0.294	0.223	0.027	-0.230	-0.041	-0.434	-0.132	-0.090	0.023	-0.040	-0.126	0.013

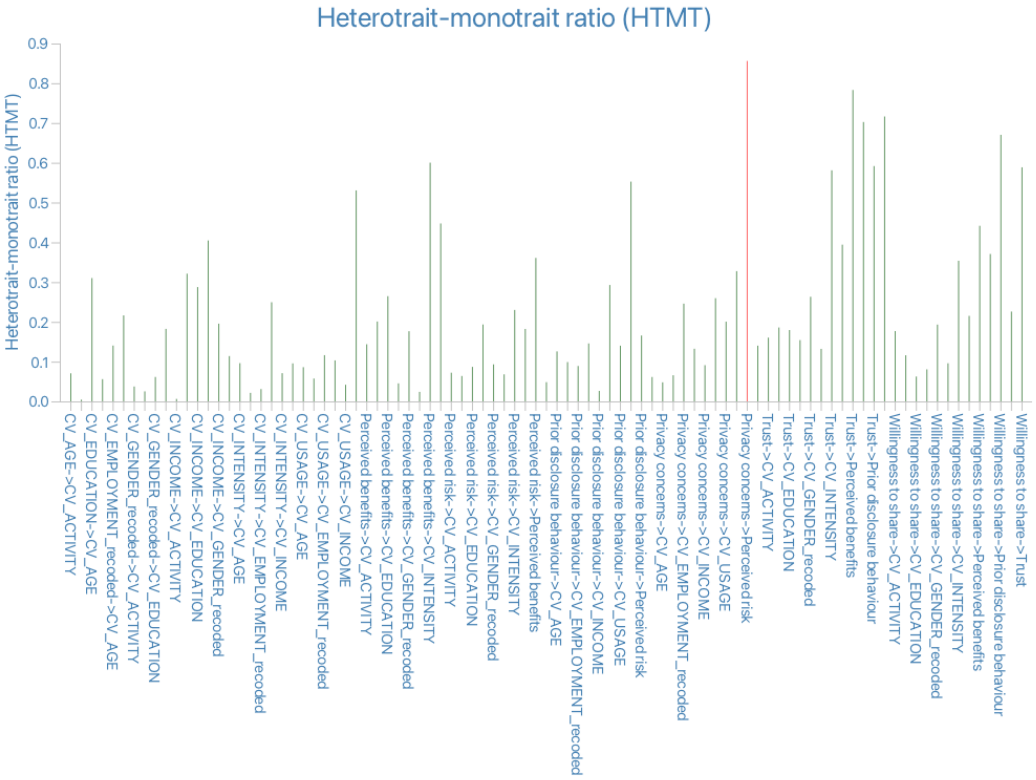
A-16a: Analysis of correlations between indicators (Part 1)

	TRUST_1	TRUST_2inv	TRUST_3inv	TRUST_4	TRUST_5	WTS_1	WTS_2	WTS_3inv	WTS_4	WTS_5	CV_USAGE	CV_INTENSITY	CV_ACTIVITY	CV_GENDER_recoded	CV_AGE	CV_EDUCATION	CV_EMPLOYMENT_recoded	CV_INCOME
CV_ACTIVITY	-0.124	-0.133	-0.059	-0.203	-0.066	-0.147	-0.167	-0.128	-0.148	-0.107	-0.096	-0.114	1.000	0.000	0.000	0.000	0.000	0.000
CV_AGE	-0.105	-0.194	-0.099	-0.015	-0.047	-0.066	-0.040	-0.164	0.085	0.184	-0.086	-0.096	0.071	0.026	1.000	0.000	0.000	0.000
CV_EDUCATION	-0.085	-0.173	-0.072	-0.100	-0.100	0.068	0.036	-0.029	0.025	0.073	-0.058	-0.022	-0.004	-0.062	0.312	1.000	0.000	0.000
CV_EMPLOYMENT_recoded	0.121	-0.100	-0.118	0.130	0.031	0.044	0.060	-0.128	0.007	0.150	-0.117	0.031	-0.056	0.185	0.141	0.217	1.000	0.000
CV_GENDER_recoded	0.286	0.119	-0.045	0.142	0.188	0.110	0.181	0.027	0.136	0.205	0.105	0.254	-0.038	1.000	0.000	0.000	0.000	0.000
CV_INCOME	0.057	-0.177	-0.077	0.069	0.018	0.015	0.002	-0.011	0.103	0.202	0.044	0.074	-0.006	0.206	0.336	0.300	0.420	1.000
CV_INTENSITY	0.468	0.229	0.264	0.359	0.426	0.319	0.326	0.231	0.254	0.238	0.531	1.000	0.000	0.000	0.000	0.000	0.000	0.000
CV_USAGE	0.261	0.191	0.230	0.151	0.258	0.259	0.220	0.261	0.099	0.114	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
PBENEFITS_1	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
PBENEFITS_2	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
PBENEFITS_3	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
PBENEFITS_4	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
PBENEFITS_5	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
PC_1	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
PC_2	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
PC_3	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
PC_4	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
PC_5	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
PDB_1	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
PDB_2	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
PDB_3	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
PDB_4	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
PDB_5	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
PRISK_1	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
PRISK_2	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
PRISK_3	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
PRISK_4inv	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
PRISK_5	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
TRUST_1	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
TRUST_2inv	0.364	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
TRUST_3inv	0.339	0.490	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
TRUST_4	0.409	0.140	0.061	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
TRUST_5	0.515	0.241	0.187	0.600	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
WTS_1	0.451	0.319	0.123	0.362	0.453	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
WTS_2	0.477	0.292	0.075	0.409	0.492	0.832	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
WTS_3inv	0.115	0.243	0.232	0.030	0.079	0.453	0.369	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
WTS_4	0.318	0.187	-0.053	0.432	0.333	0.533	0.627	0.203	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
WTS_5	0.354	0.109	-0.084	0.497	0.402	0.532	0.627	0.107	0.736	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

A-16b: Analysis of correlations between indicators (Part 2)



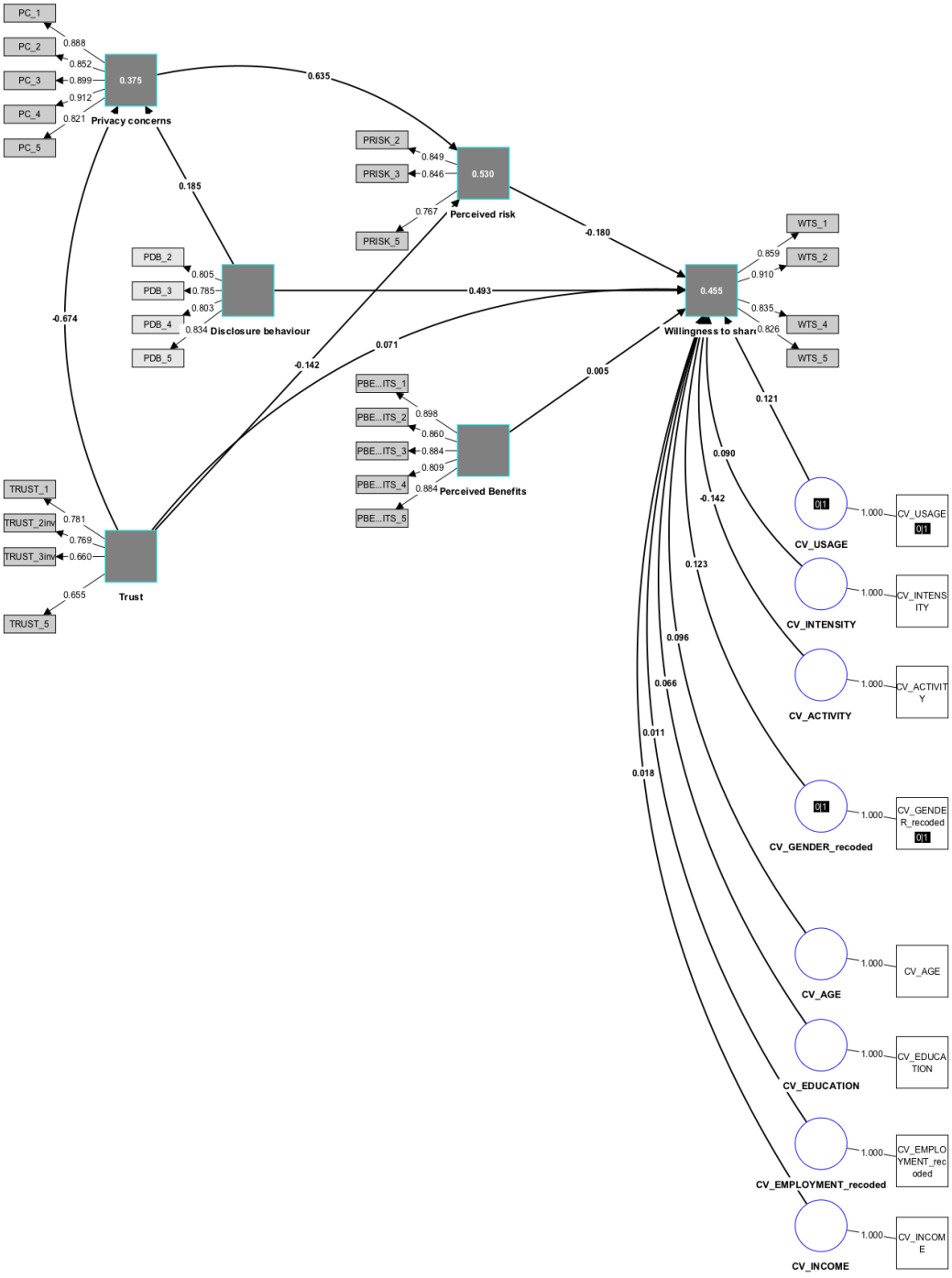
A-17: Model development in smartPLS (Note: constructs are depicted as a grey square with a blue frame; reflective items are depicted as light grey rectangles with black frames; control variables are depicted as white circles with a dark blue frame)



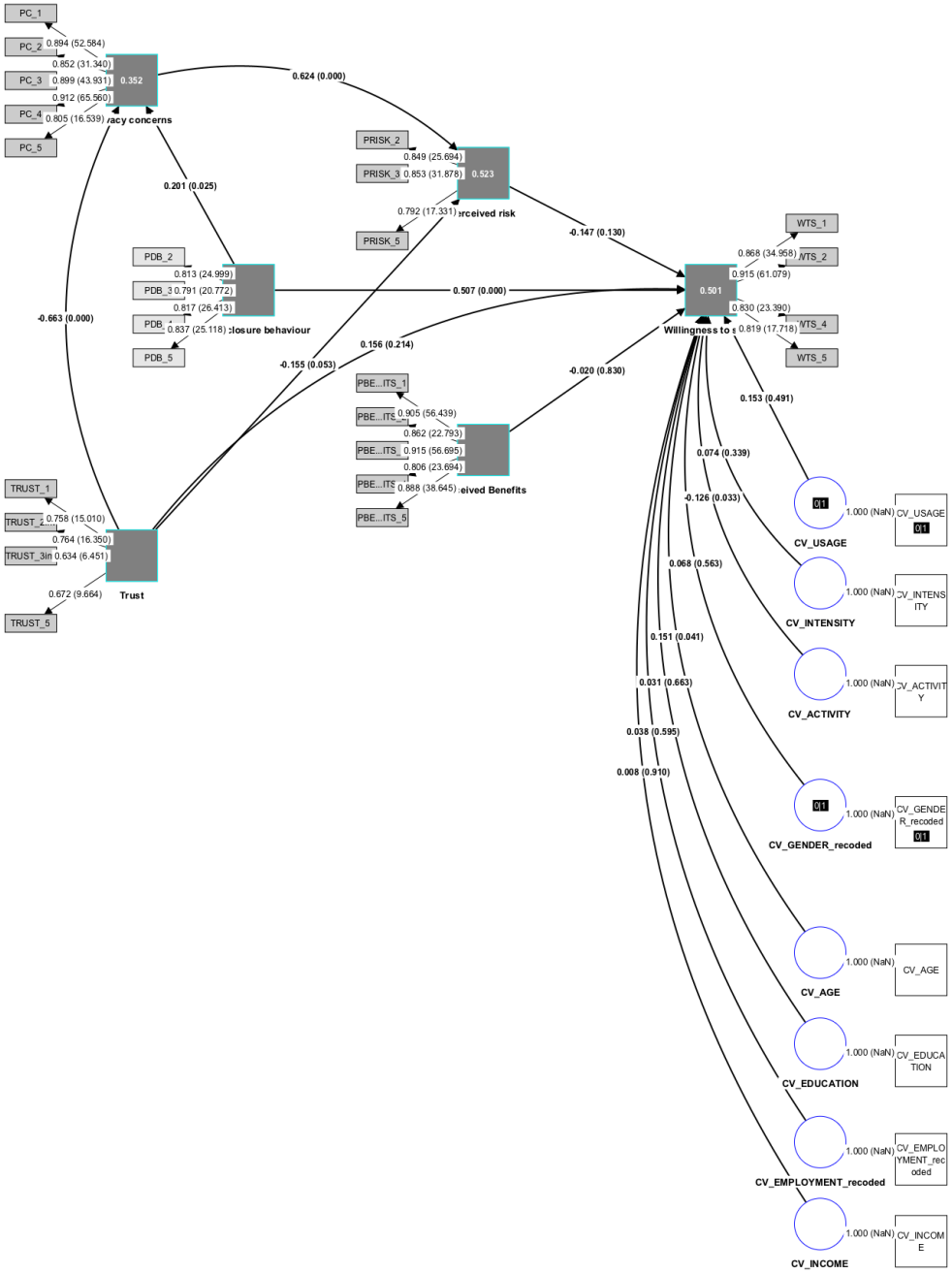
A-18: Heterotrait-montrait ratios

Constructs/Items	VIF
CV_ACTIVITY	1.000
CV_AGE	1.000
CV_EDUCATION	1.000
CV_EMPLOYMENT_recoded	1.000
CV_GENDER_recoded	1.000
CV_INCOME	1.000
CV_INTENSITY	1.000
CV_USAGE	1.000
PBENEFITS_1	3.995
PBENEFITS_2	3.129
PBENEFITS_3	3.322
PBENEFITS_4	1.913
PBENEFITS_5	3.310
PC_1	3.114
PC_2	2.619
PC_3	3.428
PC_4	3.774
PC_5	2.186
PDB_2	1.715
PDB_3	1.697
PDB_4	1.695
PDB_5	1.879
PRISK_2	1.872
PRISK_3	1.912
PRISK_5	1.290
TRUST_1	1.534
TRUST_2inv	1.403
TRUST_3inv	1.369
TRUST_5	1.367
WTS_1	3.250
WTS_2	4.003
WTS_4	2.416
WTS_5	2.420

A-19: VIF – Outer model



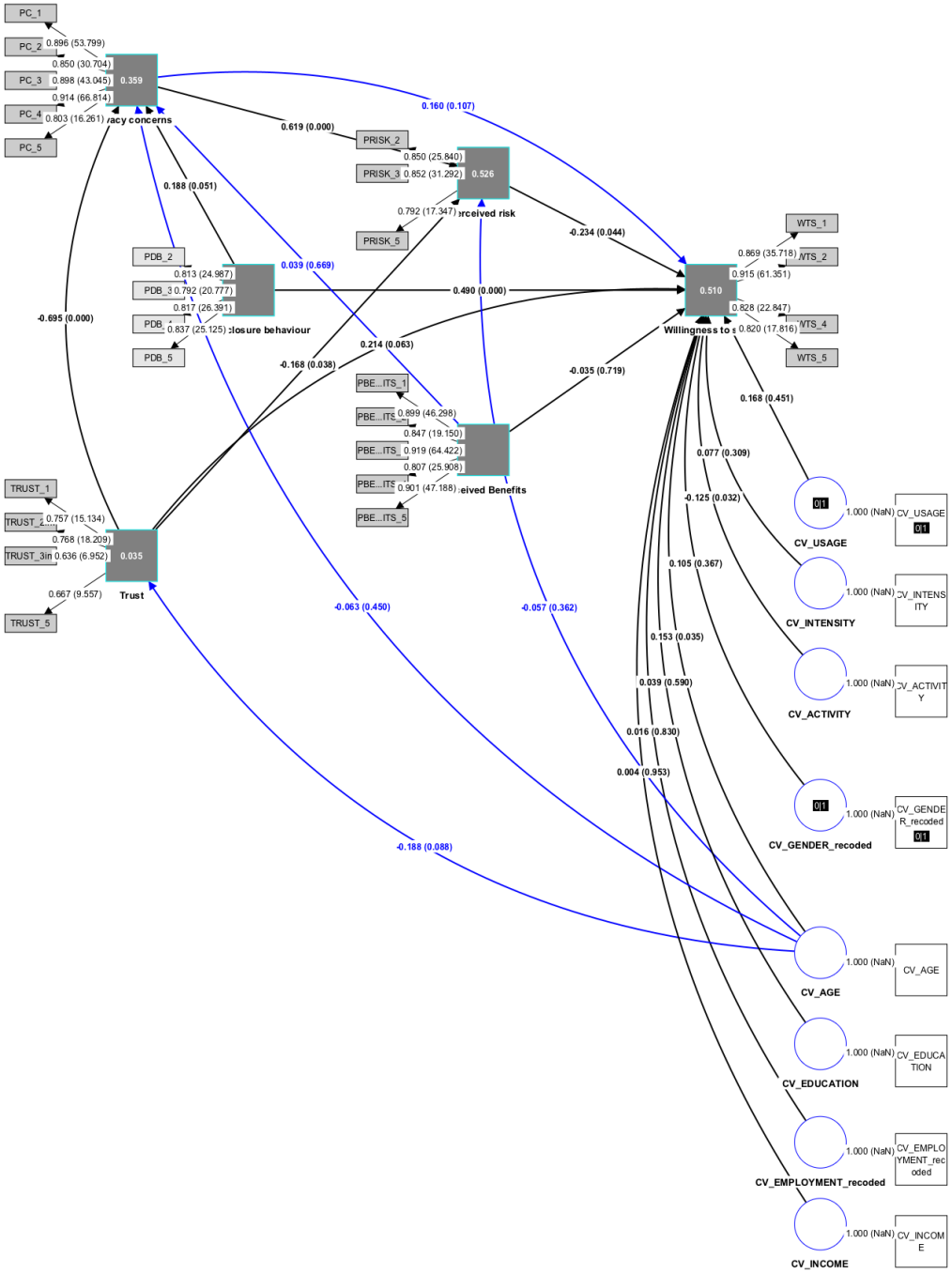
A-20: Measurement model (Note: The outer model shows outer loadings after removal of low-loaded indicators, the constructs' R-squared values and the inner model shows the path coefficients and p values for the non-bootstrapping calculation method)



A-21: Bootstrapping results in detailed version (Note: The outer model shows the outer weights/loadings and t values, the constructs' R-squared values and the inner model shows the path coefficients and p values.)

Constructs/Items	Original sample (O)	Sample mean (M)	Standard deviation (STDEV)	T statistics (|O/STDEV|)	P values
CV_ACTIVITY -> Willingness to share	-0.126	-0.119	0.059	2.137	0.033
CV_AGE -> Willingness to share	0.151	0.150	0.074	2.049	0.041
CV_EDUCATION -> Willingness to share	0.031	0.035	0.070	0.435	0.663
CV_EMPLOYMENT_recoded -> Willingness to share	0.038	0.045	0.071	0.531	0.595
CV_GENDER_recoded -> Willingness to share	0.068	0.060	0.118	0.578	0.563
CV_INCOME -> Willingness to share	0.008	0.005	0.067	0.113	0.910
CV_INTENSITY -> Willingness to share	0.074	0.074	0.077	0.956	0.339
CV_USAGE -> Willingness to share	0.153	0.138	0.223	0.689	0.491
Perceived benefits -> Willingness to share	-0.020	-0.016	0.092	0.215	0.830
Perceived risk -> Willingness to share	-0.147	-0.146	0.097	1.515	0.130
Prior disclosure behaviour -> Privacy concerns	0.201	0.200	0.090	2.245	0.025
Prior disclosure behaviour -> Willingness to share	0.507	0.497	0.079	6.431	0.000
Privacy concerns -> Perceived risk	0.624	0.623	0.085	7.352	0.000
Trust -> Perceived risk	-0.155	-0.159	0.080	1.935	0.053
Trust -> Privacy concerns	-0.663	-0.670	0.058	11.358	0.000
Trust -> Willingness to share	0.156	0.166	0.125	1.243	0.214

A-22: Path coefficients – Mean, STDEV, T values, p values (significance level of 0.05)



A-23: Post-hoc analysis (notes: same settings as in previous bootstrapping calculations; blue arrows signalling post-hoc)