

Transparency in Privacy Policies

Bianca Bartelt
Erik Buchmann

WEB 2024: The Twelfth International Conference on Building and Exploring Web Based Environments

March 10, 2024 to March 14, 2024 - Athens, Greece



About Me



✉ bianca.bartelt@igd.fraunhofer.de

- B.Sc. in Data Science at Philipps University of Marburg (2021), Germany
- M.Sc. in Data Science at Leipzig University (2023), Germany
- Ph.D. student at Fraunhofer Institute for Computer Graphics Research (IGD), Darmstadt, Germany
- Current field of research: digital twins in healthcare

GDPR Mandates Transparency in Privacy Policies

Discoverability?



Privacy Policy of NY Times, accessed March 2024

Readability?

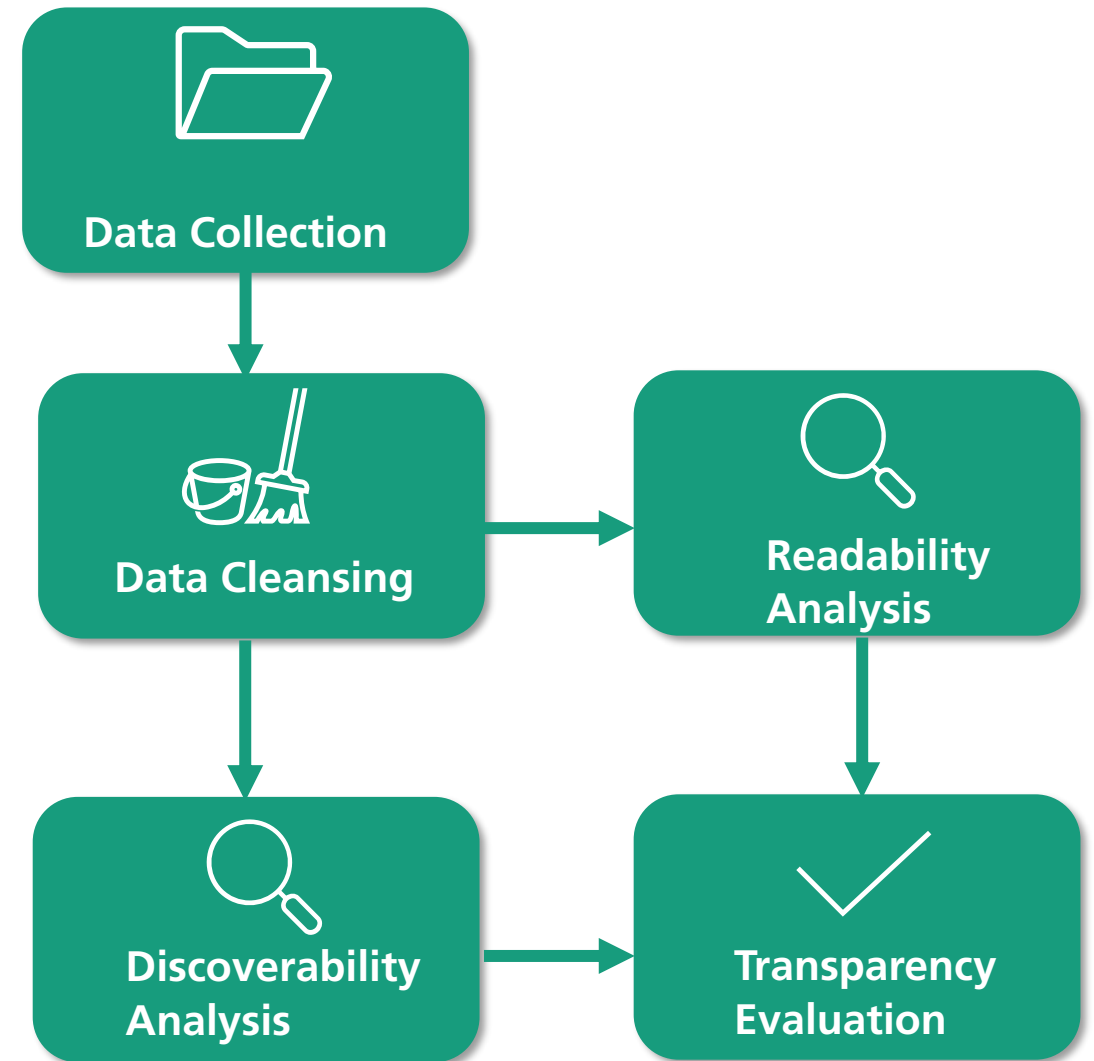
Windows is a personalized computing environment that enables you to seamlessly roam and access services, preferences, and content across your computing devices from phones to tablets to the Surface Hub. Rather than residing as a static software program on your device, key components of Windows are cloud-based, and both cloud and local elements of Windows are updated regularly, providing you with the latest improvements and features. In order to provide this computing experience, we collect data about you, your device, and the way you use Windows. And because Windows is personal to you, we give you choices about the personal data we collect and how we use it. Note that if your Windows device is managed by your organization (such as your employer or school), your organization may use centralized management tools provided by Microsoft or others to access and process your data and to control device settings (including privacy settings), device policies, software updates, data collection by us or the organization, or other aspects of your device. Additionally, your organization may use management tools provided by Microsoft or others to access and process your data from that device, including your interaction data, diagnostic data, and the contents of your communications and files. For more information about data collection in Windows, see [Data collection summary for Windows](#). This statement discusses Windows 10 and Windows 11 and references to Windows in this section relate to those product versions. Earlier versions of Windows (including Windows Vista, Windows 7, Windows 8, and Windows 8.1) are subject to their own privacy statements.

[Learn more](#)

Part of the Privacy Policy of Microsoft, accessed March 2024

Method

- Implementation of a Python script that automatically downloads policies from the Internet Archive
- Collected 534 HTML-format policies from the German Top-100 web shops, 2016 to 2022
- Removed non-essential elements like headers and footers



Readability

- Quantified the readability with **text statistics** and **readability metrics**
- Sentences more than tripled and word count more than doubled due to GDPR in 2018
- Word length remained stable over time; sentences shortened slightly post-GDPR

Windows is a personalized computing environment that enables you to seamlessly roam and access services, preferences, and content across your computing devices from phones to tablets to the Surface Hub. Rather than residing as a static software program on your device, key components of Windows are cloud-based, and both cloud and local elements of Windows are updated regularly, providing you with the latest improvements and features. In order to provide this computing experience, we collect data about you, your device, and the way you use Windows. And because Windows is personal to you, we give you choices about the personal data we collect and how we use it. Note that if your Windows device is managed by your organization (such as your employer or school), your organization may use centralized management tools provided by Microsoft or others to access and process your data and to control device settings (including privacy settings), device policies, software updates, data collection by us or the organization, or other aspects of your device. Additionally, your organization may use management tools provided by Microsoft or others to access and process your data from that device, including your interaction data, diagnostic data, and the contents of your communications and files. For more information about data collection in Windows, see [Data collection summary for Windows](#). This statement discusses Windows 10 and Windows 11 and references to Windows in this section relate to those product versions. Earlier versions of Windows (including Windows Vista, Windows 7, Windows 8, and Windows 8.1) are subject to their own privacy statements.

[Learn more](#)

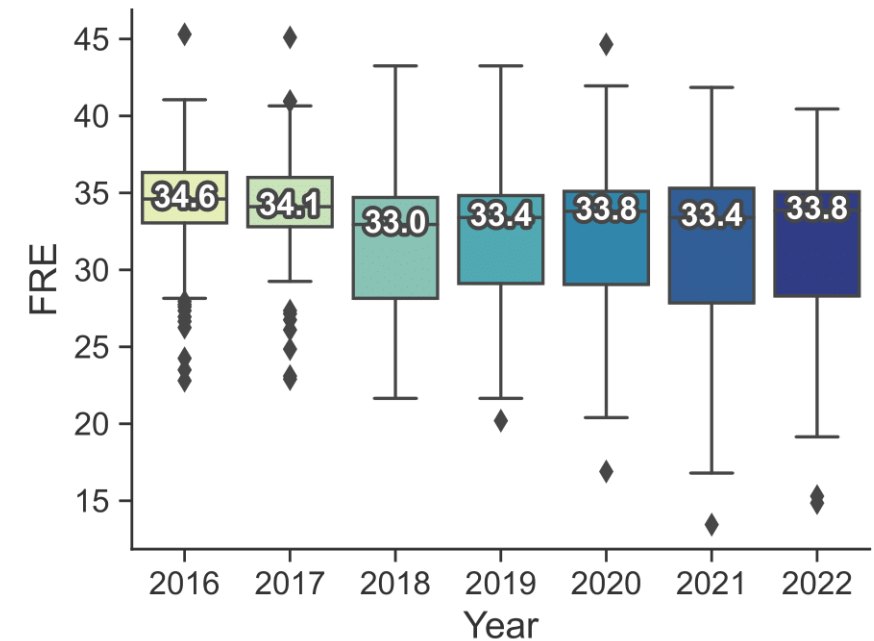
Part of the Privacy Policy of Microsoft, accessed March 2024

Readability

- Median reading competence ranges continuously from hard to very hard
- Flesch Reading Ease (FRE) median value of 33 for 2018

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

- $FRE_{\text{GERMAN}} = 180 - ASL - (ASW \cdot 58.5)$
- ASL: Average sentence length*
- ASW: Average number of syllables per word*




Discoverability


Easily discoverable:

- Information must not be hidden, but highlighted
- (Sub)heading** or **bold text**

...
You have the right to delete
personal information,...



...
§ 2 Your rights
You have the right to delete
personal information,...

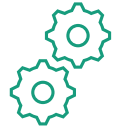


Classes of information:

Class	Description
I.	Storage period
II.	Categories of data
III.	Purpose of processing
IV.	Recipients of data
V.	Contact information
VI.	Rights of the data subject

Discoverability

Our classification toolchain:



Preprocessed highlighted text

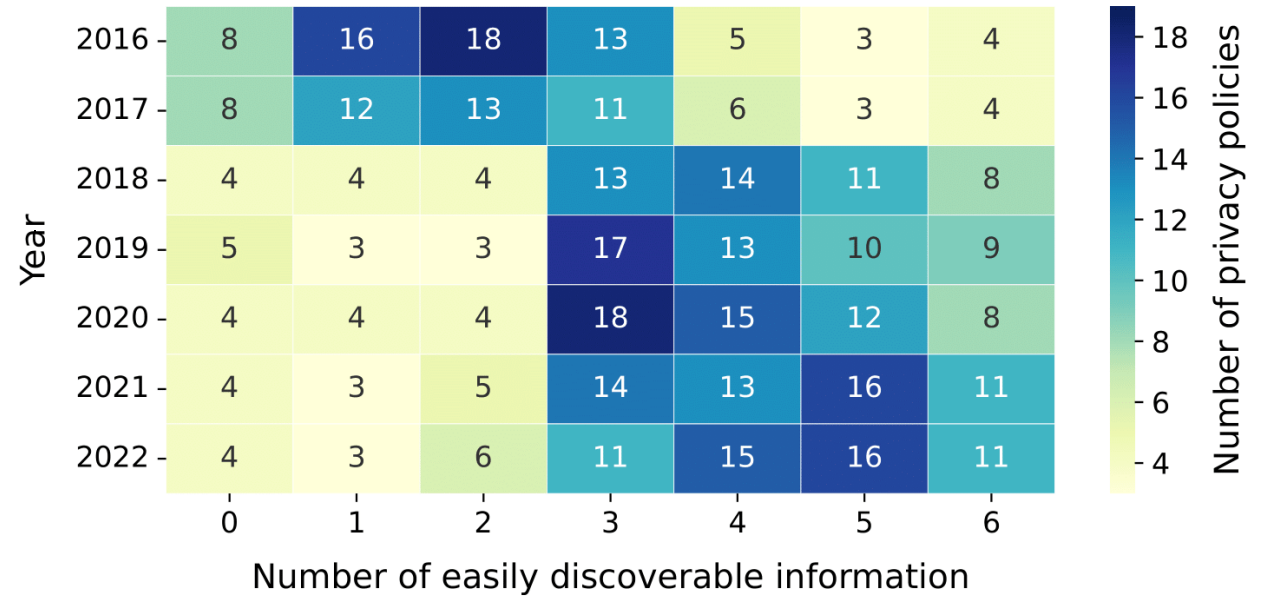
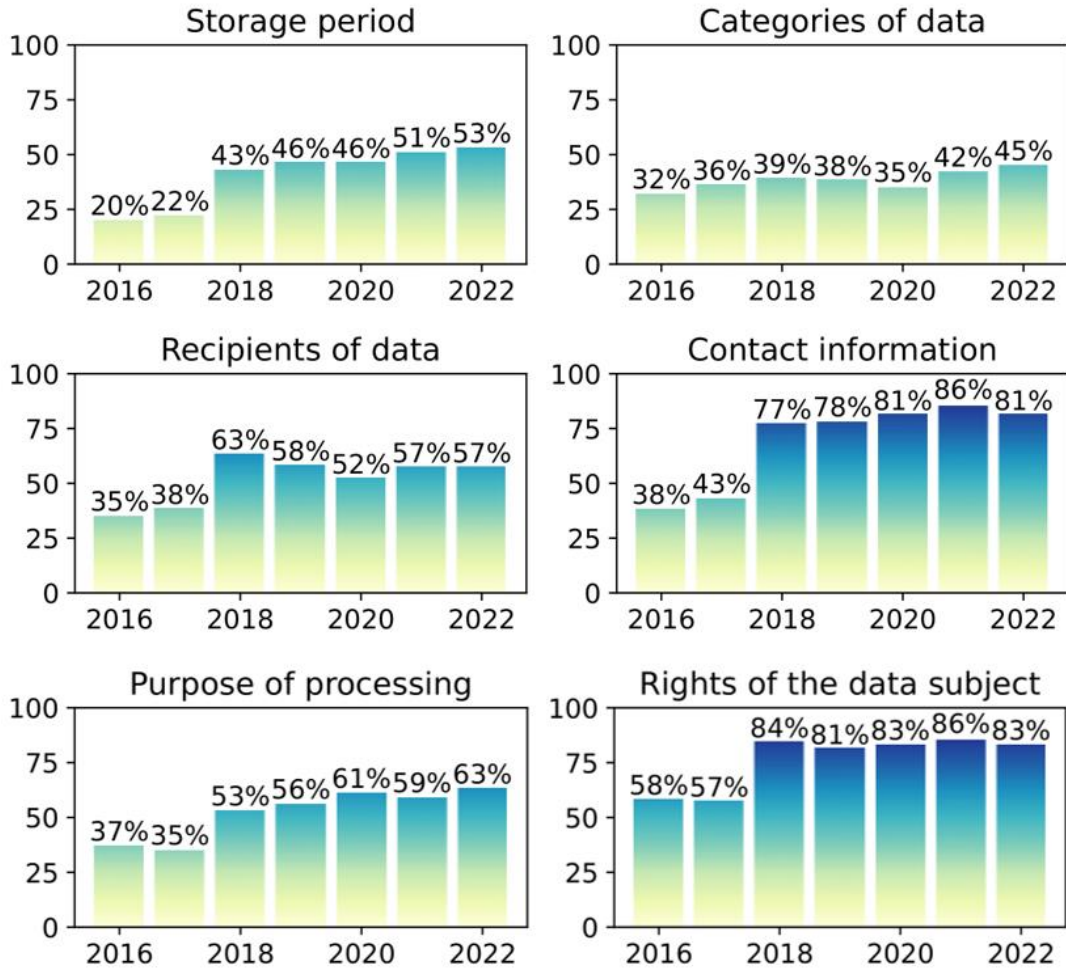
Created and manually labeled a data set of the top 250 token combinations

Calculated cosine similarities between TF-IDF vectors of labeled data set and highlighted text

Assigned a highlighted text with a class label, if cosine similarity was better than 0.75

Achieved **90% accuracy**

Discoverability



Comparison

Our Study (Transparency analysis)	Past Studies (No specific focus on transparency) <i>(Wagner et al., 2017; Amos et al., 2021; Fabian et al., 2017)</i>
Length of German policies increased significantly post-GDPR	Length of English policies increased significantly post-GDPR
Median FRE of 33 for 2018 (Germany)	Median FRE of 32.8 for 2018 (EU)
Discoverability improved	Visual Appearance improved -> Indicates that important information are easier to find

→ **GDPR standardized privacy policy content and enhanced transparency, language-independent**



Conclusion

- We **defined transparency** as readability and discoverability of mandatory information
- We **analyzed the transparency** of 534 German privacy policies from 2016 to 2022
- We **compared our results** with international findings

Our analysis shows that the GDPR ...



- ... improved information discoverability, not reading ease → **Enhancement of transparency**
- ... standardized the content of privacy policies and aligned readability and discoverability

Thank you!
