

An aerial photograph of Zurich, Switzerland, showing a river winding through the city, surrounded by dense urban buildings and greenery. A prominent domed building is visible in the lower right quadrant.

Automated Large-Scale Analysis of Cookie Notice Compliance

Ahmed Bouhoula, Karel Kubicek, Amit Zac,
Carlos Cotrini, David Basin

33rd USENIX Security Symposium
August 14, 2024

It's your choice

When we make the Guardian available to you online, we use cookies and similar technologies to help us to do this. Some are necessary to help our website work properly and can't be switched off, and some are optional but support the Guardian and your experience in other ways.

Yes, I'm happy

Manage my cookies

Allow All

Accept only
necessary

THIS WEBS

We use cookies to collect information about your use of our website and other information that you've provided

We use cookies to personalise content

But first, cookies

We use cookies to improve your user experience on our marketplace. By using Dan, you agree to our [Privacy Policy](#).

Got it

We use cookies and other technologies that are essential to provide you our services and site functionality. If you click "Accept all", we will also allow [third-party companies](#), who we partner with, to store cookies on your device to collect your personal data (e.g., IP address) for advertising personalisation, measurement and analytics. You can opt out of "Decline all" or make granular choices by clicking "Customise my choices." You may withdraw your consent at any time.

Data processing purposes: Store and/or access information on a device. Personalized ads and content, ad and content measurement, audience insights and product development.

Customise my choices

Decline All

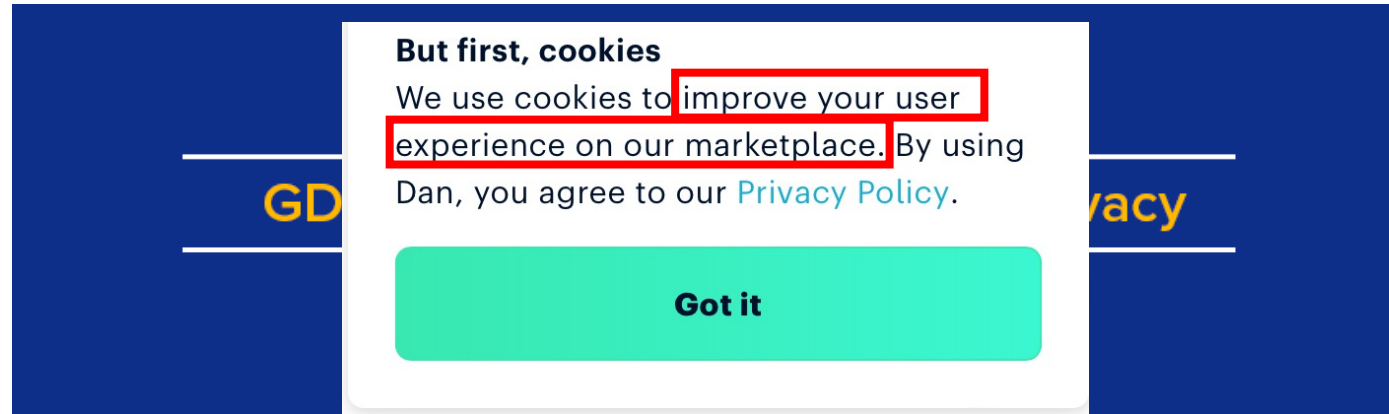
Accept All

Deny

Allow selection

Allow all

Privacy regulations in the EU



- Consent for non-essential cookies should be freely given and unambiguous.
- Personal data should be collected for specified and explicit purposes.

Do websites comply with these regulations?

Bloomberg

Subscribe

Technology

Amazon Gets Record \$888 Million EU Fine Over Data Violations


- Luxembourg watchdog wields new powers with highest data fine
- Fine follows probe based on complaint by French privacy group

30 July 2021, 13:03 CEST *Updated on 30 July 2021, 13:43 CEST*



CNIL.

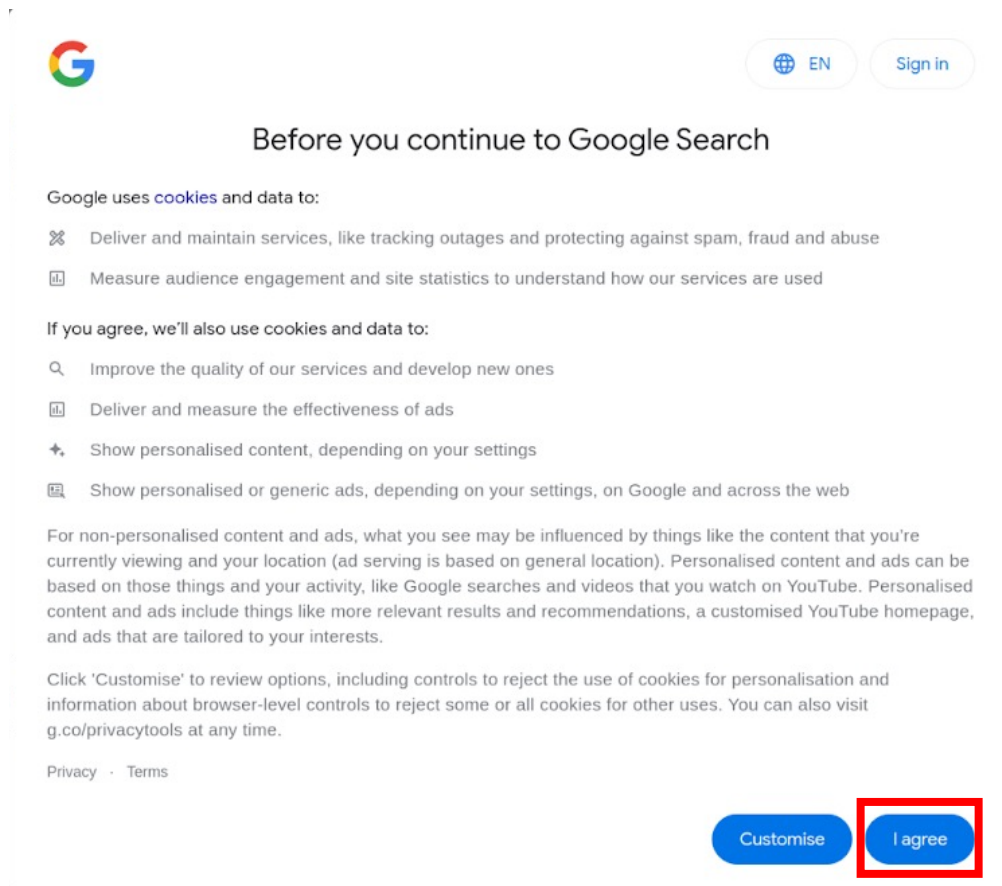


 > Cookies: the CNIL fines GOOGLE a total of 150 million euros and FACEBOOK 60 million euros for non-compliance with French legislation

Cookies: the CNIL fines GOOGLE a total of 150 million euros and FACEBOOK 60 million euros for non-compliance with French legislation

06 January 2022

Do websites comply with these regulations?



Google uses cookies and data to:

- Deliver and maintain services, like tracking outages and protecting against spam, fraud and abuse
- Measure audience engagement and site statistics to understand how our services are used

If you agree, we'll also use cookies and data to:

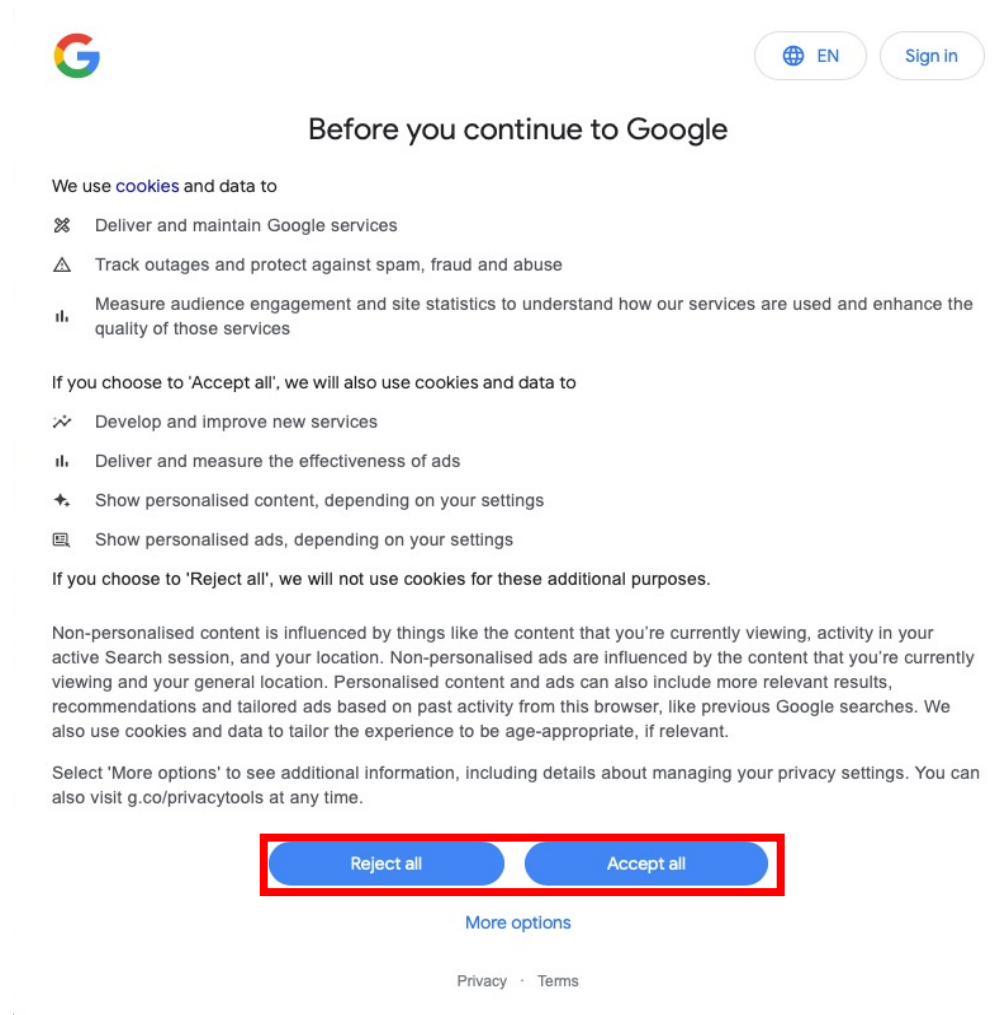
- Improve the quality of our services and develop new ones
- Deliver and measure the effectiveness of ads
- Show personalised content, depending on your settings
- Show personalised or generic ads, depending on your settings, on Google and across the web

For non-personalised content and ads, what you see may be influenced by things like the content that you're currently viewing and your location (ad serving is based on general location). Personalised content and ads can be based on those things and your activity, like Google searches and videos that you watch on YouTube. Personalised content and ads include things like more relevant results and recommendations, a customised YouTube homepage, and ads that are tailored to your interests.

Click 'Customise' to review options, including controls to reject the use of cookies for personalisation and information about browser-level controls to reject some or all cookies for other uses. You can also visit [g.co/privacytools](https://www.google.com/privacytools) at any time.

Privacy · Terms

Buttons: Customise, I agree



We use cookies and data to:

- Deliver and maintain Google services
- Track outages and protect against spam, fraud and abuse
- Measure audience engagement and site statistics to understand how our services are used and enhance the quality of those services

If you choose to 'Accept all', we will also use cookies and data to:

- Develop and improve new services
- Deliver and measure the effectiveness of ads
- Show personalised content, depending on your settings
- Show personalised ads, depending on your settings

If you choose to 'Reject all', we will not use cookies for these additional purposes.

Non-personalised content is influenced by things like the content that you're currently viewing, activity in your active Search session, and your location. Non-personalised ads are influenced by the content that you're currently viewing and your general location. Personalised content and ads can also include more relevant results, recommendations and tailored ads based on past activity from this browser, like previous Google searches. We also use cookies and data to tailor the experience to be age-appropriate, if relevant.

Select 'More options' to see additional information, including details about managing your privacy settings. You can also visit [g.co/privacytools](https://www.google.com/privacytools) at any time.

Buttons: Reject all, Accept all

More options

Privacy · Terms


Previous studies

- Studies focused on websites (non-)compliance:
 - Matte et al. (2020), Santos et al. (2021), Bollinger et al. (2022)
 - Studies focused on deceptive design patterns (a.k.a dark patterns):
 - Nouwens et al. (2020), Soe et al. (2020), Gray et al. (2021)
- **Some studies were manual, the others were limited to specific websites.**
- Low coverage of websites (Matte et al. 5%, Nouwens et al. 6.8%, Bollinger et al. 0.6%).
 - Potential selection bias.
- **No cookie notice interaction.**

Contributions

- ML for a general method of cookie notice compliance analysis.
 - declared purpose detection, interactive elements classification, cookie classification
- Our crawler interacts with cookie notices, determines which cookies are declared, and contrasts them with the cookies the website actually uses.
- Detection of 6 types of violations and 2 types of dark patterns
 - Large-scale analysis over 97k websites
 - Bias analysis of previous studies

Interactive elements classification

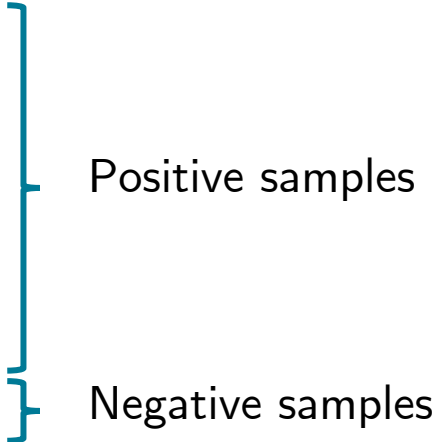
Zoom uses cookies and similar technologies as strictly necessary to make our site work. We  and our partners would also like to set additional cookies to analyze your use of our site, to personalize and enhance your visit to our site and to show you more relevant content and advertising. These will be set only if you accept.

You can always review and change your cookie preferences through our cookie settings page. For more information, please read our [Privacy Statement](#).

[COOKIES SETTINGS](#) [DECLINE COOKIES](#) [ACCEPT COOKIES](#)

The image shows a screenshot of a Zoom cookie consent banner. The banner is enclosed in a black border. The text is in a sans-serif font. The words 'We' and 'Privacy Statement' are highlighted with red boxes. The 'Privacy Statement' is a blue link. At the bottom, there are three buttons: 'COOKIES SETTINGS' (a blue link), 'DECLINE COOKIES' (a blue button), and 'ACCEPT COOKIES' (a blue button). A red box highlights a close button icon in the top right corner of the banner.

Interactive elements classification

- We collect and annotate 2353 samples with one of 6 labels:
 - Cookie settings (290 samples)
 - Accept (229 samples)
 - Reject (153 samples)
 - Close/Continue without accepting (51 samples)
 - Save cookie settings (101 samples)
 - Other (1529 samples)

The diagram uses blue curly braces to group the six labels into two categories. The top three labels (Cookie settings, Accept, and Reject) are grouped together and labeled 'Positive samples'. The bottom three labels (Close/Continue without accepting, Save cookie settings, and Other) are grouped together and labeled 'Negative samples'.

Label	Count	Category
Cookie settings	290	Positive samples
Accept	229	
Reject	153	
Close/Continue without accepting	51	Negative samples
Save cookie settings	101	
Other	1529	
- We train a BERT model with a 95.1% accuracy and a 90.9% F1 score.

Consent options detection

Zoom uses cookies and similar technologies as strictly necessary to make our site work. We and our partners would also like to set additional cookies to analyze your use of our site, to personalize and enhance your visit to our site and to show you more relevant content and advertising. These will be set only if you accept.

You can always review and change your cookie preferences through our cookie settings page. For more information, please read our [Privacy Statement](#).

[COOKIES SETTINGS](#) [DECLINE COOKIES](#) [ACCEPT COOKIES](#)

Cookies settings Reject Accept

Close

zoom [Close](#)

You can accept or decline all but Strictly Necessary Cookies, or customize your cookie settings below. You can change your cookie settings at any time. Some Strictly Necessary Cookies may transfer personal data to the United States. To learn more about how Zoom processes personal data, please visit our [privacy statement](#).

For California Residents, you may exercise your right to "Opt-Out of the Sale of Personal Information" by toggling the button labeled "Targeting" below to off. [More information](#) [Accept](#)

[ACCEPT ALL COOKIES](#)

Manage Consent Preferences


+ Targeting Cookies	<input type="checkbox"/>
+ Functional Cookies	<input type="checkbox"/>
+ Performance Cookies	<input type="checkbox"/>
+ Strictly Necessary Cookies	Always Active

[CONFIRM MY CHOICES](#)

Save cookies settings

Cookie extraction

[Close](#)

Zoom uses cookies and similar technologies as strictly necessary to make our site work. We  and our partners would also like to set additional cookies to analyze your use of our site, to personalize and enhance your visit to our site and to show you more relevant content and advertising. These will be set only if you accept.

You can always review and change your cookie preferences through our cookie settings page. For more information, please read our [Privacy Statement](#).

[COOKIES SETTINGS](#) [DECLINE COOKIES](#) [ACCEPT COOKIES](#)

[Cookies settings](#) [Reject](#) [Accept](#)

We browse the website after clicking on each of the detected consent options to extract the corresponding set of cookies

Analysis of observed cookies

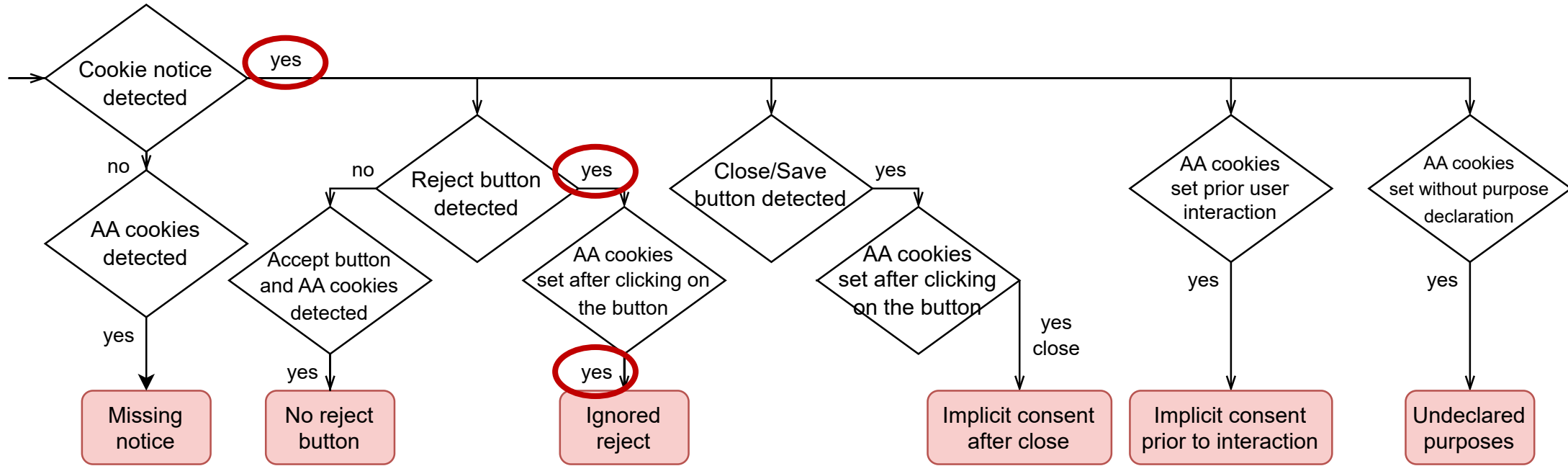
Name	Value	Domain	Expires	...	Bollinger et al. classification	AA cookie
__adroll	4179457951bc9a1a6c...	.adroll.com	2025-09-06	...	Advertising	Yes
_fbp	fb.1.1723024326113...	.zoom.us	2024-11-05	...	Advertising	Yes
_gid	GA1.2.1648944394.17...	.zoom.us	2024-08-08	...	Analytics	Yes
_ga	GA1.2.714463772.17...	.zoom.us	2025-09-11	...	Analytics	Yes
_zm_ssid	aw1_c_hWj7dYsnSj2c...	.zoom.us	Session	...	Essential	No
clientpref	theme-clientpref-dark	.zoom.us	2024-09-06	...	Functional	No

➔ The website uses AA cookies

- A website is classified as “uses AA cookies” if we detect 2 or more AA cookies
- The model achieves an accuracy of **92.9%** and a precision of **98.7%**

AA cookie = Analytics or Advertising cookie

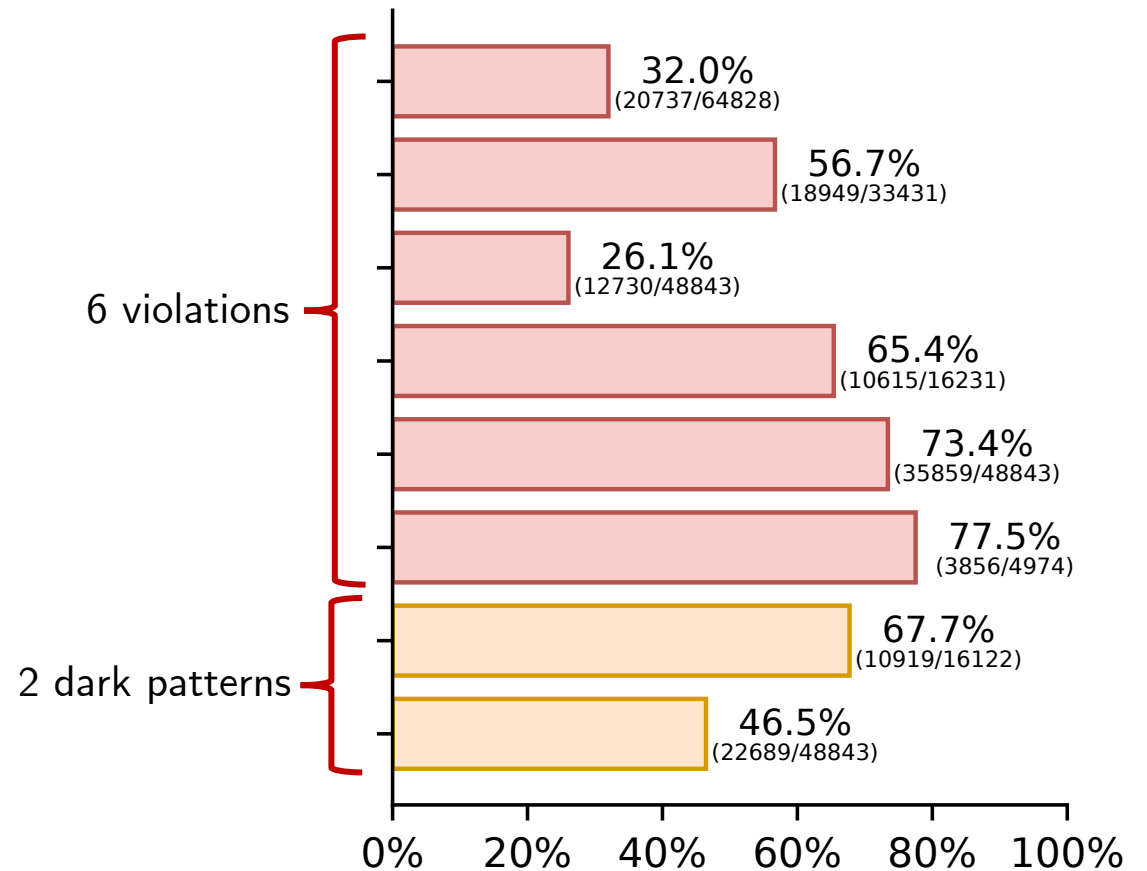
Detecting potential violations



Results: Observed violations

- We crawled 97k websites popular in 15 European countries according to the Chrome User Experience report (CrUX)

→ Aggregation of top 10k websites across these countries



Results: Observed violations



No notice despite AA cookies detected

Article 5(3) of the ePrivacy Directive



Undeclared purposes in the notice text

GDPR informed consent requirement



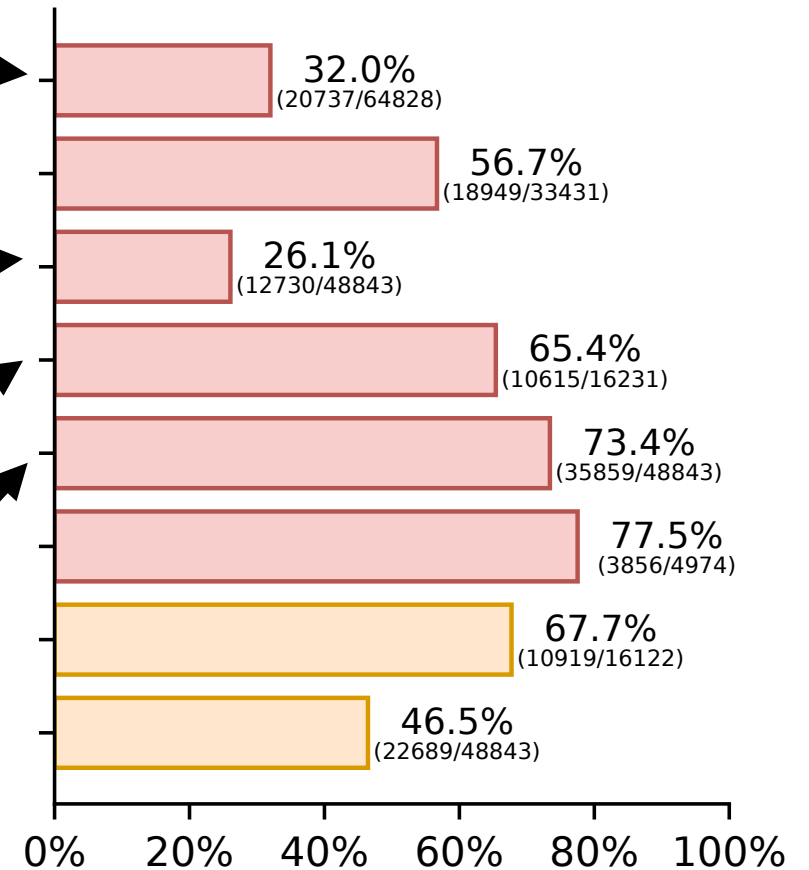
AA cookies set despite negative consent

Article 5(3) of the ePrivacy Directive



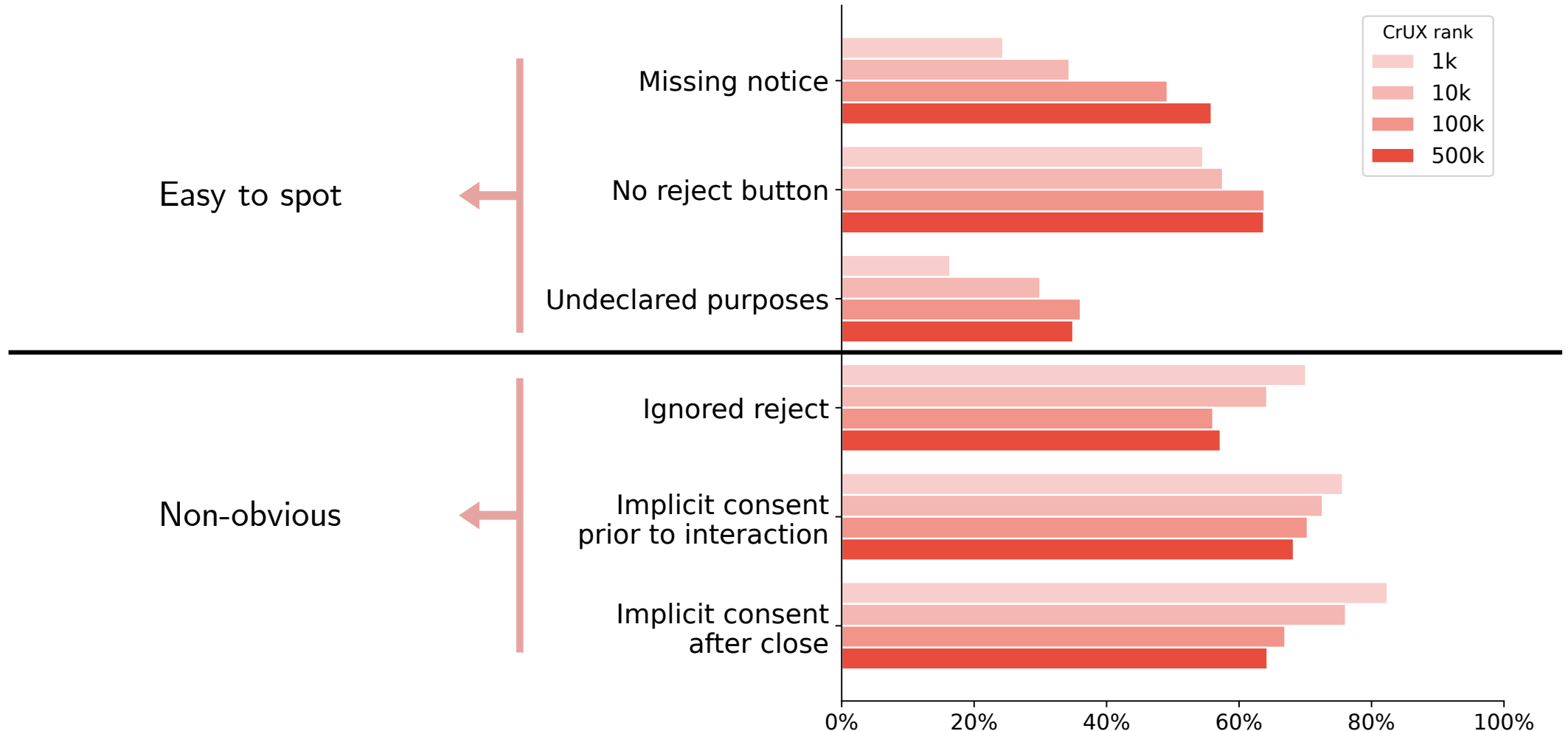
AA cookies set prior to user's consent

Article 5(3) of the ePrivacy Directive



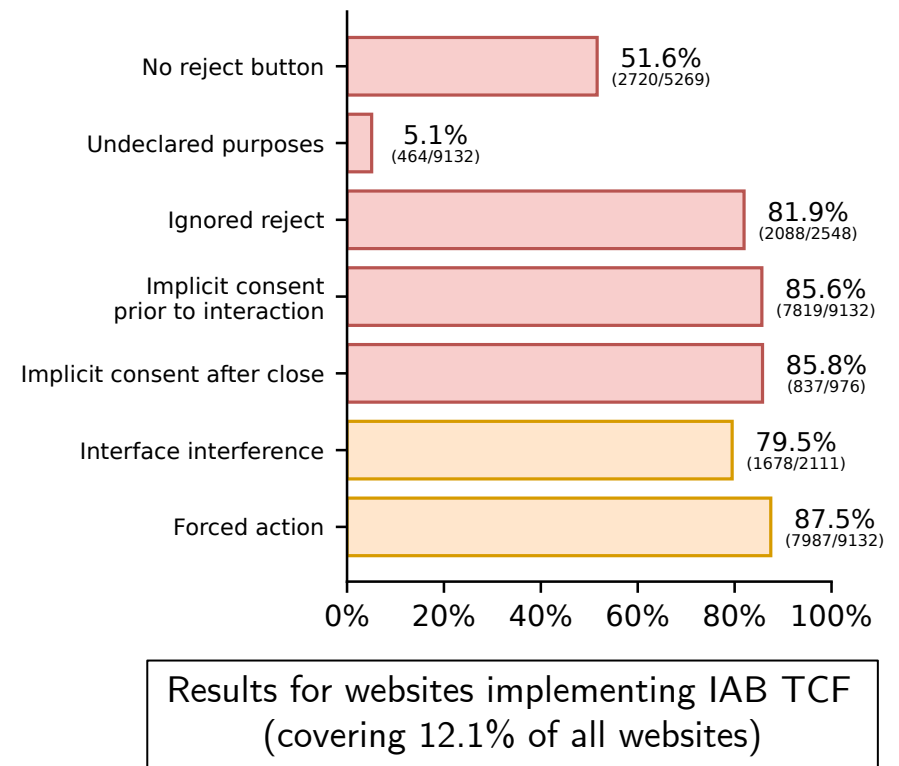
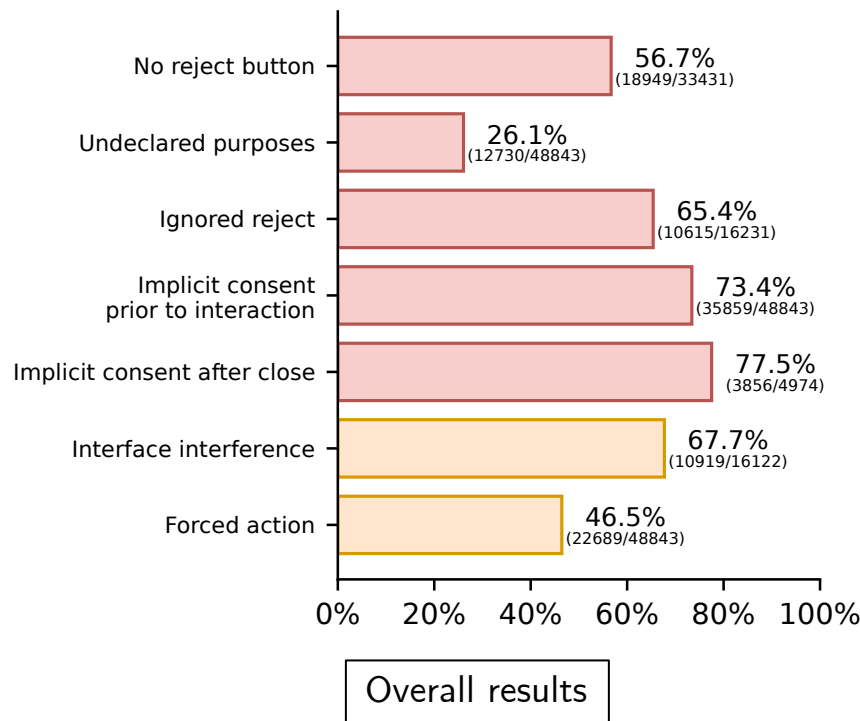
Results: Popularity-based analysis

We crawled an additional 20k websites in the ranks 100k and 500k sampled evenly across the same countries.



Results: Bias evaluation of previous studies

Websites implementing IAB TCF are more susceptible of “ignored reject” and “implicit consent” violations, as well as relying on dark patterns.



→ Matte et al. (2020) restricted their analysis to this type of websites.

Enforcement

- Our results show that consent is broken
- What's next? Enforcement!
- We started collaborations with the CNIL and NOYB.
- Our method for low false positive rates
 - As shown by manual end-to-end evaluation on 500 websites



Violation/Dark pattern	Precision	Recall
Missing notice	0.78	0.94
No reject button	0.92	0.65
Ignored reject	0.91	0.63
Implicit consent after close	0.95	0.58
Implicit consent prior to interaction	0.91	0.87
Undeclared purposes	0.98	0.91
Interface interference	1.00	0.72
Forced action	1.00	0.84

Conclusion

- A general method for detecting cookie notice violations.
- Analyzed 97k websites, addressing selection bias of previous studies.
- 72.2% of websites violate legal requirements.
 - Current focus on automated enforcement with CNIL

Thank you for your attention! Questions?

More information/source code:

<https://ahmedbouhoula.github.io/post/automated>

