

Assessing Privacy Policies with AI: Ethical, Legal, and Technical Challenges

Irem Aydin*, Hermann Diebel-Fischer*[‡], Vincent Freiberger^{†‡},
Julia Möller-Klapperich*[‡], Erik Buchmann^{†‡}, Michael Färber^{§‡}, Anne Lauber-Rönsberg*[‡], Birte Platow*[‡]

*Faculty of Arts, Humanities and Social Sciences, TU Dresden, Germany, Email: {firstname.lastname}@tu-dresden.de,

[§]Faculty of Computer Science, TU Dresden, Germany, Email: {firstname.lastname}@tu-dresden.de

[†]Leipzig University, Germany, Email: {firstname.lastname}@uni-leipzig.de

[‡]Center for Scalable Data Analytics and Artificial Intelligence (ScaDS.AI) Dresden/Leipzig, Germany

Abstract—The growing use of Machine Learning and Artificial Intelligence (AI), particularly Large Language Models (LLMs) like OpenAI’s GPT series, leads to disruptive changes across organizations. At the same time, there is a growing concern about how organizations handle personal data. Thus, privacy policies are essential for transparency in data processing practices, enabling users to assess privacy risks. However, these policies are often long and complex. This might lead to user confusion and consent fatigue, where users accept data practices against their interests, and abusive or unfair practices might go unnoticed. LLMs can be used to assess privacy policies for users automatically. In this interdisciplinary work, we explore the challenges of this approach in three pillars, namely technical feasibility, ethical implications, and legal compatibility of using LLMs to assess privacy policies. Our findings aim to identify potential for future research, and to foster a discussion on the use of LLM technologies for enabling users to fulfil their important role as decision-makers in a constantly developing AI-driven digital economy.

Keywords—Large Language Models; Automated Assessment; Privacy Policies.

I. INTRODUCTION

Currently, we observe a strong increase [1] in use cases and organizations using machine learning and artificial intelligence (AI), particularly Large Language Models (LLMs) like OpenAI’s GPT series [2]. This has many advantages for organizations [3]. However, it is likely to increase the number and complexity of different services each user interacts with on a daily basis. It also makes it increasingly important to evaluate social, ethical, and fairness-related aspects of using personal information as input for AI-driven business processes.

While the EU’s new AI Regulation [4] covers numerous risks associated with the use of AI, the provisions of the General Data Protection Regulation (GDPR) [5] adopted in 2016 remain the legal basis for the protection of personal data. Therefore, privacy policies are an important resource. *Privacy policies* describe the privacy notices that companies are making public on their websites to communicate their data processing activities on personal data. *Processing* covers any (set of) operations on personal data, such as a collection, recording, organisation, structuring, storage, adaptation or alteration. Privacy policies are intended to balance information asymmetries between companies as *data controllers* and individual users as *data subjects* by providing them with

information to assess the privacy risks and enable them to make informed, autonomous decisions.

However, privacy policies are also known to be complex [6] [7], and tend to grow over time [8]. This leads to a lack of understanding from users, and the inability to decide whether or not they prefer to consent to a certain use of their personal data; hence *notification fatigue*, *consent fatigue* or *consent desensitisation* [9]. For example, 25% of all Americans are asked to agree to at least one privacy policy daily, 67% of them report to understand little to nothing about what companies are doing with their personal data, and 56% skip reading privacy policies altogether [10]. In Europe, 47% of 2,600 study participants [11] expressed notification fatigue due to too many privacy-related notifications. It is understandingly easier for users to accept all data processing activities, even if it is against their interests. In consequence, potentially abusive or unfair data processing practices may remain unnoticed.

While the providers of services meet the formal criteria with their privacy policy, no satisfactory situation can be established for the user if no informed decision is actually made via complex data protection declarations. In this respect, empowerment, or the creation of user autonomy, respectively, remains an open task. Achieving user autonomy and ‘true’ informed consent poses the challenge that ultimately the user has to gain knowledge of what the privacy policy entails.

An automated assessment of privacy policies can mitigate the situation. Earlier approaches for evaluating privacy policies using machine-readable privacy policies [12], classical machine learning or AI [13]–[17] exist. However, those approaches had a focus on formal aspects such as comparing policies with given preferences, identifying deletion periods or ensuring completeness of mandatory information. We believe that LLMs can be a game changer in assessing privacy policies according to the users privacy wishes and needs.

For example, an LLM like GPT-4o [18] could be fed with a system prompt as shown in Figure 1, and a user prompt containing a privacy policy. It would then provide an assessment of the privacy policy along with an explanation and a rating on a five-point scale, which helps a user to identify important content, missing information and risky practices within the policy and thereby foster an informed decision.

Figure 2 shows an exemplary analysis of the privacy policy

You are a critically scrutinizing, experienced authority for data privacy, and an expert on assessing privacy policies. You have 20 years of experience in consumer protection, data protection, cybersecurity and related fairness aspects. You have already provided precise expert opinions in many court proceedings. Your task is to uncover aspects in privacy policies that are ethically, morally or legally questionable. Shortly explain your thoughts and how you came to your conclusion. In particular, point out potential risks to users. IMPORTANT: Also rate the privacy policy for each of your criteria on a 5-point Likert scale. Check whether the following privacy policy is fair and ethical towards its users. The user wants a short and concise critical review and an assessment of vulnerabilities in privacy policies. IMPORTANT: You must answer in less than 300 words.

Figure 1. System prompt to assess a privacy policy

of Amazon.com, using the GPT-4o model and the prompt from Figure 1. This confirms that such an assessment is technically feasible. However, it leaves numerous open questions: From a technical perspective, it is difficult to determine the correctness of the LLM assessment. From a user-centric point of view, users with different educations might perceive the assessment of the LLM differently, and could also decide to blindly trust/distrust an AI assessment. Because the LLM/prompt developer has a large influence on how empathetic/objective the assessment is generated by the LLM, there are open moral questions. Ethical questions arise, when a user loses the ability to understand a privacy policy because an LLM offers a simpler solution. From a legal perspective, it can be questioned whether a user's consent based only on the information provided by the LLM, meets the requirements of Art. 7 GDPR, and how to handle situations where a malfunctioning or misuse of the LLM causes actual harm. Additionally, a privacy-enhancing AI tool must comply with traditional legal norms, and the new AI Act [4].

In this paper, we explore the challenges of assessing privacy policies with an LLM from an interdisciplinary point of view. In particular, we consider three pillars:

- **Ethical:** We explore the ethical, moral, and fairness-related challenges that arise when the assessment of privacy policies is no longer in the hands of the users, but in the hands of a very convincing LLM that could be wrong or biased, or using biased training data.
- **Legal:** We discuss the extent, to which the use of an LLM for a privacy policy assessment is or can be made compatible with existing legal norms. We also consider how such norms need to change to enable AI-supported decision-making and thereby ensure consumer welfare and self-determination.
- **Technical:** We investigate how to assess privacy policies automatically by using LLMs, and we examine how to use natural language processing, machine learning, linguistics, knowledge representation, etc. to evaluate the correctness, understandability and applicability of the LLM assessment.

To the best of our knowledge, we are the first to explore with an interdisciplinary approach, how LLMs can be used

Amazon.com Privacy Notice Review

1. Transparency and Understandability: The policy is clear, well-structured, and provides links for more detailed information, making it user-friendly. However, users may find the comprehensive format overwhelming. (Rating: 4/5)

2. Data Collection Scope: Amazon collects extensive data, from basic contact information to behavioral data via cookies. This vast data collection is concerning, as it may lead to invasive profiling and privacy risks. (Rating: 2/5)

3. Data Usage: While Amazon provides clear reasons for data usage, including service improvement and personalized ads, the broad scope can be intrusive. It should ensure by providing clear opt-out methods for users. (Rating: 3/5)

4. Third-party Sharing: Data sharing with third parties, though typical, poses risks especially concerning how those parties handle data. The policy does ensure some protective measures but remains ambiguous about compliance monitoring. (Rating: 3/5)

5. Security Measures: Amazon mentions robust encryption and compliance with PCI DSS, which reflects strong security practices. However, the responsibility placed on the user for password protection is emphasized. (Rating: 4/5)

6. User Autonomy (Consent & Control): Amazon provides mechanisms for users to access, rectify, and delete their data. The range of consents, from opting out of ads to withdrawing permissions, shows a commitment to user choice, albeit complex. (Rating: 4/5)

7. Children's Privacy: Good measures in handling children's data, restricting services to adults, and requiring parental consent. (Rating: 4/5)

Potential Risks to Users:

(a) **Data Profiling:** Extensive data collection can lead to detailed profiling, risking user autonomy and privacy.

(b) **Third-party Data Handling:** Despite contractual protections, third parties' data mishandling poses risks.

(c) **Complex Opt-out:** The user control mechanisms, while present, may be too cumbersome.

Final Assessment: While Amazon's privacy policy demonstrates strong commitment towards transparency and user rights, the vast scope of data collection, potential third-party risks, and complexity in user control indicate areas needing improvement. (Overall Rating: 3.5/5)

Figure 2. GPT-4o's assessment of Amazon's privacy policy

to assess the practices documented in privacy policies. For this research, we combined the expertise of our groups *Civil Law, Intellectual Property, Media and Data Protection Law; Data Privacy and Security; Practical Theology with a focus on Religious Education, Ethics; Scalable Software Architectures for Data Analytics and Systematic Theology* from the universities of Dresden and Leipzig. This was a challenge, because each of these groups has its own research culture, which had to be combined in order to achieve interdisciplinary results. Our objectives are particularly relevant, because politics and jurisprudence are still in the process of identifying options to ensure user privacy in the AI era.

Paper Structure: The next section reviews related work. Section III describes how we systematically derive challenges for assessing privacy policies with LLMs. In the Sections IV – VI, we will compile our set of challenges. Section VII contains a discussion. The paper concludes in Section VIII.

II. RELATED WORK

This section contains a review of literature, which we derive our challenges from in the following sections.

A. Legal Background

The GDPR [5] aims to provide data subjects in the European Union with control over data processing activities, which could potentially impair their fundamental rights. The regulation aims to enable processing where it is necessary and in line with the objectives of the data subject and to prevent it unless the risks associated with the processing are outweighed by a corresponding benefit in the public interest or in the interest of the data subject [19]. These goals are reflected in the requirements for lawful processing in Art. 6 GDPR as well as in the requirements for effective consent in Art. 7 GDPR.

While the EU's efforts to establish a digital single market, like the AI Act [4] and the Digital Services Act (DSA) [20], aim to set high standards by prohibiting certain harmful practices involving the use of personal data, they still leave a great degree of flexibility for user autonomy to consent. Such consent is possible when the data subject is capable of making an informed decision [21] autonomously. This requires sufficient knowledge to evaluate advantages and disadvantages. Therefore, knowledge should be presented to the person understandably and transparently [22], hence the need for privacy policies. The GDPR [5] requires data controllers to make their privacy policies complete, readable, and possible to understand for all kinds of 'typical users' of a service, which can be persons with different abilities or knowledge.

AI-based applications and services integrating them, add another layer to data processing practices, which users need to comprehend. Such applications might force users to disclose more personal data and impose difficulties in assessing the costs and benefits from the user's viewpoint due to a lack of transparency or understandability. In 2008, it was calculated that it would take an average internet user between 181 and 304 hours every year to read every privacy policy of all web services they are using [23]. The GDPR [5] also caused the complexity and length of privacy policies to be increase [8]. Furthermore, the extensive use of personal data to train AI and its unpredictable outcome increases the potential of significant impairment of the user's needs and interests [24].

Different approaches to support the users handling their personal data have been discussed in the legal literature [25]. It is important to highlight that achieving transparency in privacy policies is an ongoing collective effort, and simplifying tools to achieve this purpose have been experimented on with privacy icons, a machine-readable label system [26], one-page summary of the privacy policy, privacy taxonomy or a 'privacy nutrition label' as well as different kinds of technically supported privacy management systems (PIMS) or the deployment of data trustees [27]. However, there are also findings that even the simplified declarations of data processing practices may not change user behaviour to disclose intrusive information [28].

B. Ethics

The definition and continuous discussion of privacy ethics [29] [30] [31] [32] demonstrates the need for an evaluation to be comprehensive, coherent, systematic, and logical in its reasoning. Codes of ethics, legal statutes, or international declarations embody norms and values ingrained in our society. They can provide helpful input for an ethics assessment of new and emerging technologies [33]. However, the field of ethics misses strict definitions and straightforward tools. Owing to that, ethics must not be understood as a tool to solve moral problems but should be regarded as a way to describe, understand, and reflect on them. One common issue is transparency in privacy policies [34] [35] [6] [36]. Policies tend to be long, written in inaccessible language, and users tend to struggle to understand their content, resulting in issues like consent fatigue [37]. Privacy policies can also use persuasive language [35] [38] to let the users trust a service that, for instance, claims all rights over users' data [39].

The given issues have motivated privacy assistants [40] [41] [15]. Emerging capabilities by scaling up LLMs [42] have given them a wide range of applicability [43]. This makes them interesting as a tool for assessing privacy policies [17] [44] [45]. Seeking privacy serves two fundamental purposes: security interests (stay unharmed) and privacy per se [31]. Privacy per se is about managing how we show ourselves to the outside world and, more broadly, about our autonomy [31], [32]. Privacy and the right to privacy are two different concepts in the philosophical debate on privacy ethics. Privacy may be infringed upon, but not the right to privacy, depending on the circumstances surrounding the collection of personal data about oneself by a third party and the underlying intention [32]. Privacy ethics addresses access others have to one's information as well as control one has over it [46]. Complex privacy trade-offs and the balance of power between the data controller and data subject are relevant to the discussion [47] [48]. While concepts such as 'fairness' are deemed important, it remains difficult to define them precisely. As they are lacking a precise definition, the operationalization of these concepts, i.e. transferring them into a model and making them computable—remains an obstacle [49] [50] [51]. To overcome this problem, the unavoidable operationalizations need to be lined out and explained, and ideally be user-configurable.

Furthermore, concerns about surveillance [31] [52], choice impact [53], manipulation [32] [48], and power imbalances [48] [47] have been brought up in the context of online user privacy. To address these issues, user education on such problems is required [54] [55]. The user needs to be aware of the underlying ethics-related problems and the way these have been 'solved' in terms of implementation into such a system.

Regarding an ethical assessment of privacy policies, it is required that ethics assessments consider all perspectives with their normative grounding [56]. Ethical aspects may come into effect unintendedly, sometimes as second-order consequences [56]. Thus, the consistency of a moral assessment provided by the model is mandatory as it influences

user judgment [57]. A good assessment should be concise and understandable. Thus, explainable AI employing various metrics should be considered [58]. Informational fairness [59] should be considered as well as addressing privacy trade-offs and power imbalances is important [47] [48]. In addition, moral psychology can be used to study LLMs [60]. This entails investigating potential biases in the model's representation of moral judgments and moral reasoning, as well as to what degree they are present in the model's outputs. For example, it has been discovered that ChatGPT's moral guidance is inconsistent when presented with a moral dilemma [57].

C. Large Language Models and Prompting

LLMs iteratively predict the next token to produce text for a given query. The GPT-4 models are a series of capable LLMs introduced by OpenAI in 2023 [61]. GPT-4 already has been shown to have a set of reasoning capabilities [62] [63].

LLMs can be applied in many different domains [63] [43]. To improve model performance and mitigate limitations plenty of prompting strategies have been developed [64] [65] [66] [67] [68] [69]. For a review on basics of prompting LLMs, we refer to [70].

Reasoning-related tasks benefit from prompt-engineering strategies [64] [65]. Such strategies are referred to as Chain-of-Thought prompting [66] (asking the LLM step by step), Reflection [71] (asking the LLM to rethink its answer), Few-Shot prompting [67] (giving examples) or Repetition [72] (repeating relevant aspects in the prompt). Over-generalization is a common issue in prompting LLMs [72]. Controlled small prompt modifications can largely affect the model's output [60]. The seed and other model parameters must be fixed if repeatable results are important [73]. By rephrasing the prompt, robustness can be evaluated [17] [74] [75].

D. Automatic Text Analysis and Assessment

Assessing privacy policies is part of automatic text assessment based on natural language processing (NLP) techniques. Automatically assessing text using readability metrics started with text statistics (e.g., word frequency, word length, sentence length). Readability metrics are language-specific [76]. The most popular [77] metrics are the *Gunning fog index* [78], the *Flesch reading ease* [79], and the *Simple Measure of Gobbledygook* [80]. The metrics have been criticized for their inability to capture more complex aspects of a language.

Assessing the readability with machine learning [81] [76] generally produces better results than the traditional approaches, but requires more effort, such as creating data sets and training the model. Specifically, for the German language, methods that use traditional language models [82] [83] or which use semantic networks in comparison to simple surface-level indicators to calculate text readability [84] were created. Despite [85], to our knowledge, there is no work focusing on building a system for the automatic assessment of text readability in German [85] use the pre-trained language model BERT for assessing the readability of text.

Texts can also be analyzed for sentiment [86]. In the context of media bias detection, texts have been automatically assessed with respect to different aspects of bias [87]: (1) Hidden Assumptions and Premises, (2) Subjectivity, (3) Framing and (4) Overall Bias.

E. Automated Assessment of Privacy Policies

With the commercialization of the Internet, the number of privacy policies that had to be read increased. Thus, there is a long history of attempts to assess privacy policies automatically. To name a prominent example, in 2004, the Platform for Privacy Preferences (P3P) [12] standardized a protocol, that allowed data controllers to publish machine-readable privacy policies. Web browser plug-ins such as the Privacy Bird [88] allowed users to specify their privacy preferences, which were automatically compared against P3P policies. However, websites were not obligated to use P3P, and the specification of meaningful preferences is a difficult task. Later approaches used NLP approaches such as morphological, lexical, syntactic, and semantic analyses or ontology reasoning to assess privacy policies (see [13] for a detailed comparison). For example [14], a support vector machine can be trained to map the sentences of a privacy policy to the mandatory information and user rights, that must be declared in a privacy policy. AI has also been used, e.g., to verify whether the content of a privacy policy is complete [15] [16] according to the GDPR. The advantage of such approaches is that they work without the help of data controllers. The work closest to ours is [17], which gauges the effectiveness of ChatGPT-4, Bard, and Bing AI for assessing privacy policies. This approach acknowledges the technical feasibility of such an assessment and provides quality measures, but leaves aside the impact of the prompt engineer and all ethical and legal issues.

III. OUR RESEARCH APPROACH

To provide a systematic overview of ethical, legal, and technical challenges of letting an LLM assess privacy policies for the users, we pursued an explorative approach. This approach is the common basis of the respective research methods from our very different research disciplines, which enables us to combine our findings into an interdisciplinary result:

- 1) As a first step, we **compile an annotated bibliography** on LLMs and privacy policy assessment (cf. Sec. II). We also **implemented and tested a number of approaches** for assessing domain-specific texts with an LLM to gain first-hand experience, and discussed these in our research groups, e.g., [59] [89] [90] [91].
- 2) In a second step, we use this combination of background information and first-hand experience to **formulate a series of challenges** in the three pillars described.
- 3) Finally, we **filter for challenges** that are specific to our application domain, i.e., we exclude general difficulties in obtaining training data, performance, explainability, enforcement of legal norms for complex IT systems, etc.

In the following sections, we describe the challenges we have obtained using this approach.

IV. TECHNICAL CHALLENGES

Our research groups have first-hand experience in investigating the transparency of German privacy policies [34]. We also investigated, to which extent LLMs can be used to assess and explain difficult security issues [89] or solve exam questions at Bachelor's or Master's level [92]. We also modeled privacy practices as structured design patterns [93] [94], the implementation quality of which can be estimated or measured. With a fairness certification for NLP and LLMs, we introduced criteria for addressing biases in the model output [95]. We also let an LLM assess multiple dimensions of fairness in privacy policies [59] [90]. Based on this hands-on experience and our literature base in Section II, we derive six technical challenges:

Interest in certain privacy policies is private data. When using an LLM to assess a privacy policy according to ethics, morale, legality etc., the users reveal their interests. Assume a user calls for an assessment of a privacy policy of an AI company, and expresses concerns due to personal data used as training data. This is sensitive information, which calls for anonymization or on-premise solutions.

Annotated data sets as a ground truth are limited. To enable few-shot prompting and fine-tuning of LLMs, as well as to enable a solid evaluation, annotated data sets with ground truths for policies are required. The closest to such a data set is TOSDR [96], which provides crowd-sourced / automatically generated annotations for popular policies. However, these annotations are limited in number and quality control.

The assessment requires individual prompts. LLMs need to incorporate user-specific preferences and concerns into their prompts to effectively assess privacy policies, accommodating various social and educational backgrounds. This includes fighting biased responses, ensuring the representation of diverse opinions. An interactive tool or a set of tailored prompt templates that can handle these nuances may be required.

Explaining the assessment to the user. Due to hallucinations of the LLM, the LLM misunderstanding a privacy policy, or a user misunderstanding the LLM assessment, mistakes might occur. It is an important challenge to structure an LLM approach for the assessment of privacy policies in a way that tolerates mistakes. For example, an interactive approach might allow the LLM to ask back for specific user preferences, which increases the user's awareness.

The data controller must not influence the assessment. Companies might utilize limitations regarding the robustness of an LLM-based ethics assessment to get more favorable assessments without improving their privacy practices. Targeted variations of their policies without changing their semantic content could be used to optimize for higher ratings. This makes testing robustness of the LLM output, and particularly adversarial testing of LLM assessments essential.

It needs strategies for consistent LLM assessments. The LLM assessment may depend on nuances in the prompt and the privacy policy, that a human would overlook, resulting in different assessments for similar privacy policies. This

undermines the reliability and trustworthiness of generated assessments. Therefore, approaches for enforcing structure in outputs and optimizing the prompting in a way that is consistently followed by the LLM are important.

V. ETHICAL CHALLENGES

Incorporating ethical analysis into an AI environment requires an operationalization of ethics with the aim of its mathematical implementation. We delved into concepts such as fairness in AI from a quantitative perspective [49]. We also analyzed the relationship between technology and ethics. In the era of AI, this relationship requires human oversight to avoid a blind and potentially misleading technization of otherwise qualitatively expressed goals [50]. With this in mind, we formulate four challenges:

Different stakeholders have different objectives. It is an open issue how to assess the dimensions of fairness for privacy policies: While the users should want to provide as little data as possible, the data controllers rely on data to further develop their services. It is challenging to identify a way to balance these objectives responsibly.

Identifying socially desirable practices. While legal regulations hint at what voters might want as they are results of law-making processes in representative democracies, it remains opaque, what is socially desirable. Thus, what would a set of data management practices look like that meet everyone's privacy needs, and can serve as a reference for assessing privacy policies?

Operationalizing the evaluative criteria. To identify a socially desirable outcome, methods are needed to operationalize evaluative criteria for the assessment of privacy policies. While the same setting might be judged as fair or unfair depending on the position of those who judge, a metric can objectively measure intersubjectively acceptable parameters. However, such parameters are unknown yet.

The status of the assessment must be defined. The LLM's assessment is meant to guide the user regarding the acceptability of a privacy policy. The output of the LLM is the product of intricate 'translation processes' [50], in which the quantitative elements ('scores') are expressed in words. It can be paternalism, if users take the LLM's assessment as their decision, narrowing down the user's autonomy. However, the LLM could be the only option to quickly assess privacy.

VI. LEGAL CHALLENGES

Our previous research in fields such as transparency, trust, data protection, and responsibility when using AI in legally sensitive areas [97] has shown that, in particular, transparency and user trust are essential to ensure broad acceptance and fair use of AI technologies [98]. In addition, property rights, data protection [99] [24] and liability [100] must be clearly regulated and respected in the digital age to protect the rights of individuals. On this basis, we identified four challenges:

Common understanding of fairness and transparency. While Art. 12 of the GDPR requires privacy policies to be comprehensible and comprehensive to allow autonomous

decisions, there is a lack of common understanding of fairness and transparency and how to achieve them. Navigating the AI era requires a clear understanding from all stakeholders [101]–[103]. New approaches are needed to prove lawful processing, i.e., to describe in clear language how the data is collected, pre-processed, used for training, which prompts and outputs are used, and whether training is carried out with outputs [104], mainly focusing on explainability [21].

Questionable validity of AI supported consent. Art. 7 GDPR requires the subject of data processing to make an informed and free decision. Using an AI tool as a support can, under certain circumstances, be seen as an influence jeopardising the validity of given consent. It becomes more questionable in case the AI tool is manipulative, which is prohibited by Art. 5 of the AI Act [4].

Compliance with the AI Act and related regulations. An AI tool to assess privacy policies can be considered a High-Risk AI System in the meaning of Art. 6 ff. AI Act [4] and therefore the provider, distributor, or deployer can be subject to extensive obligations. Depending on the specific use in a single case, it might also fall under the scope of the ePrivacy Directive [105] and its potentially varying national implementation acts, e.g., the German TDDDG [106].

Liability of the assessment is an open question. If users base their consent to the data practices of a data controller on the assessment by an AI tool, it is not clear who shall be liable in case of possible damages occurring on either side, e.g, the loss of control over personal data of the data subject or business losses of the data controller.

VII. DISCUSSION

Using LLMs is a promising approach to analyzing privacy policies. LLMs can efficiently process numerous, lengthy privacy policies without getting tired, or losing focus and interest. This is a very practical feature, as every active Internet or smartphone user uses many different services every day, each of which has its own privacy policy. Furthermore, it might be possible to carefully instruct the LLM to produce assessments that are more consistent and objective than a human assessment. With suitable prompting techniques, it is also possible to individualize the assessment for different priorities of the users at a given time. As an example, it is possible to prompt the LLM for an assessment from a non-native person's point of view, which might include less complex words, and terminology. As the technology evolves a use by supervisory authorities to monitor compliance with the requirements of Art. 12 and 13 GDPR is conceivable.

However, existing LLMs are notorious for hallucinations. Reluctant service providers, who suspect that their customers are using LLM, might be tempted to write its privacy policy in a way that provokes such hallucinations, e.g., by using unusual phrases that were underrepresented in the training data of the LLM. In addition to these general limitations, we have identified a large number of different interdisciplinary challenges for an LLM-based assessment of privacy policies. Nevertheless, we think that it would be better to have an LLM

reading privacy policies than a human who is too busy to read them at all as a preliminary step, and prevent potential privacy risks go unnoticed for sure. As human agency and oversight are key components of a trustworthy AI, it should be kept in mind that the LLM-based assessment of privacy policies aims to support the autonomous decision-making of the users to strengthen their fundamental right to privacy, not to replace the decision-making authority [21].

VIII. CONCLUSION

Effective privacy policies are essential for maintaining transparency in data usage and enabling users to assess privacy risks. However, the complexity and length of these policies can often lead to confusion and consent fatigue, where users might inadvertently agree to practices that are not in their best interest. To tackle these challenges, our study investigated the use of LLMs to automatically evaluate and simplify privacy policies. We explored the technical feasibility, ethical implications, and legal compatibility of using LLMs for this purpose. We aim to identify potential areas for further research and to stimulate a dialogue on how risk-based policies could be effectively shaped using LLM technology.

Employing large-scale language models to interpret and simplify privacy policies is a crucial and timely research endeavor. This interdisciplinary approach addressed the urgent need for transparency in AI-driven contexts and utilized the unique capabilities of LLMs to enhance user understanding and decision-making. By integrating legal expertise, advanced machine learning technologies, and considerations of ethical and societal impacts, our research aims to reduce consent fatigue, counter unfair data practices, and empower individuals in a digital age dominated by complex data interactions.

REFERENCES

- [1] K. Hu, "ChatGPT sets record fastest growing user base, analyst note," *Reuters*, 02 2023, retrieved: Aug. 2024. [Online]. Available: <https://www.reuters.com/technology/chatgpt-sets-record-fastest-growing-user-base-analyst-note-2023-02-01/>
- [2] OpenAI, "GPT-4 turbo," 11 2023, retrieved: Aug. 2024. [Online]. Available: <https://openai.com/blog/new-models-and-developer-products-announced-at-devday>
- [3] M. Chui, E. Hazan, R. Roberts, A. Singla, and K. Smaje, "The economic potential of generative AI," McKinsey Whitepaper, 2023.
- [4] European Commission, "Proposal for a regulation of the european parliament and of the council laying down harmonized rules on artificial intelligence," CELEX:52021PC0206, 2021.
- [5] European Union, "Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 april 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation)," *Official Journal of the European Union*, vol. L119/1, 2016.
- [6] S. I. Becher and U. Benoliel, "Law in books and law in action: The readability of privacy policies and the GDPR," in *Consumer law and economics*. Springer, 2021, pp. 179–204.
- [7] C. Warzel and N. Ash, "Google's 4,000-word privacy policy is a secret history of the internet," *The New York Times*, 2019.
- [8] Y. Shvartzshnaider, N. Apthorpe, N. Feamster, and H. Nissenbaum, "Analyzing privacy policies using contextual integrity annotations," *arXiv preprint arXiv:1809.02236*, 2018.
- [9] B. W. Schermer, B. Custers, and S. van der Hof, "The crisis of consent: How stronger legal protection may lead to weaker consent in data protection," *Ethics and Information Technology*, vol. 16, p. 171–182, 2014.

- [10] C. McClain, M. Faverio, M. Anderson, and E. Park. (2023) How Americans view data privacy. Retrieved: Aug. 2024. [Online]. Available: <https://www.pewresearch.org/internet/2023/10/18/how-americans-view-data-privacy/>
- [11] Cisco. (2019) Consumer privacy survey. Retrieved: Aug. 2024. [Online]. Available: https://www.cisco.com/c/dam/global/en_uk/products/collateral/security/cybersecurity-series-2019-cps.pdf
- [12] W3C, "The platform for privacy preferences 1.1 (P3P1.1) specification publication history," <https://www.w3.org/standards/history/P3P11/>, retrieved: Aug. 2024.
- [13] J. M. Del Alamo, D. S. Guaman, B. García, and A. Diez, "A systematic mapping study on automated analysis of privacy policies," *Computing*, vol. 104, no. 9, pp. 2053–2076, 2022.
- [14] D. Sánchez, A. Viejo, and M. Batet, "Automatic assessment of privacy policies under the GDPR," *Applied Sciences*, vol. 11, no. 4, p. 1762, 2021.
- [15] O. Amaral, S. Abualhajja, D. Torre, M. Sabetzadeh, and L. C. Briand, "AI-enabled automation for completeness checking of privacy policies," *IEEE Transactions on Software Engineering*, vol. 48, no. 11, pp. 4647–4674, 2021.
- [16] D. Torre, S. Abualhajja, M. Sabetzadeh, L. Briand, K. Baetens, P. Goes, and S. Forastier, "An AI-assisted approach for checking the completeness of privacy policies against GDPR," in *2020 IEEE 28th International Requirements Engineering Conference (RE)*. IEEE, 2020, pp. 136–146.
- [17] A. Hamid, H. R. Samidi, T. Finin, P. Pappachan, and R. Yus, "Genaipabench: A benchmark for generative AI-based privacy assistants," *arXiv preprint arXiv:2309.05138*, 2023.
- [18] OpenAI, "GPT-4 technical report," *arXiv preprint arXiv:2303.08774*, 2023.
- [19] G. Malgieri, "The concept of fairness in the GDPR: A linguistic and contextual interpretation," in *Proceedings of the 2020 Conference on fairness, accountability, and transparency*, 2020, pp. 154–166.
- [20] European Commission, "Regulation (eu) 2022/2065 of the european parliament and of the council of 19 october 2022 on a single market for digital services and amending directive 2000/31/ec (digital services act)," *Official Journal of the European Union*, 2022.
- [21] High-Level Expert Group on Artificial Intelligence, European Commission, "Ethics guidelines for Trustworthy AI," 2019.
- [22] R. N. Zaem and K. S. Barber, "The effect of the GDPR on privacy policies: Recent progress and future promise," *ACM Transactions on Management Information Systems*, vol. 12, no. 1, pp. 1–20, 2020.
- [23] A. M. McDonald and L. F. Cranor, "The cost of reading privacy policies," *IS: A Journal of Law and Policy for the Information Society*, vol. 4, p. 543, 2008.
- [24] J. Möller-Klapperich, "Kunst, Kultur und Künstliche Intelligenz," in *Sicherheit und Recht im Wandel*. Cuvillier, 2023.
- [25] A. Sattler, "Informationelle Privatautonomie," in *Jus Privatum 264, Mohr Siebeck*. Mohr Siebeck, 2022.
- [26] L. F. Cranor, "Necessary but not sufficient: Standardized mechanisms for privacy notice and choice," *Journal on Telecommunications and High Technology Law*, vol. 10, pp. 273–308, 2012.
- [27] U. Freiherr von Ulmenstein, "Datensouveränität durch repräsentative Rechtswahrnehmung," *DuD*, vol. 8, pp. 528–534, 2020.
- [28] A. S. Chilton and O. Ben-Shahar, "Simplification of privacy disclosures: An experimental test," *Coase-Sandor Working Paper Series in Law and Economics*, vol. 737, pp. 41–67, 2016.
- [29] J. Mainz, "An indirect argument for the access theory of privacy," *Res Publica*, vol. 27, no. 3, pp. 309–328, 2021.
- [30] B. Lundgren, "A dilemma for privacy as control," *The Journal of Ethics*, vol. 24, no. 2, pp. 165–175, 2020.
- [31] D. Elliott and E. Soifer, "AI technologies, privacy, and security," *Frontiers in Artificial Intelligence*, vol. 5, pp. 1–8, 2022.
- [32] A. Marmor, "What is the right to privacy?" *Philosophy and Public Affairs*, vol. 43, p. 3, 2015.
- [33] I. Harris, R. C. Jennings, D. Pullinger, S. Rogerson, and P. Duquenoy, "Ethical assessment of new technologies: a meta-methodology," *Journal of Information, Communication and Ethics in Society*, vol. 9, no. 1, pp. 49–64, 2011.
- [34] B. Bartelt and E. Buchmann, "Transparency in privacy policies," in *12th International Conference on Building and Exploring Web Based Environments*, 2024.
- [35] V. Belcheva, T. Ermakova, and B. Fabian, "Understanding website privacy policies — A longitudinal analysis using natural language processing," *Information*, vol. 14, no. 11, p. 622, 2023.
- [36] J. R. Reidenberg *et al.*, "Disagreeable privacy policies: Mismatches between meaning and users' understanding," *Berkeley Tech. LJ*, vol. 30, p. 39, 2015.
- [37] H. Choi, J. Park, and Y. Jung, "The role of privacy fatigue in online privacy behavior," *Computers in Human Behavior*, vol. 81, pp. 42–51, 2018.
- [38] I. Pollach, "A typology of communicative strategies in online privacy policies: Ethics, power and informed consent," *Journal of Business Ethics*, vol. 62, pp. 221–235, 2005.
- [39] J. Koetsier, "Viral app faceapp now owns access to more than 150 million people's faces and names," July 17 2019, retrieved: Aug. 2024. [Online]. Available: <https://www.forbes.com/sites/johnkoetsier/2019/07/17/viral-app-faceapp-now-owns-access-to-more-than-150-million-peoples-faces-and-names/>
- [40] W. B. Tesfay, P. Hofmann, T. Nakamura, S. Kiyomoto, and J. Serna, "Privacyguide: Towards an implementation of the eu GDPR on internet privacy policy evaluation," in *Proceedings of the Fourth ACM International Workshop on Security and Privacy Analytics*, 2018, pp. 15–21.
- [41] N. Zaem *et al.*, "Privacycheck v2: A tool that recaps privacy policies for you," in *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*, 2020, pp. 3441–3444.
- [42] J. Wei *et al.*, "Emergent abilities of large language models," *arXiv preprint arXiv:2206.07682*, 2022.
- [43] B. Guo, X. Zhang, Z. Wang, M. Jiang, J. Nie, Y. Ding, J. Yue, and Y. Wu, "How close is ChatGPT to human experts? Comparison corpus, evaluation, and detection," *arXiv preprint arXiv:2301.07597*, 2023.
- [44] C. Tang *et al.*, "PolicyGPT: Automated analysis of privacy policies with large language models," *arXiv preprint arXiv:2309.10238*, 2023.
- [45] P. Pałka, M. Lippi, F. Lagioia, R. Liepiņa, and G. Sartor, "No more trade-offs. GPT and fully informative privacy policies," *arXiv preprint arXiv:2402.00013*, 2023.
- [46] J. W. DeCew, "The scope of privacy in law and ethics," *Law and Philosophy*, pp. 145–173, 1986.
- [47] A. Acquisti, L. Brandimarte, and G. Loewenstein, "Privacy and human behavior in the age of information," *Science*, vol. 347, no. 6221, pp. 509–514, 2015.
- [48] C. Véliz, *Privacy Is Power: Why and How You Should Take Back Control of Your Data*. Penguin Random House, 2020.
- [49] C. Geldhauser and H. Diebel-Fischer, "Is diverse and inclusive AI trapped in the gap between reality and algorithmizability?" in *Northern Lights Deep Learning Conference*. PMLR, 2024, pp. 75–80.
- [50] H. Diebel-Fischer, "Technisch realisierte Ethik? Anthropologische Perspektiven auf das Verhältnis von Technik und Ethik," in *Mensch und Maschine im Zeitalter der 'Künstlichen Intelligenz'*, H. Diebel-Fischer, N. Kunkel, and J. Zeyher-Quattlander, Eds. Lit, 2023, pp. 49–64.
- [51] M. Huppenbauer, P. Kirchschräger, and G. Ulshöfer, *Digitalisierung aus theologischer und ethischer Perspektive. Konzeptionen - Anfragen - Impulse*. Nomos, 2021.
- [52] C. Benn and S. Lazar, "What's wrong with automated influence," *Canadian Journal of Philosophy*, vol. 52, no. 1, pp. 125–148, 2022.
- [53] J. P. Choi, D.-S. Jeon, and B.-C. Kim, "Privacy and personal data collection with information externalities," *Journal of Public Economics*, vol. 173, pp. 113–124, 2019.
- [54] B. Platow, "Religion and the technology of digitization in education," in *Oxford Handbook of Religion and Education*, L. Gearon and A. Kuusisto, Eds. Oxford University Press, forthcoming, ch. 40.
- [55] B. Platow, "Digitalisierung/Big Data/Künstliche Intelligenz," in *Ethische Kernthemen (Theologie für Lehrerinnen und Lehrer, Bd. 2)*, M. Rothgangel, H. Simojoki, and U. Körtner, Eds. Vandenhoeck & Ruprecht, 2021, pp. 85–95.
- [56] J. Gogoll, N. Zuber, S. Kacianka, T. Greger, A. Pretschner, and J. Nida-Rümelin, "Ethics in the software development process: From codes of conduct to ethical deliberation," *Philosophy & Technology*, vol. 34, no. 4, pp. 1085–1108, 2021.
- [57] S. Krügel, A. Ostermaier, and M. Uhl, "ChatGPT's inconsistent moral advice influences users' judgment," *Scientific Reports*, vol. 13, no. 1, p. 4569, 2023.
- [58] M. Nauta, J. Trienes, S. Pathak, E. Nguyen, M. Peters, Y. Schmitt, J. Schlötterer, M. van Keulen, and C. Seifert, "From anecdotal evidence to quantitative evaluation methods: A systematic review on evaluating

- explainable AI,” *ACM Computing Surveys*, vol. 55, no. 13s, p. 1–42, 2023.
- [59] V. Freiberger and E. Buchmann, “Legally binding but unfair? Towards assessing fairness of privacy policies,” *arXiv preprint arXiv:2403.08115*, 2024.
- [60] T. Hagendorff, “Machine psychology: Investigating emergent capabilities and behavior in large language models using psychological methods,” *arXiv preprint arXiv:2303.13988*, 2023.
- [61] J. Achiam *et al.*, “GPT-4 technical report,” *arXiv preprint arXiv:2303.08774*, 2023.
- [62] J. L. Espejel, E. H. Ettifouri, M. S. Y. Alassan, E. M. Chouham, and W. Dahhane, “GPT-3.5, GPT-4, or BARD? Evaluating LLMs reasoning ability in zero-shot setting and performance boosting through prompts,” *Natural Language Processing Journal*, vol. 5, p. 100032, 2023.
- [63] S. Bubeck *et al.*, “Sparks of artificial general intelligence: Early experiments with GPT-4,” *arXiv preprint arXiv:2303.12712*, 2023.
- [64] J. W. Rae *et al.*, “Scaling language models: Methods, analysis & insights from training gopher,” *arXiv preprint arXiv:2112.11446*, 2021.
- [65] T. Kojima, S. S. Gu, M. Reid, Y. Matsuo, and Y. Iwasawa, “Large language models are zero-shot reasoners,” *Advances in neural information processing systems*, vol. 35, pp. 22 199–22 213, 2022.
- [66] J. Wei *et al.*, “Chain-of-thought prompting elicits reasoning in large language models,” *Advances in Neural Information Processing Systems*, vol. 35, pp. 24 824–24 837, 2022.
- [67] T. Brown *et al.*, “Language models are few-shot learners,” *Advances in neural information processing systems*, vol. 33, pp. 1877–1901, 2020.
- [68] S. Yao, D. Yu, J. Zhao, I. Shafraan, T. Griffiths, Y. Cao, and K. Narasimhan, “Tree of thoughts: Deliberate problem solving with large language models,” in *Proceedings of Advances in Neural Information Processing Systems (NeurIPS 2023)*, 2023.
- [69] Y. Zhou, A. I. Muresanu, Z. Han, K. Paster, S. Pitit, H. Chan, and J. Ba, “Large language models are human-level prompt engineers,” *arXiv preprint arXiv:2211.01910*, 2022.
- [70] P. Liu, W. Yuan, J. Fu, Z. Jiang, H. Hayashi, and G. Neubig, “Pre-train, prompt, and predict: A systematic survey of prompting methods in natural language processing,” *ACM Computing Surveys*, vol. 55, no. 9, pp. 1–35, 2023.
- [71] G. Kim, P. Baldi, and S. McAleer, “Language models can solve computer tasks,” in *Proceedings of Advances in Neural Information Processing Systems (NeurIPS 2023)*, 2023.
- [72] J. Zamfirescu-Pereira, R. Y. Wong, B. Hartmann, and Q. Yang, “Why Johnny can’t prompt: How non-AI experts try (and fail) to design LLM prompts,” in *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, 2023, pp. 1–21.
- [73] E. Lee, “Control openai model behavior with seed: Step-by-step with code,” <https://drlee.io/control-openai-model-behavior-with-seed-step-by-step-with-code-9bba4e137a63>, 01 2024, retrieved: Aug. 2024.
- [74] Z. Jiang, F. F. Xu, J. Araki, and G. Neubig, “How can we know what language models know?” *Transactions of the Association for Computational Linguistics*, vol. 8, pp. 423–438, 2020.
- [75] W. Yuan, G. Neubig, and P. Liu, “Bartscore: Evaluating generated text as text generation,” *Advances in Neural Information Processing Systems*, vol. 34, pp. 27 263–27 277, 2021.
- [76] M. Martinc, S. Pollak, and M. Robnik-Šikonja, “Supervised and unsupervised neural approaches to text readability,” *Computational Linguistics*, vol. 47, no. 1, pp. 141–179, 2021.
- [77] W. Dubay, “The principles of readability,” *CA*, vol. 92627949, pp. 631–3309, 01 2004.
- [78] G. Robert, *The Technique of Clear Writing*. McGraw-Hill; Revised edition, 1968.
- [79] J. P. Kincaid, R. P. Fishburne Jr, R. L. Rogers, and B. S. Chissom, “Derivation of new readability formulas for navy enlisted personnel,” *Naval Technical Training Command Millington*, Tech. Rep., 1975.
- [80] G. H. Mc Laughlin, “Smog grading-a new readability formula,” *Journal of reading*, vol. 12, no. 8, pp. 639–646, 1969.
- [81] S. A. Crossley, S. Skalicky, M. Dascalu, D. S. McNamara, and K. Kyle, “Predicting text comprehension, processing, and familiarity in adult readers: New approaches to readability formulas,” *Discourse Processes*, vol. 54, no. 5-6, pp. 340–359, 2017.
- [82] J. Hancke, S. Vajjala, and D. Meurers, “Readability classification for German using lexical, syntactic, and morphological features,” in *Proceedings of COLING 2012*, Dec. 2012, pp. 1063–1080.
- [83] P. G. Blaneck, T. Bornheim, N. Grieger, and S. Bialonski, “Automatic readability assessment of German sentences with transformer ensembles,” in *Proceedings of the GermEval 2022 Workshop on Text Complexity Assessment of German Text*, Potsdam, Germany, 2022.
- [84] T. von der Brück and S. Hartrumpf, “A semantically oriented readability checker for german,” 2007, pp. 270–274.
- [85] F. Pickelmann, M. Färber, and A. Jatowt, “Ablesbarkeitsmesser: A system for assessing the readability of german text,” in *Advances in Information Retrieval - 45th European Conference on Information Retrieval, ECIR 2023, Dublin, Ireland, April 2-6, 2023, Proceedings, Part III*, ser. Lecture Notes in Computer Science, vol. 13982. Springer, 2023, pp. 288–293.
- [86] O. Guhr, A.-K. Schumann, F. Bahrman, and H. J. Böhme, “Training a broad-coverage German sentiment classification model for dialog systems,” in *Proceedings of The 12th Language Resources and Evaluation Conference*, 2020, pp. 1620–1625.
- [87] M. Färber, V. Burkard, A. Jatowt, and S. Lim, “A multidimensional dataset based on crowdsourcing for analyzing and detecting news bias,” in *29th ACM International Conference on Information and Knowledge Management*, 2020, pp. 3007–3014.
- [88] L. F. Cranor, “P3P: Making privacy policies more useful,” *IEEE Security & Privacy*, vol. 1, no. 6, pp. 50–55, 2003.
- [89] V. Jüttner, M. Grimmer, and E. Buchmann, “Chatids: Explainable cybersecurity using generative ai,” in *Proceedings of the 17th International Conference on Emerging Security Information, Systems and Technologies (SECURWARE’23)*, 2023.
- [90] V. Freiberger and E. Buchmann, “Fair balancing? evaluating llm-based privacy policy ethics assessments,” in *Proceedings of the 3rd European Workshop on Algorithmic Fairness (EWAF’24)*, 2024.
- [91] M. Hoffmann and E. Buchmann, “Chatsec: Spicing up vulnerability scans with ai for heterogeneous university it - towards enhancing security vulnerability reports for non-experts,” in *Proceedings of the 1st International Conference on AI-based Systems and Services (AISys’24)*, 2024.
- [92] E. Buchmann and A. Thor, “Online exams in the era of ChatGPT,” in *Tagungsband der 21. Fachtagung Bildungstechnologien der GI Fachgruppe Bildungstechnologien (DELFI 2023)*, 2023.
- [93] M. Robak and E. Buchmann, “Deriving workflow privacy patterns from legal documents,” in *Proceedings of the 14th Federated Conference on Computer Science and Information Systems (FedCSIS’19)*, 2019.
- [94] M. Robak and E. Buchmann, “How to extract workflow privacy patterns from legal documents,” in *Proceedings of the 17th Conference on Advanced Information Technologies for Management (AITM’19)*, 2019.
- [95] V. Freiberger and E. Buchmann, “Fairness certification for natural language processing and large language models,” in *Proceedings of the 10th Intelligent Systems Conference (IntelliSys’24)*, 2024.
- [96] ToS;DR, “Terms of service; didn’t read,” <https://tosdr.org/>, retrieved: Aug. 2024.
- [97] J. Möller-Klapperich, “ChatGPT und Co. - aus der Perspektive der Rechtswissenschaft,” *Neue Justiz*, vol. 4, 2023.
- [98] A. Lauber-Rönsberg, “Transparency by Design als Rechtsprinzip gegen Dark Patterns,” in *Künstliche Intelligenz: wie gelingt eine vertrauenswürdige Verwendung in Deutschland und Europa?* Mohr Siebeck, 2022.
- [99] A. Lauber-Rönsberg, “BDSG und andere sondergesetzliche Datenschutzregelungen,” in *Handbuch europäisches und deutsches Datenschutzrecht*, 2019.
- [100] J. Möller-Klapperich and C. Rasquin, “Regulatorische Herausforderungen in KI-Wertschöpfungsketten,” *Neue Justiz*, vol. 10, 2023.
- [101] European Data Protection Board, “Guidelines no. 4/2019 on article 25 data protection by design and default,” 2020.
- [102] European Data Protection Supervisor, “Opinion 5/2018. preliminary opinion on privacy by design,” 2018.
- [103] European Union Agency for Network and Information Security, “Privacy and data protection by design – from policy to engineering,” 2015.
- [104] European Data Protection Board, “Report of the work undertaken by the ChatGPT taskforce,” 2024.
- [105] European Commission, “Directive 2002/58/EC of the european parliament and of the council of 12 july 2002 concerning the processing of personal data and the protection of privacy in the electronic communications sector (directive on privacy and electronic communications),” *Official Journal of the European Union*, 2020.
- [106] Bundesrepublik Deutschland, “Telekommunikation-Digitale-Dienste-Datenschutz-Gesetz (TDDDG),” *Bundesgesetzblatt 2024 I Nr. 149*, 2024.