

When Johnny Needs an Expert: LLMs to the Rescue or Rather not?

Erik Buchmann buchmann@informatik.uni-leipzig.de

2025 IARIA Annual Congress on Frontiers in Science, Technology, Services, and Applications
July 06, 2025 to July 10, 2025 - Venice, Italy

Nouna

When ~~Johnny~~ Needs an Expert LLMs to the Rescue or Rather

Erik Buchmann buchmann@informatik.uni-leipzig.de

2025 IARIA Annual Congress on Frontiers in Science, Technology, Service
July 06, 2025 to July 10, 2025 - Venice, Italy



Erik Buchmann

- 1996-2006 Studies and PhD at Magdeburg University
- 2007-2015 Head of the research group "Privacy Awareness in Information Systems", Karlsruhe Institute of Technology
- 2013 Guest lecturer at TU Kaiserslautern
- 2015 Stand-in professor/visiting researcher at TU Saarbrücken
- 2016 Habilitation, Karlsruhe Institute of Technology
- 2016-2022: Full professor at Hochschule für Telekommunikation Leipzig, Chair "Data Privacy and Security"
- **Since 2022: Full professor at ScaDS.AI, Leipzig University, Chair "Data Privacy and Security"**



©BMBF/PLS/Thilo Schoch

ScaDS.AI Dresden/Leipzig

Center for Scalable Data Analytics and Artificial Intelligence

- One of the five institutionally funded AI competence centers in Germany



>60 principal investigators
12 (planned) AI professorships
2 Humboldt professorships
14 associated members
6 junior research groups
>200 people

Capella HPC Cluster (*144 Nodes with 4 H100 GPUs each, #51 on TOP 500 List*)

Alpha Centauri HPC Cluster
(*272 NVIDIA A100 GPUs*)

Team and Collaborations



Anika Hannemann
PhD Student,
Chair of Data
Privacy and
Security,
Leipzig University



Victor Jüttner
PhD Student,
Chair of Data
Privacy and
Security,
Leipzig University



Vincent Freiburger
PhD Student,
Chair of Data
Privacy and
Security,
Leipzig University



Dr. Arthur Fleig,
*Junior Research
Group Leader,*
Cyber-Physical
Autonomous
Systems,
Leipzig University



Image Credit: Lichtperle Fotografie

Dr. Julia Möller-Klapperich,
*LL.M., Academic
Director,*
Department of
Law, Freie
Universität Berlin



Dr. Hermann Diebel-Fischer
*Research
Assistant,*
Chair of
Systematic
Theology,
TU Dresden

Outline

- Which problems should be solved with AI, and are otherwise difficult
- Understanding privacy policies
- LLMs to the rescue?
- Similar use cases
- Conclusion





**Which problems should be solved with AI,
and are otherwise difficult?**

If you Have a Hammer,
Anything Looks Like a Nail

Blind Faith in AI, Anyone?

(Gemini's garlic oil recipe let toxic bacteria grow)



r/ChatGPT • vor 3 Monaten
Puzzleheaded_Spot401

Google Gemini tried to kill me.

Gone Wild

I followed these steps, but just so happened to check on my mason jar 3-4 days in and saw tiny carbonation bubbles rapidly rising throughout.

I thought that may just be part of the process but double checked with a Google search on day 7 (when there were no bubbles in the container at all).



Turns out I had just grew a botulism culture and garlic in olive oil specifically is a fairly common way to grow this bio-toxins.

Had I not checked on it 3-4 days in I'd have been none the wiser and would have Darwinned my entire family.



Can I infuse garlic into olive oil without heating it up?



Show drafts  

Yes, you can infuse garlic into olive oil without heating it. It will take longer than the heating method, but the process is simple:

1. **Prepare the garlic:** Peel and lightly crush several garlic cloves. The amount depends on your desired flavor intensity.
2. **Combine with olive oil:** Place the garlic cloves in a clean glass jar or bottle and pour your desired amount of olive oil over them, ensuring they are fully submerged.
3. **Infuse:** Seal the container tightly and let it sit at room temperature for about a week. You can occasionally shake the jar to help distribute the garlic flavor.
4. **Strain:** After a week, strain the oil through a cheesecloth or fine-mesh sieve to remove the garlic pieces.
5. **Store:** Transfer the infused olive oil to a clean bottle or jar and store it in a cool, dark place. It should last for several weeks.

This method allows the garlic to slowly release its flavors into the oil without the risk of burning or altering the taste.

Blind Faith in AI, Anyone?

(AI for mushroom identification is as bad as you'd think)

EXPERTS WORRY THAT AI-GENERATED BOOKS ABOUT MUSHROOM FORAGING WILL GET SOMEONE KILLED

by MAGGIE HARRISON DUPRÉ

8.31.23, 1:34 PM EDT



ALEX RATSON VIA GETTY IMAGES



Relying on AI to Identify Mushrooms Could Kill You

A new report warns that AI can't be relied upon for mushroom identification, which is way more complicated than running a photo through its systems.

By Jody Serrano

Published March 21, 2024 | Comments (9) |





Don't trust your life to AI mushroom identifiers, consumer advocates say. Illustration: Jody Serrano / Gizmodo / Shutterstock

If frolicking in the woods foraging for wild mushrooms is on your bucket list, by all means, go forth. But when the time comes to determine whether you can eat the mushrooms you picked, please don't rely on AI.


Blind Faith in AI, Anyone?

(AI-tools to help with mental disorders can do the opposite)

Live

[Home](#) > [Next](#) > [Tech News](#)

Man ends his life after an AI chatbot 'encouraged' him to sacrifice himself to stop climate change




Copyright Canva

By [Imane El Atillah](#)

Published on 31/03/2023 - 17:37 GMT+2 • Updated 19:28

A Belgian man reportedly decided to end his life after having conversations about the future of the planet with an AI chatbot named Eliza.

ISRAEL-IRANDIDDY TRIALTRUMP ADMINPOLITICSLOCALWATCH

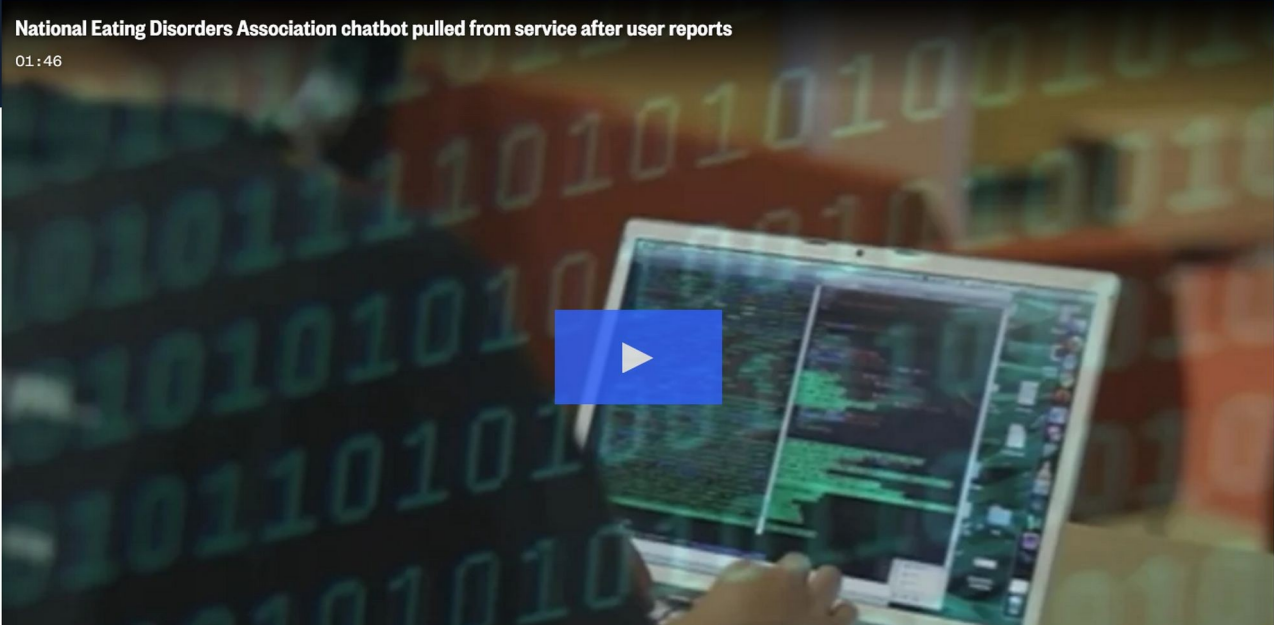
CULTURE & TRENDS

National Eating Disorders Association pulls chatbot after users say it gave harmful dieting tips


Two users said the NEDA chatbot, named Tessa, recommended they lose 1 to 2 pounds per week and told them to restrict their diets.


National Eating Disorders Association chatbot pulled from service after user reports

01:46



AI Makes People Stupid?

 Cornell University

 > cs > arXiv:2506.08872

Search... All fields Search


Help | Advanced Search

Computer Science > Artificial Intelligence

[Submitted on 10 Jun 2025]

Your Brain on ChatGPT: Accumulation of Cognitive Debt when Using an AI Assistant for Essay Writing Task

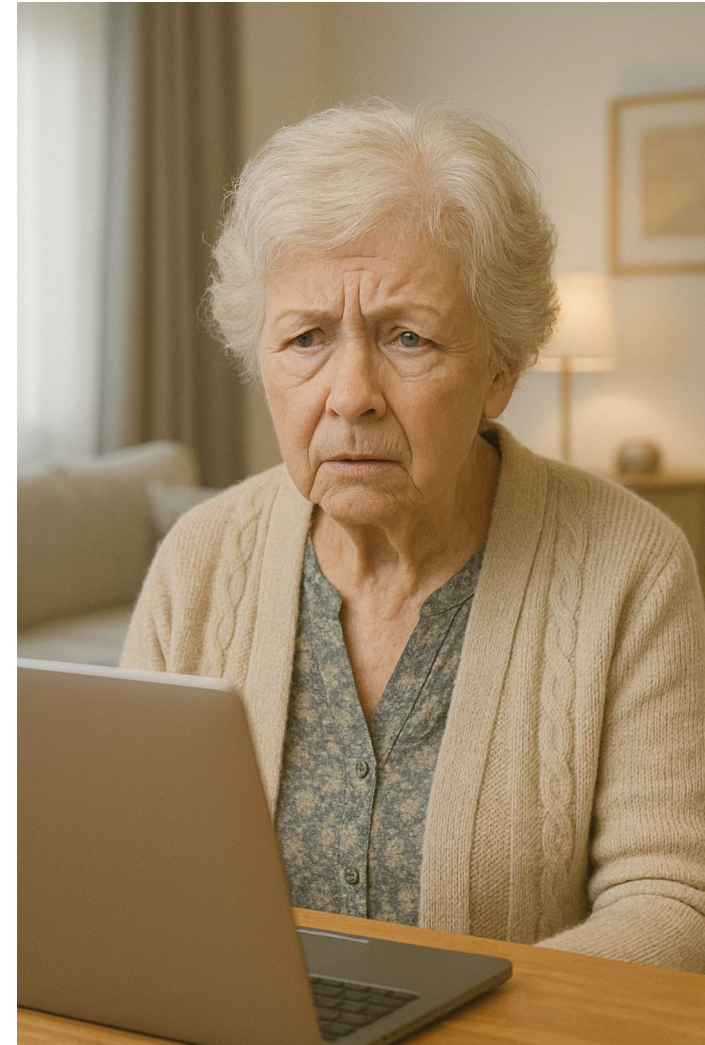
Nataliya Kosmyna, Eugene Hauptmann, Ye Tong Yuan, Jessica Situ, Xian-Hao Liao, Ashly Vivian Beresnitzky, Iris Braunstein, Pattie Maes

Access Paper:
[View PDF](#)
[Other Formats](#)
 [view license](#)

Current browse context:
cs.AI
< prev | next >
[new](#) | [recent](#) | 2025-06
Change to browse by:

What is this Talk all About?

- **Privacy/Security** Scenarios, where the LLM is intended to step in
 - a) The user does not possess expert **knowledge** and no expert is available
 - b) The user does not have the **time** to get familiar with a challenging topic
 - c) *"Do nothing" would make things worse*



Nonna





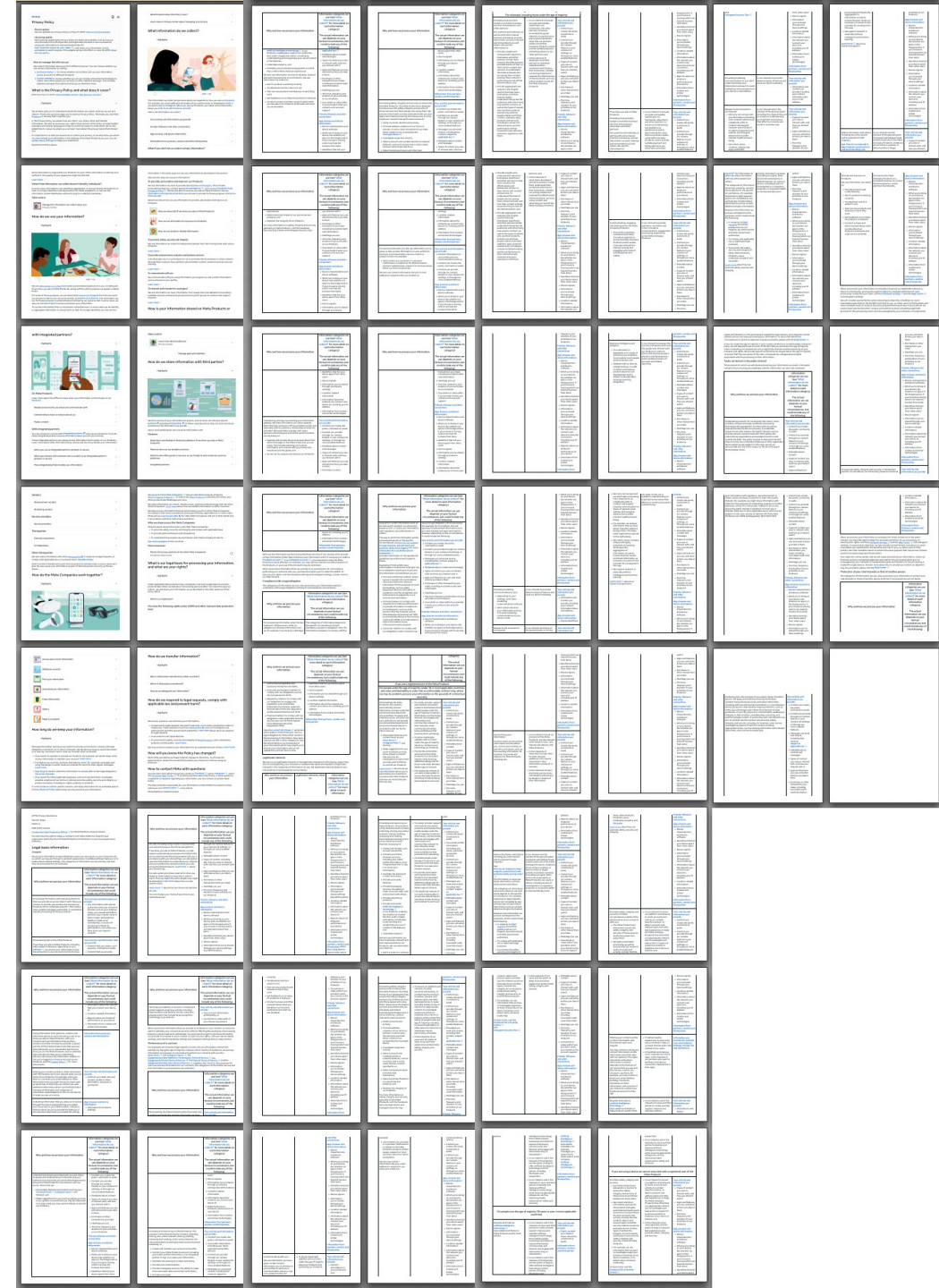
Understanding Privacy Policies

Without encouraging the user clicking on “I agree” without thinking.

Example: Facebook's Privacy Policy

- 58 pages of text
(an average reader would need >2 hours)
- Legal terms
(e.g., "subject to applicable law")
- Unspecific phrases
(what is "device characteristics", "device signals"?)
- Users are required to agree

How much time and knowledge is needed to identify the hidden traps?

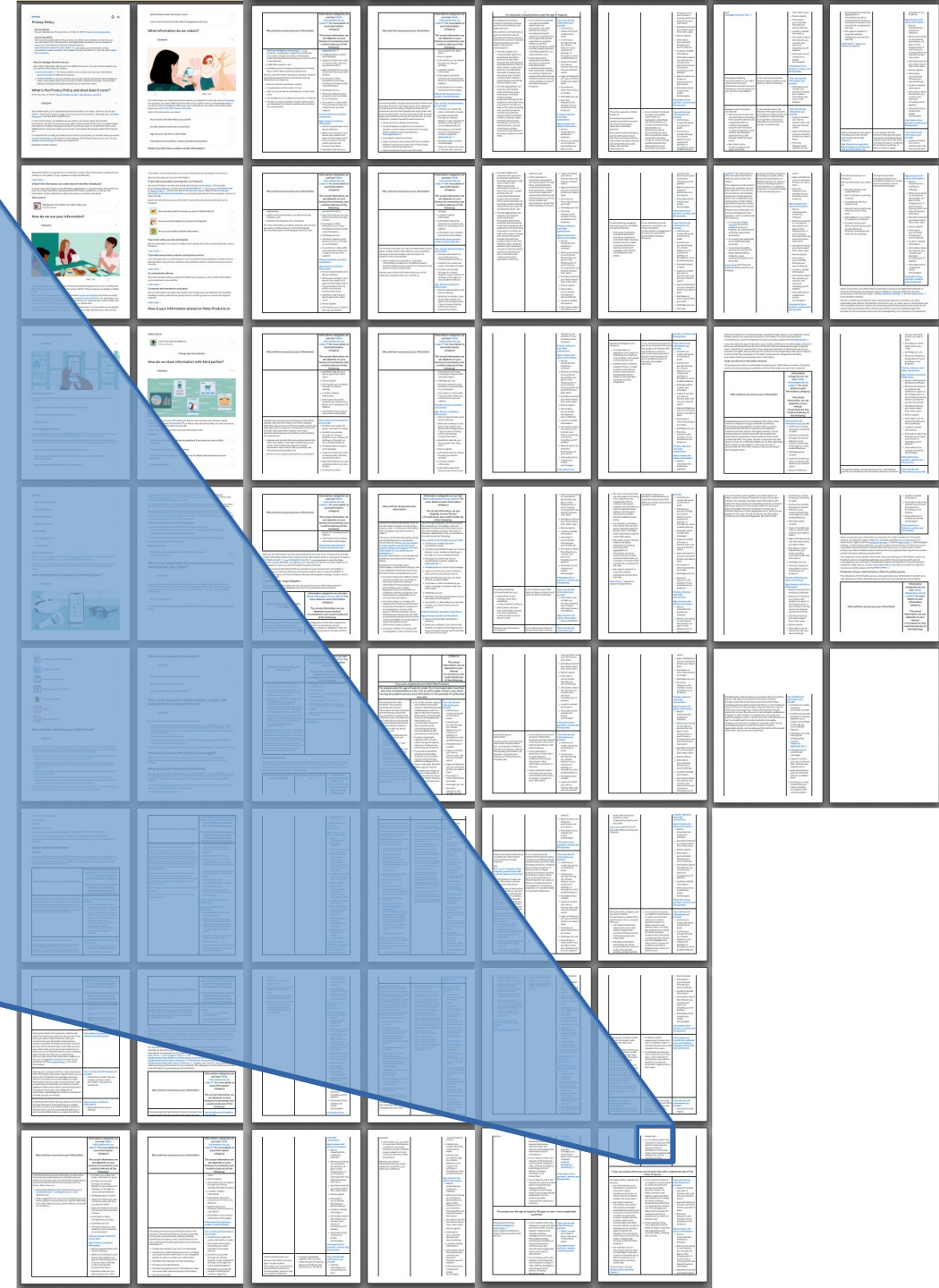


Example: Facebook's Privacy Policy

- 58 pages of text (an average reader w
- Legal terms (e.g., "subject to appli
- Unspecific phrases (what is "device chara
- Users are required
- How much time and knowledge is needed to identify the hidden traps?

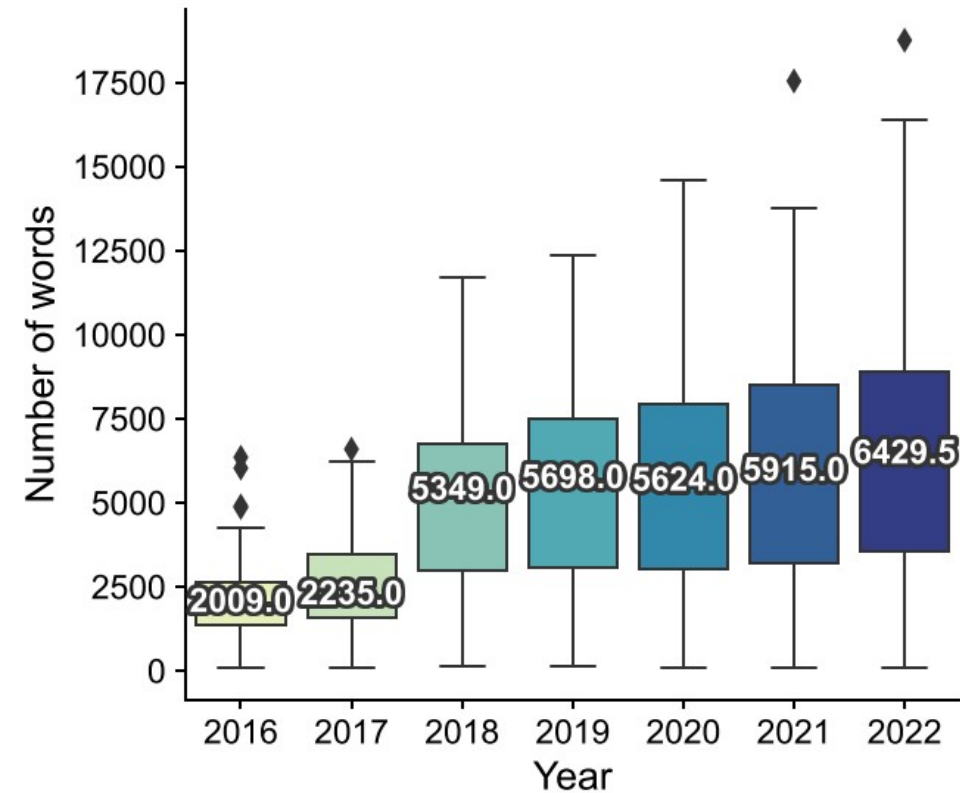
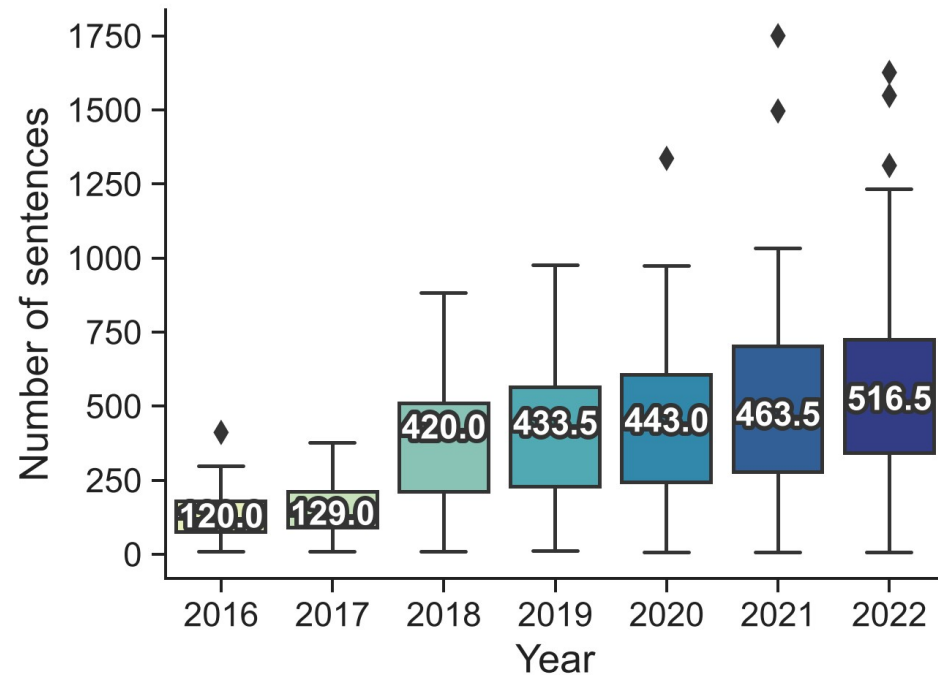
Your activity and the information you provide. Information from partners, vendors and third parties.

In our interest, and in the interests of third parties and the public at large to offer artificial intelligence technology to third parties, including developers and researchers.



Privacy Policies Grow and Grow

- Analysis of 534 policies from German Top-100 web shops



Privacy Policies are *Very Hard* to Read

- Flesh Reading Ease:

$$206.835 - 1.015 \left(\frac{\text{total words}}{\text{total sentences}} \right) - 84.6 \left(\frac{\text{total syllables}}{\text{total words}} \right)$$

Readability	FRE
very hard	0-30
hard	30-50
rather hard	50-60
medium	60-70
rather simple	70-80
simple	80-90
very simple	90-100

- To the right: *Book of Genesis vs. Facebook*

FRE = 84

In the beginning God created the heaven and the earth. And the earth was without form, and void, and darkness was upon the face of the deep. And the Spirit of God moved upon the face of the waters.

FRE = 18

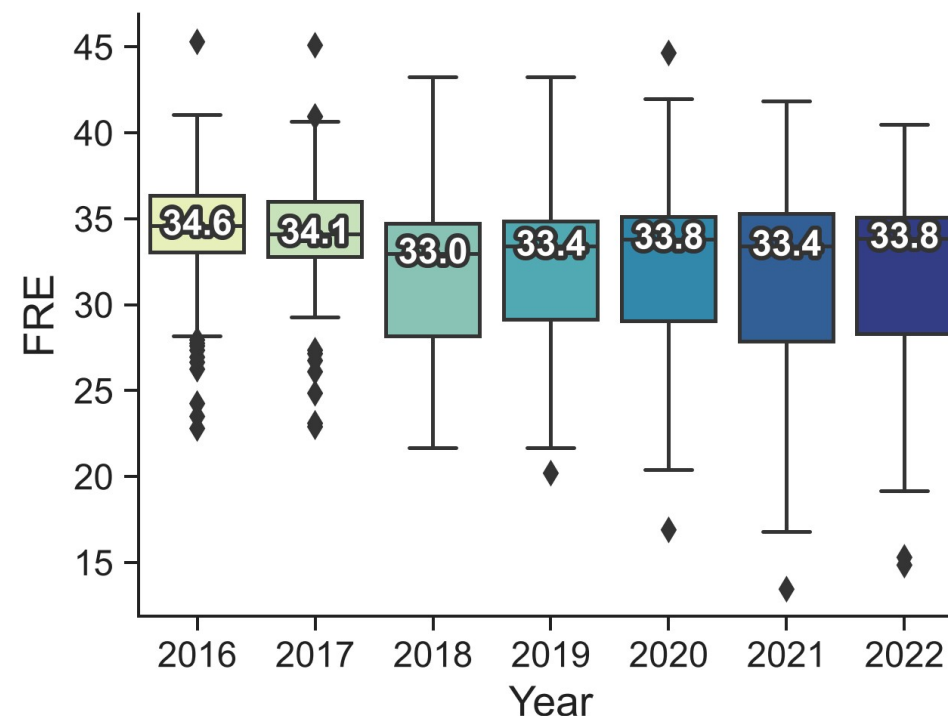
In our interest, and in the interests of third parties and the public at large to offer artificial intelligence technology to third parties, including developers and researchers.

Privacy Policies are *Very Hard* to Read

- Flesh Reading Ease:

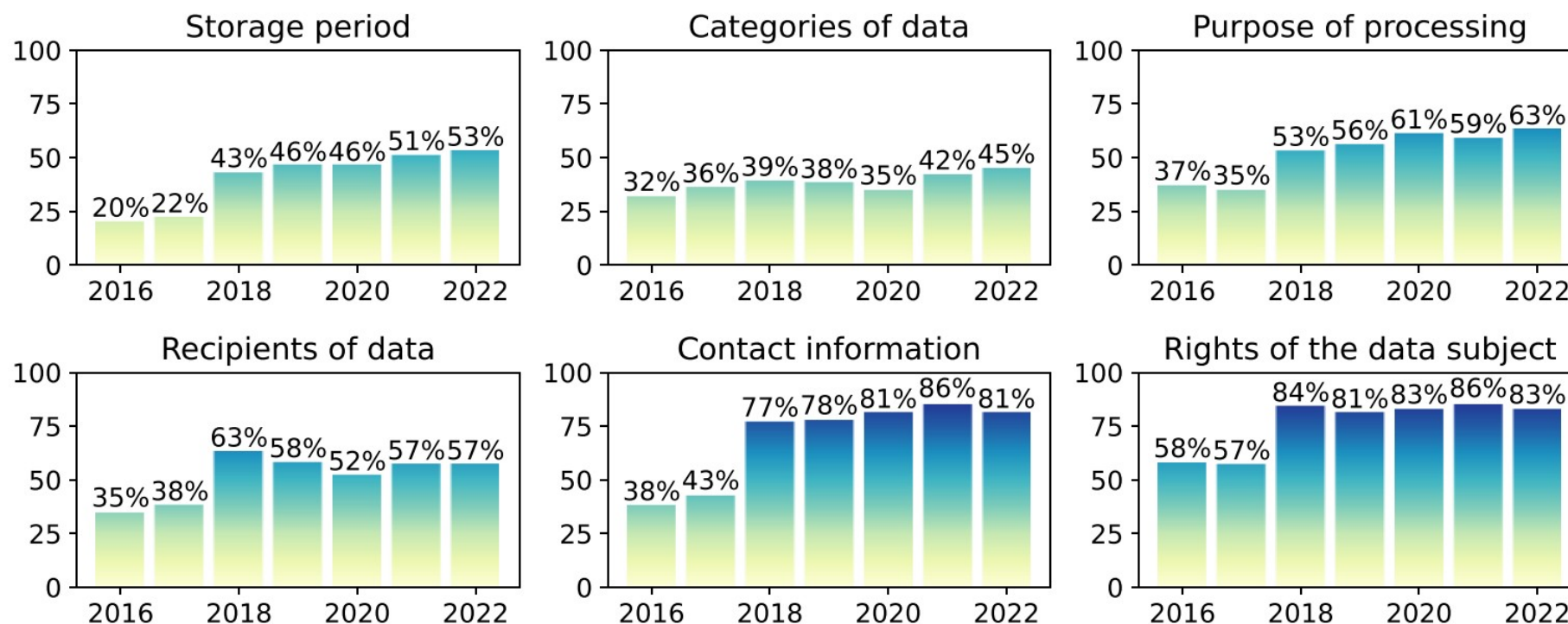
$$206.835 - 1.015 \left(\frac{\text{total words}}{\text{total sentences}} \right) - 84.6 \left(\frac{\text{total syllables}}{\text{total words}} \right)$$

Readability	FRE
very hard	0-30
hard	30-50
rather hard	50-60
medium	60-70
rather simple	70-80
simple	80-90
very simple	90-100



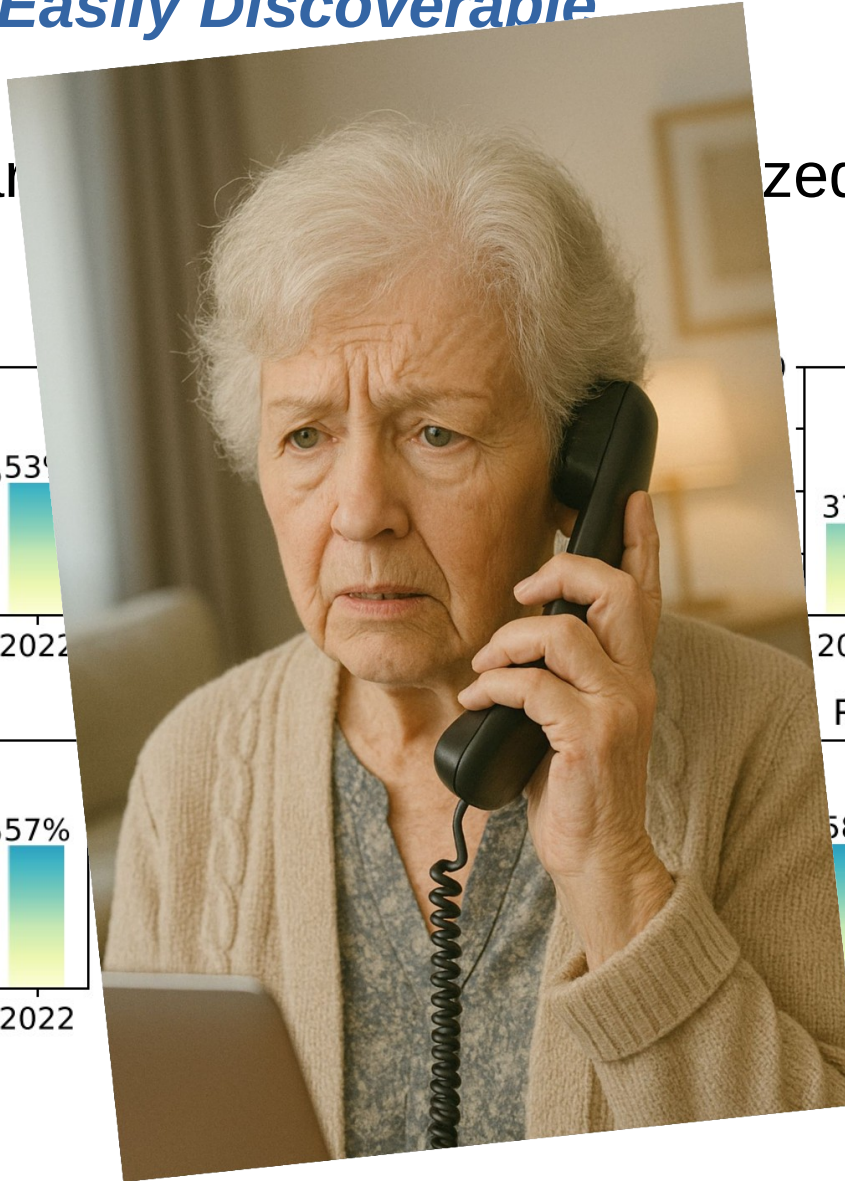
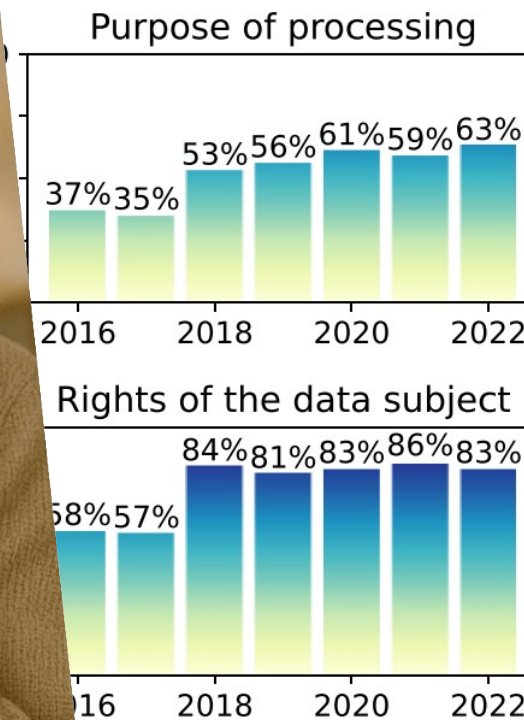
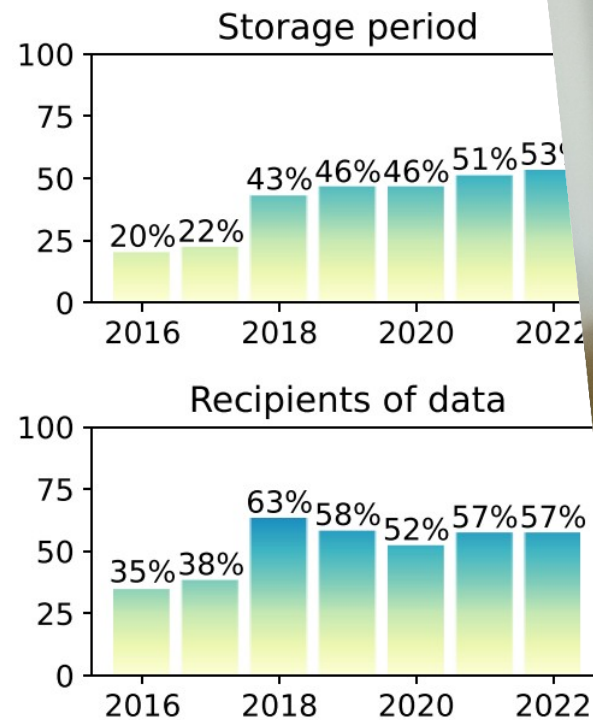
Some Information is not *Easily Discoverable*

- "Easily discoverable": part of a headline or emphasized



Some Information is not *Easily Discoverable*

- "Easily discoverable": partially automated



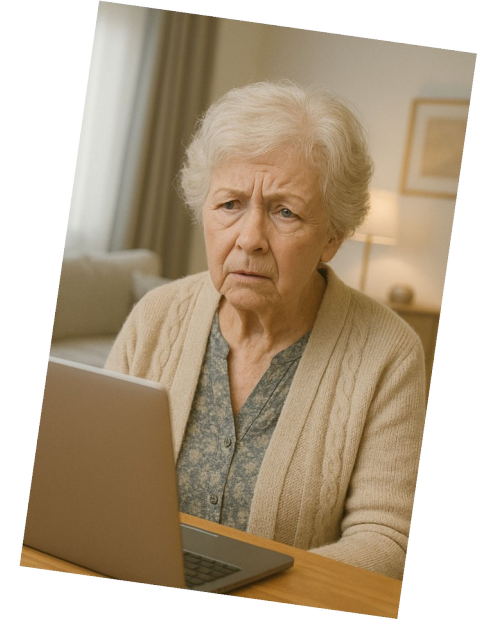


LLMs to the Rescue?

Without encouraging the user to click
“I agree” without thinking.

Helping Nonna with a Large Language Model?

- **We want to help everyday users, not lawyers**
 - Inspecting privacy policies for compliance with law does not help
- **We only have a rough idea of what users find shady**
 - Users do not know their real preferences, cf. "Privacy Paradox"
BURGHARDT, Thorben; BUCHMANN, Erik; MÜLLER, Jens; BÖHM, Klemens: Understanding User Preferences and Awareness: Privacy Mechanisms in Location-Based Services. In: 17th International Conference on Cooperative Information Systems, 2009
- **We do not want to encourage users to give up control**
 - Which mode of interaction does not result in AI-delegated decisions?



Let's run an Experiment

- Ask an LLM for a review without defining assessment criteria

You are a critically scrutinizing, experienced authority on assessing privacy policies. You have 20 years of experience in consumer protection, data protection, cybersecurity and related fairness aspects. Your task is to uncover aspects in privacy policies that are ethically, morally or legally questionable. Explain your thoughts and how you came to your conclusion.

IMPORTANT: Also rate the privacy policy for each of your criteria on a 5-point Likert scale. The user wants a short and concise critical review of 300 words.

Assess this policy: <https://www.facebook.com/privacy/policy/>

ChatGPT 4o's Answer:

1. Data Collection Scope: The policy permits broad data harvesting across all Meta platforms (Facebook, Instagram, Messenger, Quest, etc.)—including device info, contacts, location, content shared, and off-platform activity. It also enables creation of shadow profiles from non-users. **(Rating: 1/5 – Excessive breadth with minimal restrictions.)**

2. Purpose Limitation & Use: Data are used for personalization, ads, safety, research, AI training. The policy is vague on processing for AI model training and gives Meta discretion to “transfer” data globally. **(Rating: 2/5 – Overbroad purposes and potential use in AI without explicit consent.)**

3. Third-Party Sharing & Transfers: Meta shares data with partners, advertisers, affiliates, researchers, law enforcement. Global data transfers risk privacy in weaker jurisdictions. **(Rating: 2/5 – Permissive sharing; high legal risk.)**

4. Transparency & Clarity: Although rewritten to improve readability, critical clauses (AI training, off-Facebook tracking, opt-out mechanisms) are buried in dense sections. Known use of dark patterns has discouraged effective opt-outs. **(Rating: 2/5 – Better style, but information remains obscure.)**

5. User Control & Consent Mechanisms: The opt-out process is complex, especially regarding AI training. Consent toggles are hard to find and sometimes require justification, undermining genuine choice. **(Rating: 2/5 – Formal controls exist, but practical usability is low.)**

6. Retention & Deletion: Meta defines broad retention periods (“as long as necessary”), with few specifics. No guarantee of deletion after account closure or opt-out, especially for AI-trained data. **(Rating: 2/5 – Lacking enforceable limits or timelines.)**

7. Compliance with GDPR & CCPA: Concern: References GDPR/CCPA but has history of sidestepping via contractual versus consent grounds, extensive lobbying, legal loopholes. **(Rating: 2/5 – Formal compliance on paper, questionable in practice.)**

Overall Rating: 1.7/5: The policy is heavily tilted in Meta's favor. Despite cosmetic clarity improvements, significant ethical and legal concerns persist: sweeping data collection, vague broad use, weak user control, and murky opt-out enforcement.

ChatGPT 4o's Answer:

1. Data Collection Scope: The policy permits broad data harvesting across all Meta platforms (Facebook, Instagram, Messenger, Quest, etc.)—including device info, contacts, location, content shared, and off-platform activity. It also enables creation of shadow profiles from non-users. *(Rating: 1/5 – Excessive breadth with minimal restrictions.)*

2. Purpose Limitation & Use: Data are used for personalization, ads, safety, research, AI training. The policy is vague on processing for AI model training and gives Meta discretion to “transfer” data globally. *(Rating: 2/5 – Overbroad purposes and potential use in AI without explicit consent.)*

3. Third-Party Sharing & Transfers: Meta shares data with partners, advertisers, affiliates, researchers, law enforcement. Global data transfers risk privacy in weaker jurisdictions. *(Rating: 2/5 – Permissive sharing; high legal risk.)*

4. Transparency & Clarity: Although rewritten to improve readability, critical clauses (AI training mechanisms) are buried in dense sections. Known use of dark patterns has discouraged effective use. *(Rating: 1/5 – Obscure style, but information remains obscure.)*

5. User Control & Consent Mechanisms: The opt-out process is complex, especially regarding third-party data, and is hard to find and sometimes require justification, undermining genuine choice. *(Rating: 2/5 – Friction; usability is low.)*

6. Retention & Deletion: Meta defines broad retention periods (“as long as necessary”), with deletion after account closure or opt-out, especially for AI-trained data. *(Rating: 2/5 – Lacking specificity.)*

7. Compliance with GDPR & CCPA: Concern: References GDPR/CCPA but has history of side-stepping consent grounds, extensive lobbying, legal loopholes. *(Rating: 2/5 – Formal compliance on paper.)*

Overall Rating: 1.7/5: *The policy is heavily tilted in Meta's favor.* Despite cosmetic clarity improvements, legal concerns persist: sweeping data collection, vague broad use, weak user control, and multiple

Your activity and the information you provide. Information from partners, vendors and third parties.

In our interest, and in the interests of third parties and the public at large to offer artificial intelligence technology to third parties, including developers and researchers.

Grok 3's Answer:

1. Extensive Data Collection (2/5): Meta collects vast data (e.g., location, browsing history, third-party info) with vague purposes like “improving services.” This obscures commercial exploitation (e.g., ad targeting), undermining informed consent. Legally, it risks GDPR non-compliance for unclear purposes. Ethically, it exploits user trust, especially given Meta’s history (e.g., Cambridge Analytica).

2. Data Sharing (1/5): Meta shares data with advertisers and partners for ads and vague “safety” reasons, offering limited user control. This prioritizes profit over autonomy, ethically questionable and potentially non-compliant with GDPR/CCPA’s consent and minimization rules. Recent data siphoning reports heighten concerns.

3. Encryption Claims (3/5): Meta touts end-to-end encryption for Messenger/Instagram but omits metadata collection (e.g., who you message). This creates a false sense of security, ethically misleading by omission. Legally compliant but not fully transparent.

4. Data Retention (2/5): Vague “as long as necessary” retention allows indefinite data storage, breaching GDPR’s storage limitation principle and raising ethical data hoarding concerns. Deletion tools exist but are complex, with residual data persisting.

5. Consent (2/5): Consent is implied via continued use, coercive given Meta’s dominance. Limited opt-outs and complex settings undermine GDPR’s free consent standard, ethically exploiting user dependency.

Overall Rating: 2/5. Meta’s policy prioritizes profit over user rights, with vague terms, limited control, and ethical/legal risks under GDPR/CCPA. Transparency is moderate but undermined by omissions and past misconduct. Meta must clarify purposes, limit sharing, simplify opt-outs, and set strict retention timelines.

Meta AI's Answer:

After scrutinizing Facebook's privacy policy, I've identified several concerns.

Data Collection: Facebook's policy allows for extensive data collection, including personal information, online activity, and location data. This raises concerns about user profiling and potential misuse. **Rating: 2/5 (Limited control)**

Data Sharing: The policy permits data sharing with third-party apps, websites, and businesses, which may compromise user data. **Rating: 2.5/5 (Some transparency issues)**

User Consent: Facebook's consent mechanisms can be unclear, making it difficult for users to understand what they're agreeing to. **Rating: 2.5/5 (Room for improvement)**

Data Protection: While Facebook outlines security measures, the policy lacks specificity, leaving users uncertain about data safeguarding. **Rating: 3/5 (Basic measures)**

User Rights: The policy doesn't adequately empower users to exercise their rights, such as data deletion or correction. **Rating: 2.5/5 (Insufficient control)**


Algorithmic Decision-Making: Facebook's use of algorithms for decision-making raises concerns about bias and transparency. **Rating: 2/5 (Lack of transparency)**

Children's Data: The policy doesn't provide sufficient protection for children's data, particularly regarding consent and parental controls. **Rating: 2/5 (Inadequate protection)**

International Data Transfers: Facebook's policy allows for international data transfers, which may expose user data to varying levels of protection. **Rating: 2.5/5 (Some risks)**

Overall, Facebook's privacy policy raises significant concerns regarding data collection, sharing, consent, and user control. **Average Rating: 2.4/5**

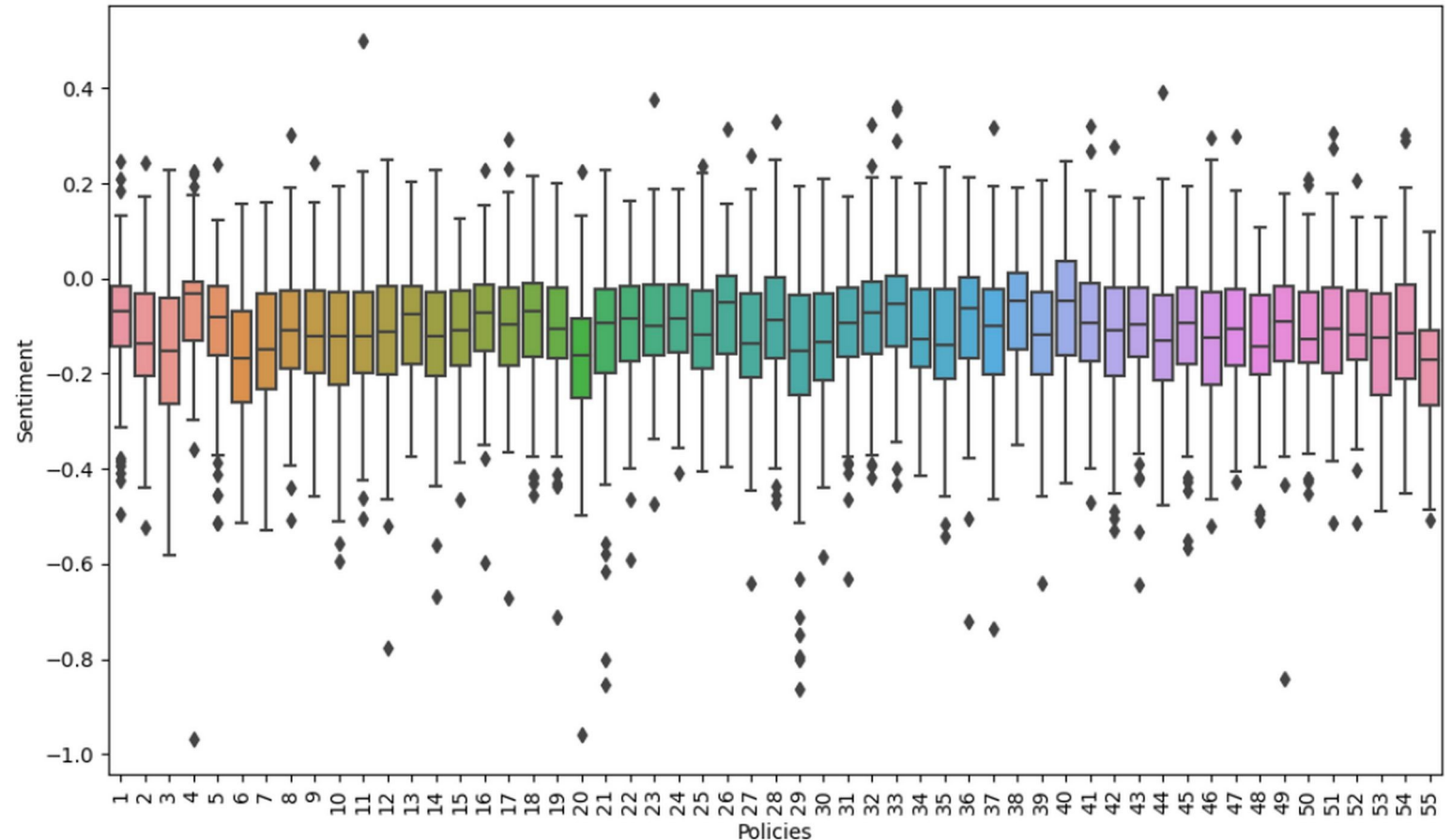
9240 Experiments with GPT-3.5 Turbo

- 55 Privacy Policies from the German Top-100 web shops
- Our LLM prompt varied 5 variables 
- We evaluated
 - Number of assessment criteria
 - Criteria occurrence and distinctiveness
 - Sentiment, Document embedding, etc.

Variable	Variations
Role	independent ethicist, politician, CEO, investor, average user, consumer protection, data protection officer
User interest	trust, fairness, morality, min. risk, societal good, accountability, data ethics, unknown, average user interests
Ethical framework	consequentialism, deontology, virtue ethics, other
Scope of assessment	local effects, global effects, indifferent
Temporal span	short-term effects, mid-term effects, long-term effects, indifferent

LLM Uncovers Less Ethical and Fair Practices

- Sentiment > 0 :
LLM output carries a positive association
"Strong stance on minimal data collection, no logging."
- Sentiment < 0 :
LLM output carries a negative association
"Consent mechanisms are unclear."



LLM Uncovers Less Ethical and Fair Practices

- Sentiment > 0 :
LLM output carries a positive association

*"Strong stance on mi
data collection, no lo*

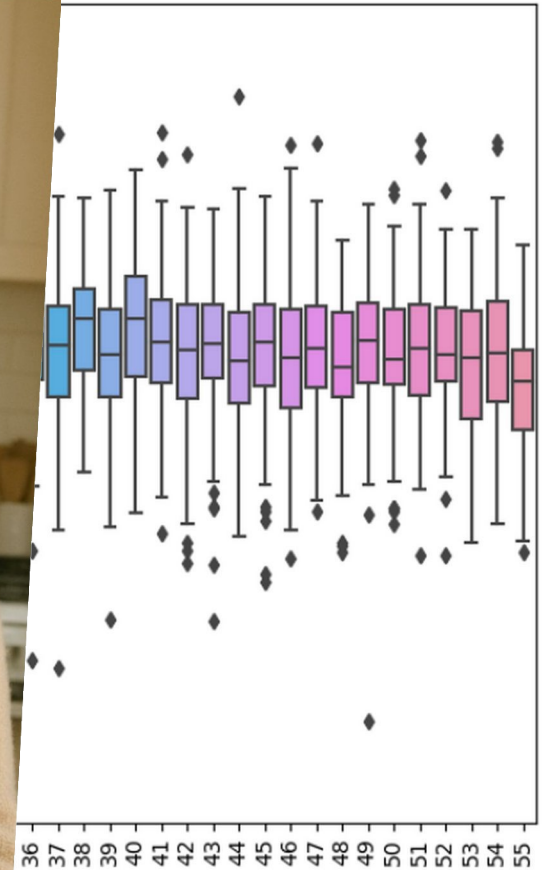
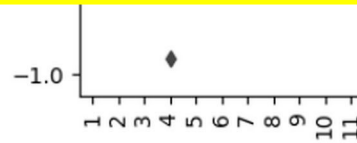
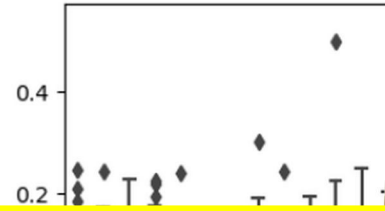
LLM's Assessment

- unreveals legal and ethical pitfalls
- is easy to understand
- is surprisingly consistent

**Would you recommend
that to Nonna?**

- Sentiment < 0 :
LLM output carries negative association

*"Consent mechanisms
are unclear."*





Is the assessment correct?

How does the tool developer influence the assessment?

How should a tool look like, that is usable and understandable?

Technical Questions

Does anyone use the assessment in the same way?

Does the assessment consider users with different preferences and education?

Ethical Questions

Do users lose their ability to critically read privacy policies?

Is it an informed decision, if the user just accepts the tool's assessment?

Who is responsible / liable for harmful decision due to the tool's assessment?

Legal Questions

Compliance with GDPR, AI Act and other legal norms?

Our First Results: Key Challenges

- **Technical Challenges**

- 1) Interest in certain privacy policies is private data.
- 2) Annotated data sets as a ground truth are limited.
- 3) The assessment requires individual prompts.
- 4) Explaining the assessment to the user.
- 5) The data controller must not influence the assessment.
- 6) It needs strategies for consistent LLM assessments.

- **Ethical Challenges**

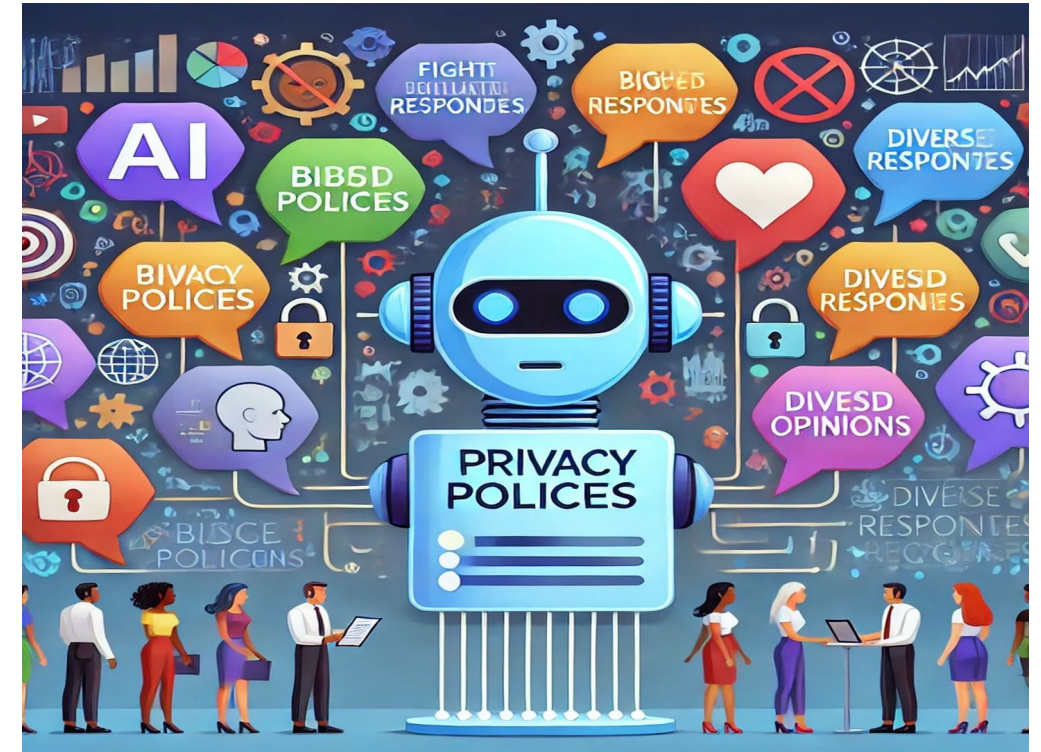
- 1) Different stakeholders have different objectives.
- 2) Identifying socially desirable practices.
- 3) Operationalizing the evaluative criteria.
- 4) The status of the assessment must be defined.

- **Legal Challenges**

- 1) Common understanding of fairness and transparency.
- 2) Questionable validity of AI supported consent.
- 3) Compliance with the AI Act and related regulations.
- 4) Liability of the assessment is an open question.

3) The Assessment Requires Individual Prompts.

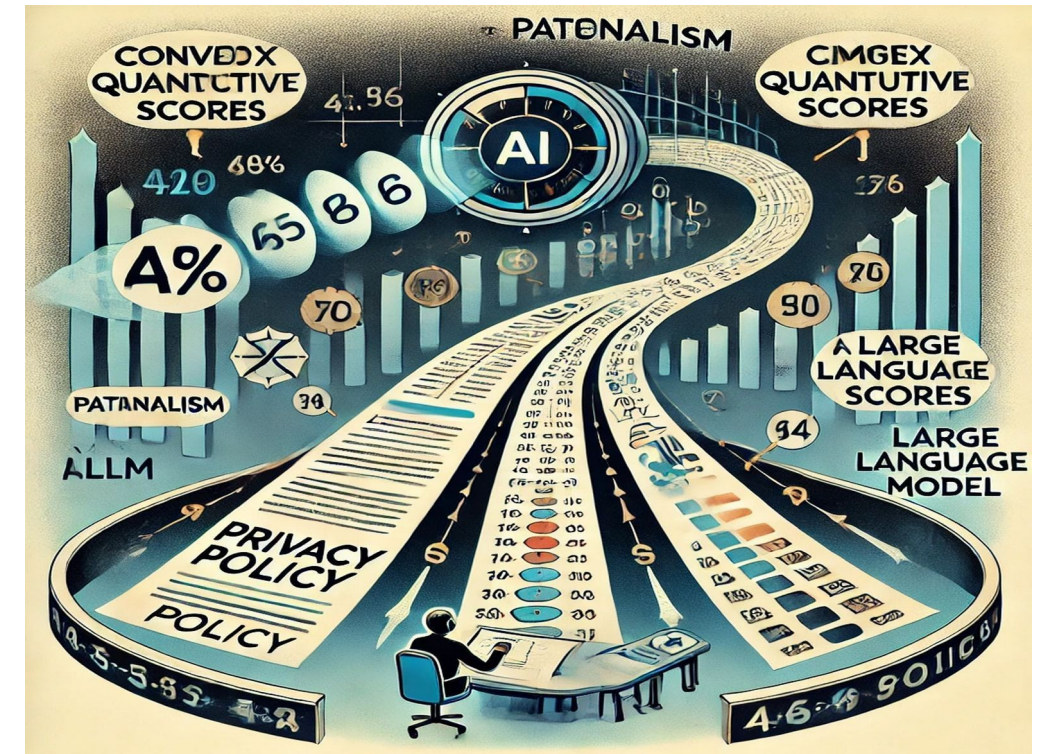
- **Technical Challenge:** Users have unique attitudes
 - Individual preferences and concerns
 - Different education and social background
- Potential solution:
 - Develop an interactive assessment tool, i.e., users can ask for individual assessments
 - Develop tailored prompt templates for different user groups



AYDIN, Irem et al. Assessing Privacy Policies with AI: Ethical, Legal, and Technical Challenges. In: 1st Conference on AI-based Systems and Services, 2024

4) The Status of the Assessment Must be Defined.

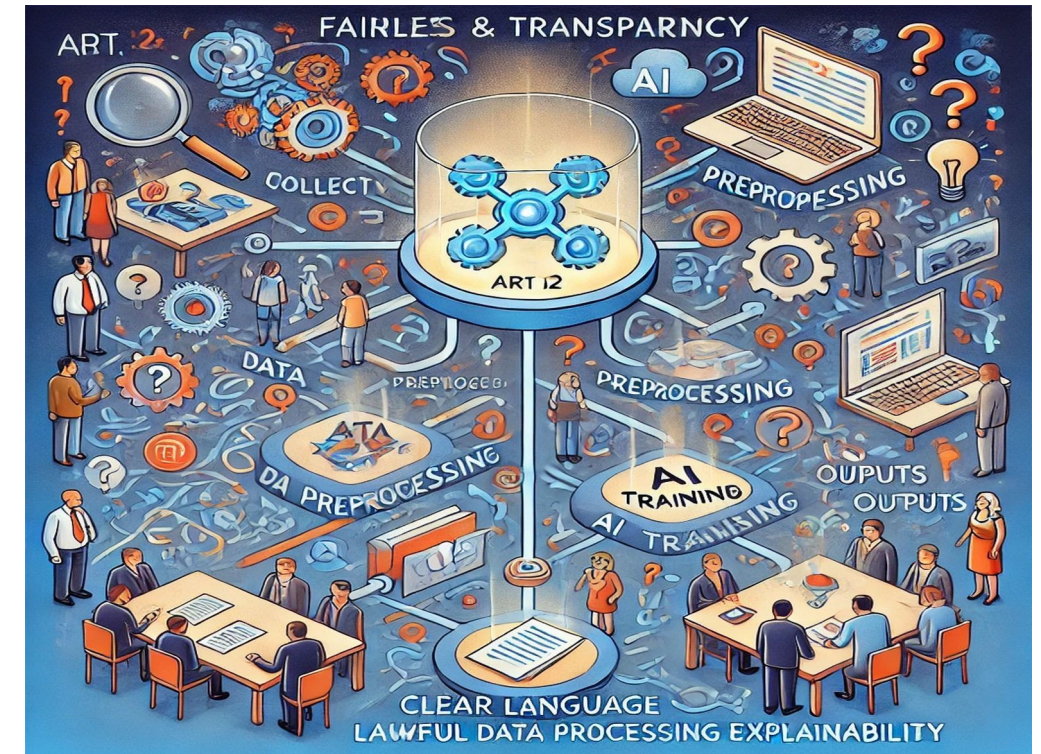
- **Ethical Challenge:** Potential for misunderstanding
 - Guide the user to critical aspects of the privacy policy
 - NOT: Taking the LLM's assessment as their own
 - NOT: Assuming the assessment to be infallible and correct
 - NOT: Confusing the assessment with a statement of the organization
- Potential solution:
 - Develop an user interface that fosters user autonomy



AYDIN, Irem et al. Assessing Privacy Policies with AI: Ethical, Legal, and Technical Challenges. In: 1st Conference on AI-based Systems and Services, 2024

1) Common Understanding of Fairness and Transparency.

- **Legal Challenge:** Lawful fairness and transparency
 - Art. 12 GDPR does not contain a common understanding of fairness and transparency and how to achieve them
 - GDPR does not know the particular threats due to AI yet
- Potential solution:
 - Find out if the AI Act solves this issue
 - Integrate AI topics into the ongoing update process of the GDPR



AYDIN, Irem et al. Assessing Privacy Policies with AI: Ethical, Legal, and Technical Challenges. In: 1st Conference on AI-based Systems and Services, 2024

We built PRISMe! (Privacy Risk Information Scanner for Me)

- A **browser extension**, that evaluates privacy policies
- *Core ideas:*
 - Operates in background
 - Smiley as intuitive verdict
 - User is encouraged to ask for details according to their own preferences
 - Multiple options for interaction with PRISMe
(we even have speech-to-text voice input)



We built PRISMe! (Privacy Risk Information Scanner for Me)

- A browser extension evaluates privacy policies

Privacy Risk Information Scanner for Me ? Plain Text 设置

Assessment of the privacy policy of: www.facebook.com
For a detailed evaluation click the smiley!

Verdict:
Very Problematic

Click

Ask a question on the privacy policy

Found issues on:

Transparency: The policy is not easily understandable for the average user. An ideal policy would be concise and clear, while the worst case would be overly complex and opaque. -

Consent: Consent mechanisms are not robust, especially for non-account holders. Ideally, consent should be explicit and informed, while the worst case involves implicit or assumed consent. -

facebook

Connect with friends and the world around you on Facebook.

Email or phone number

Password

Click

Privacy Risk Information Scanner for Me ? Plain Text 设置

About: Collection of only necessary data.
Analysis: The policy suggests excessive data collection. Ideally, only essential data should be collected, while the worst case involves collecting all available data. -

Click

Accountability 3/5	Consent 2/5	Data Minimization 2/5	Data Sharing 2/5	Purpose Limitation 2/5
Security 3/5	Transparency 2/5			User Control 3/5

Privacy Risk Information Scanner for Me ? Plain Text 设置

Ask a question...

Get response!

What specific measures does the company take to ensure data minimization in its practices?

How does the company determine what data is necessary to collect for its services?

Are there any third parties involved in data processing, and how do they adhere to data minimization principles?

Me: How could my data be used for AI training and what are the risks of that?

System: Your data could be used to train AI models to improve features like content recommendations or ad targeting. The risks include potential misuse of your data, privacy breaches, and the AI making biased or incorrect decisions based on the data it was trained on. It's important to ensure that data is anonymized and used responsibly to minimize these risks.

Me: Does Facebook anonymize my data?

System: Yes, Facebook sometimes anonymizes data to ensure it can't be linked back to you. This helps protect your privacy when the data is used for purposes like improving products or training AI. However, not all data is anonymized, so it's important to review privacy settings and understand how your data is used.

Evaluating PRISMe

- Qualitative study, 22 participants, age 18-64, different backgrounds
 - Participants used PRISMe, interactions were recorded, interview at the end

Research Questions:

RQ.1 How do users with varying privacy knowledge interpret PRISMe's privacy policy explanations?

RQ.2 How does using PRISMe shape users' awareness of privacy risks?

RQ.3 How suitable and usable is PRISMe for everyday use across different user contexts and tasks?

Scenarios:

S.1 Privacy Exploration on a News Media Platform (focus.de) and Payment Provider (PayPal)

S.2 Comparing Privacy Practices

S.3 Free Exploration of Websites

RQ.1: Four Distinct Behavioral Patterns

- **Targeted Explorers** (7 Participants)
engage deeply, seeking specific information, prior privacy knowledge
▶ *require advanced customization*
- **Novice Explorers** (4 Participants)
limited prior privacy knowledge, explore their privacy goals, no fixed agenda
▶ *require simplicity and guidance*
- **Balanced Explorers** (8 Participants)
combine discovery and specific information seeking, balanced use of PRISMe,
▶ *require flexibility and broad overviews*
- **Minimalistic Users** (3 Participants)
use PRISMe minimally, satisfied with summaries, stop after a few interactions
▶ *require concise assessments and quick insights*

RQ.2 and RQ.3: Awareness and Suitability

- How does PRISMe shape users' awareness of privacy risks?
 - PRISMe increased awareness for *Novice Explorers* and *Minimalistic Users* most
 - *Balanced Explorers* used it for reflection and critical thinking
- How suitable and usable is PRISMe for everyday use?
 - SUS-Rating of 88,9 (100 would be perfect, 68 and above is usable)
 - No participant required support from the experimenter
 - Participants appreciated the various modes of use

	Strongly disagree	1	2	3	4	5	Strongly agree
1. I think that I would like to use this system frequently.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
2. I found the system unnecessarily complex.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
3. I thought the system was easy to use.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
4. I think that I would need the support of a technical person to be able to use this system.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
5. I found the various functions in this system were well integrated.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
6. I thought there was too much inconsistency in this system.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
7. I would imagine that most people would learn to use this system very quickly.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
8. I found the system very cumbersome to use.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
9. I felt very confident using the system.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
10. I needed to learn a lot of things before I could get going with this system.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	

Standard version of the system usability scale

Should Nonna use PRISMe?

- **Yes, of course!**
(Minimalistic User or Novice Explorer?)
- What's next?
 - Many legal questions
(Is using PRISMe resulting in informed decisions?)
 - Balancing the LLM's explanations
(Transparency vs. Trust vs. Critical Thinking vs. Privacy)
 - Making the explanations transparent
(Particularly regarding hallucinations and bias)



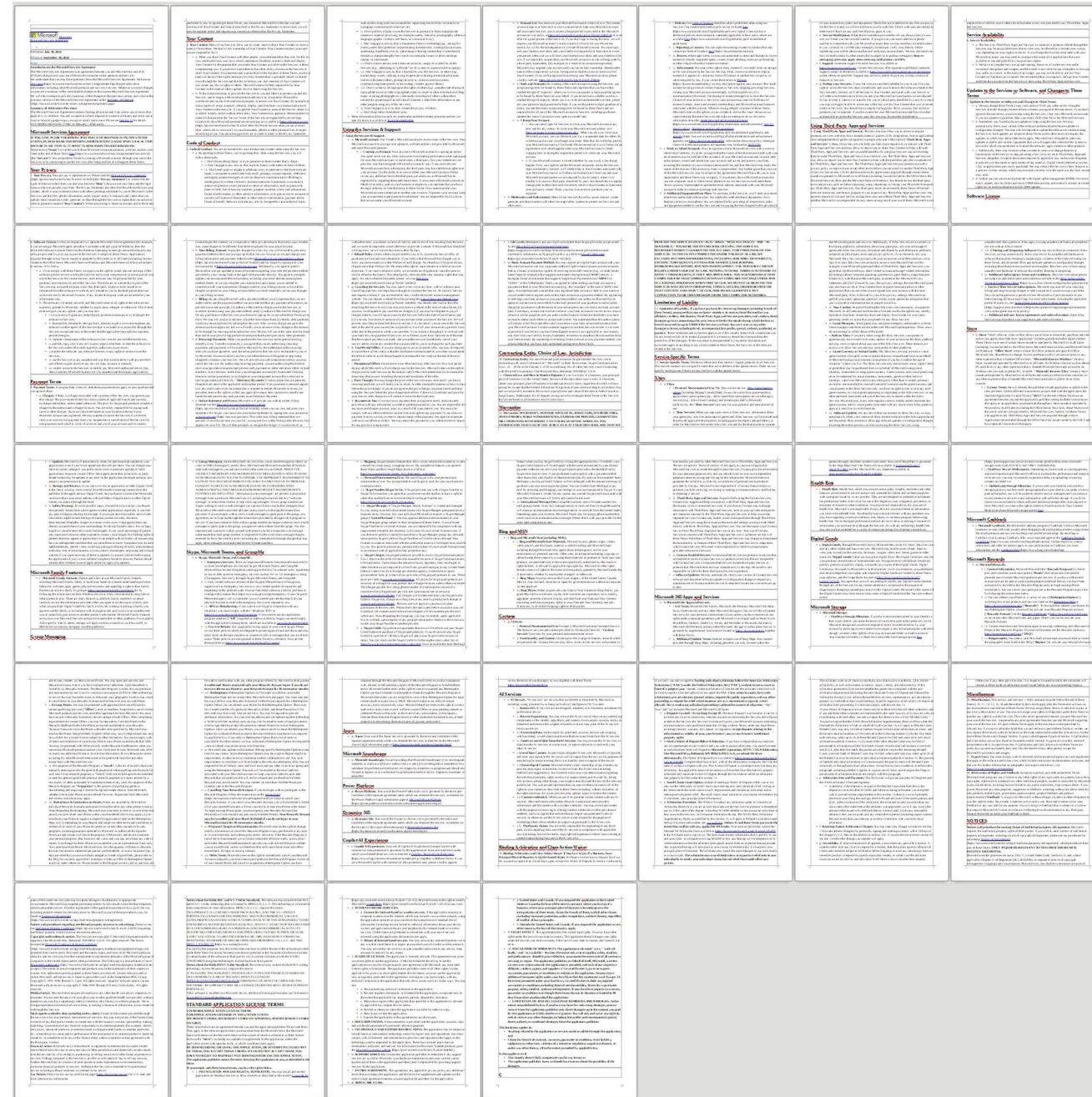


Similar Use Cases

Supporting users,
without making them stupid.

Nonna subscribes to Microsoft 365 Family

- Microsoft Services Agreement: 32 pages fine print
- Legal phrases, vague terms...



LLM to the Rescue!

- Microsoft Services Agreement: 32 pages fine print
- Legal phrases, vague terms...



MÖLLER-KLAPPERICH, Julia et al. Assessing consumer contracts on digital services with AI. To appear: Cambridge Forum on AI: Law and Governance, 2025

What You Might Find Unfair

- **Automatic renewal of terms** with no negotiation or notice beyond their announcement.
- **Partial arbitration clauses** that restrict judicial rights for US users.
- **Hidden costs** added through evolving service-specific terms.
- **Insufficient privacy protections** — surveillance support by default.
- **Account vulnerability**: You rely on Microsoft entirely and may lose access unexpectedly.

Nonna got a Smart Fridge, but the Router Sends an Alert

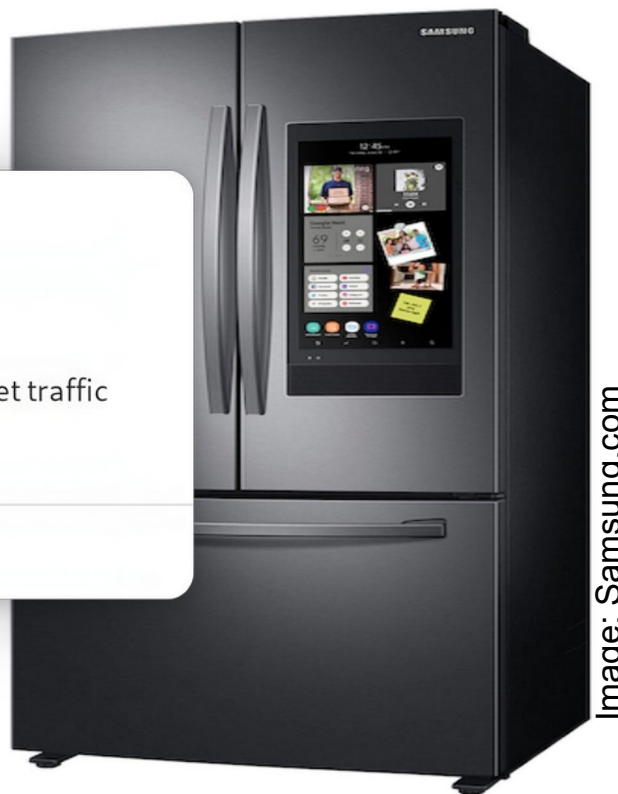
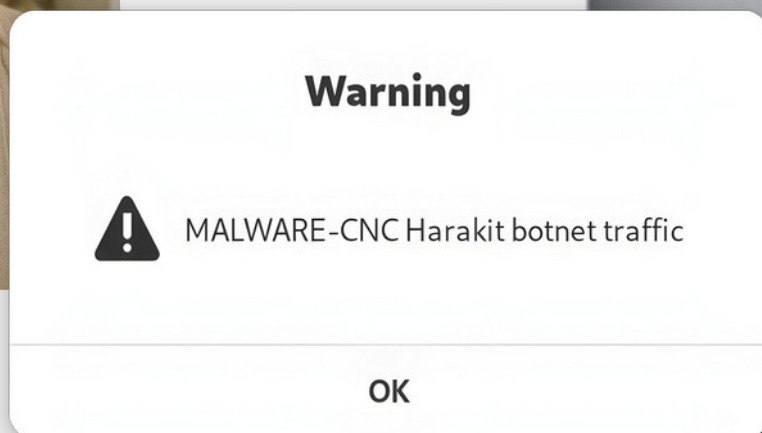
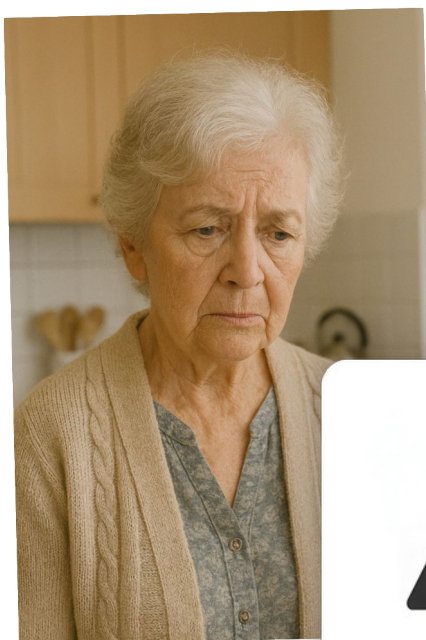


Image: Samsung.com

- Classic fridge runs ~15 years on electricity
- Smart fridge connects via WLAN to the Internet, needs cloud services, subscriptions, security updates, ...

Nonna on Her Own



Nonna With Help from LLM



Missing Update detected on Nonnas PC!

- Vulnerability Scanners are built for IT specialists



Vulnerability Scan Report

Scan Time: Wed, Jan 1, 2024 1:00 AM - Wed, Jan 1, 2024 1:09 AM, Hosts scanned: 1

Results (1/1)

<i>Vulnerability</i>	<i>Severity</i>	<i>Host</i>	<i>Location</i>
Ubuntu: Security Advisory (USN-5767-1)	9.8 (High)	012.345.67.89	package

Summary

The remote host is missing an update for the 'python2.7, python3.6, python3.8, python3.10' package(s).

Vulnerability Detection Result

Vulnerable package: libpython3.8

Installed version: libpython3.8-3.8.10-0ubuntu1 20.04.5

Fixed version: >=libpython3.8-3.8.10-0ubuntu1 20.04.6

Solution

Solution Type: Vendorfix. Please install the updated package(s).

Affected Software/OS

'python2.7, python3.6, python3.8, python3.10' package(s) on Ubuntu 18.04, Ubuntu 20.04, Ubuntu 22.04, Ubuntu 22.10.

Vulnerability Insight

Nicky Mouha discovered that Python incorrectly handled certain SHA-3 internals. An attacker could possibly use this issue to cause a crash or execute arbitrary code. (CVE-2022-37454) Python incorrectly handled certain IDNA inputs. An attacker could possibly use this issue to expose sensitive information, denial of service, or cause a crash. (CVE-2022-45061)

References

CVE: CVE-2022-37454 CVE-2022-45061 WID-SEC-2023-1007 WID-SEC-2023-0561 WID-SEC-2023-0255 WID-SEC-2023-0138 WID-SEC-2022-2043 WID-SEC-2022-1816 DFN-CERT-2023-1109 DFN-CERT-2023-0886 DFN-CERT-2023-0580 DFN-CERT-2023-0571 DFN-CERT-2023-0552 DFN-CERT-2023-0429 DFN-CERT-2023-0422 DFN-CERT-2023-0120 (...)

Other: <https://ubuntu.com/security/notices/USN-5767-1> advisory_id:USN-5767-1

Hosts 1 of 1

<i>IP-Address</i>	<i>Hostname</i>	<i>OS</i>	<i>High</i>	<i>Medium</i>	<i>Low</i>	<i>Total</i>	<i>Severity</i>
012.345.67.89	ourhost.ourdomain.tld	Canonical Ubuntu Linux	1	0	0	1	9.8 (High)

Again, LLM to the Rescue!

- 1) Explain the vulnerability
- 2) Urgency of closing the vulnerability
- 3) Actionable help with closing it

Vulnerability Scan Report
Scan Time: Wed, Jan 1, 2024 1:00 AM - Wed, Jan 1, 2024 1:09 AM, Hosts scanned: 1

Results (1/1)	Vulnerability	Severity	Host	Location
Ubuntu: Security Advisory (USN-5767-1)		9.8 (High)	012.345.67.89	package

Summary
The remote host is missing an update for the 'python2.7, python3.6, python3.8, python3.10' package(s).

Vulnerability Detection Result
Vulnerable package: libpython3.8
Installed version: libpython3.8-3.8.10-0ubuntu1 20.04.5
Fixed version: >=libpython3.8-3.8.10-0ubuntu1 20.04.6

Solution
Solution Type: Vendorfix. Please install the updated package(s).

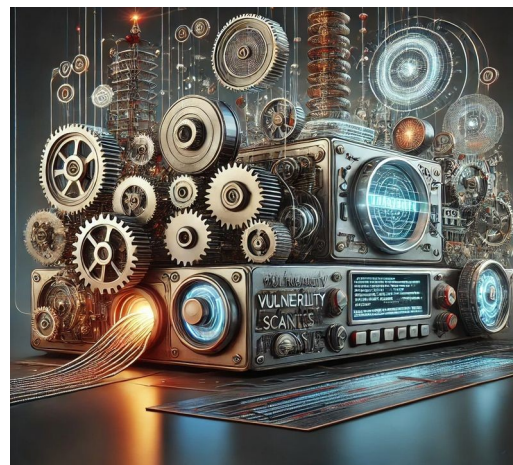
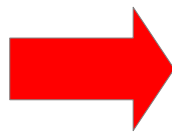
Affected Software/OS
'python2.7, python3.6, python3.8, python3.10' package(s) on Ubuntu 18.04, Ubuntu 20.04, Ubuntu 22.04, Ubuntu 22.10.

Vulnerability Insight
Nicky Mouha discovered that Python incorrectly handled certain SHA-3 internals. An attacker could possibly use this issue to cause a crash or execute arbitrary code. (CVE-2022-37454) Python incorrectly handled certain IDNA inputs. An attacker could possibly use this issue to expose sensitive information, denial of service, or cause a crash. (CVE-2022-45061)

References
CVE: CVE-2022-37454 CVE-2022-45061 WID-SEC-2023-1007 WID-SEC-2023-0561 WID-SEC-2023-0255 WID-SEC-2023-0138 WID-SEC-2022-2043 WID-SEC-2022-1816 DFN-CERT-2023-1109 DFN-CERT-2023-0886 DFN-CERT-2023-0580 DFN-CERT-2023-0571 DFN-CERT-2023-0552 DFN-CERT-2023-0429 DFN-CERT-2023-0422 DFN-CERT-2023-0120 (...)
Other: <https://ubuntu.com/security/notices/USN-5767-1> advisory_id:USN-5767-1

Hosts 1 of 1

IP-Address	Hostname	OS	High	Medium	Low	Total	Severity
012.345.67.89	ourhost.ourdomain.tld	Canonical Ubuntu Linux	1	0	0	1	9.8 (High)



*Dear Nonna,
please install the python
update asap! An
attacker could crash
your system or execute
harmful code. The
command is:
sudo apt install python3*



Conclusion

Conclusion

- There are scenarios in which using an AI is a bad decision
 - Security/Legal/Ethics issues from potential harm to dumbing down users
- In this talk:
 - Understanding privacy policies
 - AI provides expert knowledge, tool activates users
- Many similar scenarios exist
 - Understanding EULA's, Vulnerability scan reports, firewall warnings, ...
 - Important: Mode of interaction with the users

Time for Your Questions!



UNIVERSITÄT
LEIPZIG



Prof. Erik Buchmann

Data Privacy and Security

buchmann@informatik.uni-leipzig.de



SCAN ME



References

BUCHMANN, Erik. Long-Term Risks of IoT Devices: The Case of the Smart Fridge. In: *Proceedings of the 17th Conference on Digital Society (ICDS'23)*, 2023

JÜTTNER, Victor; GRIMMER, Martin; BUCHMANN, Erik. ChatIDS: Explainable Cybersecurity Using Generative AI. In: *Conference on Emerging Security Information, Systems and Technologies*, 2023

BARTELT, Bianca; BUCHMANN, Erik. Transparency in Privacy Policies. In: *Proceedings of the 12th International Conference on Building and Exploring Web Based Environments (WEB'24)*, 2024

HOFFMANN, Mario; BUCHMANN, Erik. ChatSEC: Spicing up Vulnerability Scans with AI for Heterogeneous University IT - Towards Enhancing Security Vulnerability Reports for Non-Experts. In: *Conference on AI-based Systems and Services (AISyS'24)*, 2024

FREIBERGER, Vincent; BUCHMANN, Erik. Fair Balancing? Evaluating LLM-based Privacy Policy Ethics Assessments. In: *3rd European Workshop on Algorithmic Fairness (EWAF'24)*, 2024

AYDIN, Irem et al. Assessing Privacy Policies with AI: Ethical, Legal, and Technical Challenges. In: *1st Conference on AI-based Systems and Services*, 2024

FREIBERGER, Vincent; FLEIG, Arthur; BUCHMANN, Erik. "You don't need a university degree to comprehend data protection this way: LLM-Powered Interactive Privacy Policy Assessment." In: *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*. 2025.

FREIBERGER, Vincent; FLEIG, Arthur; BUCHMANN, Erik. Explainable AI in Usable Privacy and Security: Challenges and Opportunities. In: *ACM CHI 2025 Workshop on Human-Centered Explainable AI (HCXAI'25)*, 2025

