New findings verification review Google Workspace for Education

16 May 2024
Sjoera Nas and Floor Terra

Introduction
At the request of SURF and SIVON, Privacy Company has verified whether Google had mitigated the 9 high risks for data subjects when using the paid versions of Google Workspace for Education, as those risks are described in the June 2021 update DPIA. The verification report concludes the 9 high risks are mitigated, or reduced to a low risk.

However, the verification report also mentions that the researchers saw five new processing operations that could potentially lead to high risks. These extra findings were discussed with Google, and all five of these risks have also been mitigated.

Below, the new findings are explained in five separate paragraphs.

Summary table of findings and (mitigated) risks
The table shows the five extra findings and the risks they could potentially cause for pupils, students and educators using Google Workspace for Education. The third column shows the mitigating measures taken by Google, while the fourth column shows the measures advised to schools and universities. If the measure mitigates the risk, the box is accordingly coloured green.

<table>
<thead>
<tr>
<th>Original risk(s)</th>
<th>Finding</th>
<th>Measure taken by Google</th>
<th>Measure recommended to schools and universities</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lack of purpose limitation Customer Data and Diagnostic Data</td>
<td>Google can show surveys to end users in the Core Services.</td>
<td>Google has confirmed it will not show surveys to K-12 users.</td>
<td>Schools and universities must choose the K-12 setting in their tenant, even if their users are 18+. Google has confirmed it will not check the actual age of Education tenants.</td>
</tr>
<tr>
<td>Lack of purpose limitation Customer Data</td>
<td>Introduction of Smart Features such as Smart Compose that use machine learning. Role of Google not clear/not documented. Users are nudged to enable the services.</td>
<td>Smart features are turned off by default for domains based in Europe, but users can still enable. Google has committed to process all Content and Diagnostic Data from Smart Features within the processor role (apart from the limited agreed legitimate business purposes) and results stay in customer domain.</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>No legal ground for Google and schools/universities</th>
<th>Use of the multi-purpose NID-cookie when logging in to a Workspace account and when looking up the Google Cloud Privacy Notice.</th>
<th>Google has informed SURF and SIVON that it does not use the NID-cookie set in the Workspace environment or when looking up the Google Cloud Privacy Notice for advertising purposes. When signed-out, users must provide consent for the use of the NID-cookie and other non-essential cookies for advertising on Google websites and on third party websites with Google advertising. Google improved the text of the cookie banner on the GCPN page to commit that the GCPN site does not use cookies for advertising.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lack of transparency</td>
<td>Google collects Content Data from spelling check in telemetry events and directly identifiable personal data (name/email address). The retention period of these events is unknown.</td>
<td>Google has explained, and published an explanation on 9 June 2023 why it is necessary to collect these data, and explained that the retention period is 30 days.</td>
</tr>
<tr>
<td>No legal ground for Google and schools/universities</td>
<td>New guidance from the EDPB about high risks of CSAM scanning.</td>
<td>Google has confirmed that it only scans for known CSAM, no use of machine learning/AI.</td>
</tr>
</tbody>
</table>

**Finding 1: Google asks survey questions in customer environment**

Immediately after creating the new Google Workspace for Education account in the university environment and viewing the default setting for personalised ads in Google's My Ad Center, the new user was presented with a series of survey questions from Google, about Google's ad business. In total, Google showed seven screens with questions about the use of data for personalised ads. See Figure 1 below.
In the K-12 environment, access to the Ad Centre was blocked. Privacy Company also did not see the survey questions in this environment. In reply to this finding, Google explained that the survey was shown in ‘My Ad Center’, which is not part of the Core Workspace for Education services.

As shown in Figure 1 above, the bottom of the survey screen contains hyperlinks to Google’s general privacy statement and terms and conditions. A privacy expert can infer from these references that Google permits itself to process the data for its own commercial purposes, acting as an independent data controller. For ordinary end users, this is not clear. Moreover, Google's questions are suggestive about the usefulness of personalised ads (with many more agree and don't know answers than disagree options). It appeared Google collected these answers to substantiate a legitimate interest as a legal basis for the processing of personal data for behavioural advertising. At the time, the EDPB ruled in sanction proceedings against Facebook/Meta.

2 For this report, Privacy Company only performed limited extra tests on top of the thorough testing for the Workspace verification report.
that the legal basis of contract (Article 6(1)(b) of the GDPR) could not be used for this purpose. On 27 October 2023, the EDPB also ruled out the use of necessity for a legitimate business interest as legal basis for behavioural advertising.

Figure 3: Google Feedback form as optional functionality

In reply to this finding, Google explained that it can present surveys in the Core Services, aimed at understanding how satisfied Workspace users are with the functionality of the specific Core service they are using, as part of the contractually agreed Legitimate Business Purposes. Google also explained that the questions about behavioural advertising were not commercial, as Google will not show commercial surveys in the Core Services. Google did confirm it can show other types of surveys in the core Workspace for Education services.

However, the agreed purpose was specifically limited to the use of a Feedback Form shown in Figure 3 that end users can optionally use to tell Google about their experience. As a mitigating measure, the DPIA initially concluded that schools and universities should tell their employees and students not to use this feature (not answer any Google surveys), as it results in disclosure of personal data to Google as a third party, outside of its processor role. In the discussion about these new findings, Google has explained it will not show surveys to users in K-12 environments.

Conclusion: schools and universities must and can mitigate high risk of surveys by switching to K-12 setting

If schools switch to K-12, Google won’t show surveys to the users, only the Feedback form. Google has confirmed it will not check the actual age of Education tenants.

A second reason why schools and universities must switch to the K-12 setting, is that a technical measure is always more effective than an organisational measure. For end users, the boundary between Core Services and Additional services is not very clear. It is logical and legitimate that users want to check the ad settings in their school or university Workspace account, without accidentally leaving the protected processor environment. The only technical measure to prevent this, is for schools and universities to change the nature of the tenant to K-12, because then access to the Ad Center is blocked.

Finding 2: Unclear role of Google in new Smart Features

During the verification process, Privacy Company noticed that Google had introduced a new service, Smart Compose. That service uses machine learning, Google explains: “Use Smart Compose in Google Documents and Presentations to write documents faster and easier. Smart Compose is also available for responses in Spreadsheets and Drawings. This feature uses machine learning to give suggestions as you type.” Google also explains that administrators can centrally disable this new service.

In the university Workspace for Education environment, Smart Compose was not turned on by the administrators for Gmail, but was still available to users because Google provides this functionality by default. Users are presented with a choice screen asking them to continue using the service (Continue with Smart Features). This is not entirely clear from Google’s public explanation. That explanation suggests that Smart Features would actually be off by default in the EU, and that the user has to make an effort to activate the services.

---

4 Google, Workspace administrator help, Turn Smart Compose on or off for users, URL: https://support.google.com/a/answer/10020933?hl=nl .
5 Google, Turn on or off Gmail, Chat and Meet smart features and personalization (Region specific), URL: https://support.google.com/a/answer/10095404?hl=en . See also Smart features & controls in Google products, URL: https://support.google.com/mail/answer/10007373?hl=en .

---
The new test user in the university Workspace for Education environment was given a choice screen in Gmail to continue with Smart Features, or to turn them off. For the 'off' option, Google gave a warning that turning the service off would result in the features and 'more' being disabled or degraded.

Figure 6: Question Google in Gmail to authorise Smart Features

When the test user clicked (anyway) on the bottom option 'Turn off smart features', Google displayed a new question asking if the user really wanted to turn off the service. See Figure 7 below.
Due to the unclear wording of the negative consequences of turning off, Google is steering towards consent (what does degradation of ‘more’ mean?). This is a form of pushing (nudging) on consent. Initially, it was not clear whether the data processing of these Smart Features was covered by the data processing agreement, and whether help with ‘writing’ is also a kind of spelling and grammar checker. Google has since published that it will not reuse the results of Workspace Feature Spelling and Grammar Checking for other customers. Google states in the public Workspace Data Protection Guide: “It is important to highlight that your Customer Data is not used to improve spelling & grammar services for other customers’ accounts.”

In reply to this finding, Google confirmed that the Smart Features are part of the Core Services, and governed by the data processor terms. According to Google, this ensures that there is no privacy risk in use of these features.

**Conclusion: no high risk resulting from use of the processor-service Smart Features**

Google has clarified its role as data processor for the processing of personal data from Smart Features, within the boundaries of the negotiated data processing terms. Dutch Workspace for Education users are free to use these services, without any additional data protection risks.

**Finding 3: Use of the NID-cookie**

When a Workspace for Education user logs-in, Google sets a NID-cookie without any user choice to accept or refuse the cookie. See Figure 8 below. Google also uses a NID-cookie when a Workspace user looks up

---

relevant legal information from Google in the Google Cloud Privacy Notice without offering the user an option to refuse the cookie. These two issues are discussed separately below.

Use of NID-cookie in Workspace applications

Figure 8: Screenshot of NID-cookie set when signing into a Google Workspace for Education account

The NID-cookie is also set when the school or university has set the tenant to the most privacy friendly K-12 setting.

In its public cookie statement, Google explains that it uses the NID-cookie for four purposes: (i) for functional purposes, to remember setting choices, (ii) analytics, (iii) to show ads to logged-out users, and (iv) for personalization.8

Google writes: “Some cookies and similar technologies are used to maintain your preferences. For example, most people who use Google services have a cookie called ‘NID’ or ‘_Secure-ENID’ in their browsers, depending on their cookies choices. These cookies are used to remember your preferences and other information, such as your preferred language, how many results you prefer to have shown on a search results page (for example, 10 or 20), and whether you want to have Google’s SafeSearch filter turned on. Each ‘NID’ cookie expires 6 months from a user’s last use, (...).”9

And:

“Google services also use ‘NID’ and ‘_Secure-ENID’ cookies on Google Search, and ‘VISITOR_INFO1_LIVE’ and ‘__Secure-YEC’ cookies on YouTube, for analytics.”10

And:

7 Screenshot made 16 November 2023, video of the process shared with Google on 8 January 2024.
8 Google Cookie Policy, undated, URL: https://policies.google.com/technologies/cookies?hl=en-US. Search for ‘NID’.
9 Idem.
10 Ibid.
“The ‘NID’ cookie is used to show Google ads in Google services for signed-out users,”\footnote{Ibid.}

And:

“And the ‘NID’ cookie enables personalized autocomplete features in Search as you type search terms. These cookies expire 6 months after a user’s last use.”\footnote{Ibid.}

It follows from these explanations from Google that the NID-cookie is a so-called quadruple-use cookie, with a unique identifier. Google can use this unique identifier to recognise a browser’s cookie jar until the NID cookie is deleted. Google currently uses the NID-cookie to show ads in Google services (such as YouTube and Search) for signed-out users. Earlier, Google also used the NID-cookie to show ads on third-party websites (AdSense for Search), but Google no longer uses the NID-cookie for this purpose.\footnote{Google, AdSense for Search and other Search Ads products to use new serving domains and deprecate ad personalization 3 November 2023, URL: \url{https://support.google.com/adsense/answer/14201307?hl=en}.} When Google uses the NID-cookie for the purpose of identifying a browser’s cookie jar across sites, it is a tracking cookie and, as online identifier, personal data. Based on the Dutch implementation of the ePrivacy Directive and based on the GDPR, the reading of such a cookie requires consent.

In reply to this finding, Google has explained it does not use the NID-cookie inside Workspace for Education for advertising purposes, as guaranteed in the public Google for Education Privacy & Security FAQs. In these FAQ’s Google guarantees that it will not show ads in Workspace for Education, or use any user data to make profiles to target ads.

\begin{figure}[h]
\centering
\includegraphics[width=0.5\textwidth]{figures/figure9.png}
\caption{Google’s promise of no advertising\footnote{Google Workspace for Education, in Dutch: \textit{Veelgestelde vragen over privacy en beveiliging}, URL: \url{https://edu.google.com/why-google/privacy-security/frequently-asked-questions/}.}}
\end{figure}

Google also explained it will not use the NID-cookie set in Workspace for Education for advertising purposes on external (third party) websites if the user has not provided consent for non-essential cookies. Google has technical and contractual measures to enable compliance by third parties using its advertising services with the legal consent requirement for non-essential cookies in the EU. Google’s EU user consent policy requires websites that carry Google advertising to obtain consent to both the use of the NID cookie for advertising purposes on the third-party website and (if they wish to use Google’s advertising services for the purpose of personalised advertising) the use of personal data for personalised advertising before calling Google’s advertising services. Google enforces its EU user consent policy through a combination of audits and technical
means. In May 2023, Google published a blog about stricter consent requirements for the use of its publisher-facing advertising services, including AdSense. 15

If a Workspace for Education user visits a Google website such as google.com (for which Google is a data controller), the user is automatically treated as a signed-out user and google.com will present a cookie consent banner to the user. If the user does not give consent for non-essential cookies, Google cannot use the NID-cookie set in the Workspace environment and the information associated with this NID-cookie to inform advertisements on google.com.

Similarly, Google's addition of the analytical purpose to the NID-cookie does not breach the privacy of classroom, as Google has contractually agreed not to use any Content or Diagnostic Data (including Website Data) from the Dutch Workspace for Education users for data analytics unless explicitly authorised by the agreement. If the NID-cookie is set when logging in to Workspace for Education, Google remains bound to these contractual limitations even when the end user logs out of Workspace and continues to surf to other websites where Google can read the NID-cookie.

As mentioned above, pursuant to the Dutch implementation of the ePrivacy Directive, only analytical cookies with no or little impact on privacy are exempted from the consent requirement. This exception does not apply to analytics across multiple websites.

Use of NID-cookie when reading Google Cloud Privacy Notice (GCPN)

Google also uses the NID-cookie when a user wants to read the legal information in the GCPN. Privacy Company observed the use of the NID-cookie during the sign-up process for a Google Workspace account. A test user clicked on the hyperlink to this text in the pop-up screen with the terms of the new account. At that point, Google displays a banner with a question about cookies.

Initially, Google's cookie pop-up included a choice between More information or OK, without an option to refuse. See Figure 10 below.

Figure 10: (Former) Google Cloud Privacy Notice with cookie pop-up

The user was treated as logged-out user by Google. This means Google acted as a data controller, rather than a processor, for the processing of personal data through cookies.

---

In reply to this finding, Google has confirmed that it only uses cookies on this GCPN page to offer and improve the quality of its services and to analyse traffic. Google does not use the NID cookie on the GCPN page for advertising purposes. Google launched a new version of this cookie banner on 1 February 2024, and a further redesigned banner on 13 May 2024. See Figure 11 below.

In the new version Google has changed the wording of the ‘OK’ button to ‘I understand’ to prevent the impression that Google asks for consent.

Google now mentions 3 processing purposes in this banner: (1) to provide services, (2) to improve the quality of the services, and (3) to analyse traffic. Google also explicitly excludes the use of the NID-cookie for advertising purposes (Deze website maakt geen gebruik van cookies voor advertenties). This purpose limitation is extremely important, because Google does mention advertising in the text shown to users when they click on ‘More Information’. In this paragraph Google describes it can use cookies and other unique identifiers for advertising purposes. See Figure 12 below.

Figure 11: Google new cookie banner on GCPN page since 13 May 2024 (Dutch version)

Figure 12: Google information about advertising purposes for unspecified cookies

Because Google does not provide a specific explanation in this text what cookies it uses for what purposes, with what retention period, but instead describes all possible purposes, and refers to its general (consumer) Privacy Policy, users cannot know what cookies and information Google will read, and for what purposes Google will use these personal data. The French data protection authority CNIL has issued (and upheld) two fines of 100 and 50 million euro for Google’s lack of offering an option to withdraw consent and the lack of a ‘refuse all cookies’ banner in Search and on YouTube. The CNIL explicitly warned that closure of this sanction procedure does not exclude future enforcement, in particular with regard to “the classification of cookies exempted from consent and the requirement to provide “clear and complete” information or the requirement to collect consent for each purpose.”

The purpose of ‘improving the quality’ of a service is too broad to qualify as strictly necessary processing, and hence, exemption from the consent requirement. In its Privacy Policy Google mentions it can use content data from users for improvement of features of its services: “we use your information to make improvements to our services — for example, understanding which search terms are most frequently misspelled helps us improve spell-check features used across our services.” In another example, Google combines ‘improvement’ with

17 Idem.
‘personalisation’ and ‘advertising’: “Depending on your account settings, your activity on other sites and apps may be associated with your personal information in order to improve Google’s services and the ads delivered by Google” [underlining added by Privacy Company].” These examples show that at least one of these 3 purposes is not well-defined.

Due to the improved wording in the banner, and the exclusion of advertising purposes, the ‘quality improvement’ purpose has been narrowed down.

**Conclusion: high risk mitigated**

Google has redesigned the cookie banner on the GCPN page to commit to users it won’t use the NID-cookie for advertising purposes. This measure mitigates the high risk of the lack of a legal ground for government organisations due to Google’s lack of information.

**Finding 4: Content Data in telemetry**

When the original DPIA and the update DPIA were written, it was not (yet) possible to separately examine the Telemetry Data collected by Google. One of the agreed improvement measures was the development of a telemetry viewer, in addition to more comprehensive public documentation on the nature of the personal data collected with telemetry. For this verification review, Privacy Company used Google's new telemetry viewer (DIT, abbreviation of Diagnostic Information Tool). Inspection of these data revealed that Google collects Content Data from the spelling and grammar checker, and some directly identifiable personal data in some telemetry events (the name of an accessory). This was surprising, but Google explained convincingly why the data collection about use of the spelling and grammar checker was necessary (see below for more details). In response to this report, Google also found the collection of the (directly identifying) name of the accessory was not necessary, and Google agreed to stop collecting that name. As described in the 13 November 2023 version of the Update DPIA report, in June 2023 Google has significantly expanded its documentation about Telemetry Data. The information page about the Diagnostic Information Tool (DIT) contains two sources of information: a general description with non-exhaustive examples of telemetry events, and detailed examples with the full payload of a representative browser telemetry event for each Workspace Core Service.

Google does not offer Workspace users and administrators a choice between local or cloud-based spelling services: depending on the (internet) conditions, Google chooses a local or cloud spelling service.

In a separate section about Spelling and grammar suggestions in the telemetry data, Google publicly explains that collection through telemetry is necessary because Google allows users to work with documents and emails offline. The only way for Google to still collect input about the chosen spelling suggestions is to send that information to Google via telemetry messages the moment a user comes back online.

Google also explains that it does not retain these data longer than 1 month: “These logs are temporary in nature, held for a maximum of 30 days. They are collected, anonymised or pseudonymised, and aggregated to provide the information needed to operate the spell and grammar check tool. The document itself does not retain a record of spelling suggestions and interactions.”

Google is contractually bound to collect and process only those personal data that are necessary for the agreed purposes.

---

19 Google, Diagnostic Information Tool, URL: [https://support.google.com/a/answer/12830816](https://support.google.com/a/answer/12830816)

20 Ibid, ‘Understand your search results’

21 Ibid.

Conclusion: High risk mitigated by transparency on telemetry data

By improving the transparency about this data processing, and by committing to a short retention period, Google has mitigated the high risk of a lack of control for the schools and universities.

Finding 5: Scanning for child abuse material (CSAM)

When the update DPIA was drafted, it was unclear what Google’s policy was with regard to automated scanning of Workspace content for child sexual abuse material (CSAM).

While there is a temporary exception in the ePrivacy Directive to the ban on the scanning of the content of communications, a positive legal obligation has not yet been adopted at European level. In May 2022 the European Commission published a proposal for a Regulation that would legally require designated providers to scan for CSAM (i.e. almost a year after the Update DPIA was finalised). The EDPB and the EDPS have adopted a critical joint opinion on this proposal on 28 July 2022.

The proposed Regulation has been scrutinised by civil society, human rights experts, cryptographers and many online service providers because the proposal required providers of operating systems and providers of communication services to build a back-door in the services that would allow for the scanning of data before they were encrypted with End-To-End-Encryption (E2EE). Aside from this controversial obligation for Client Side Scanning, the proposed Regulation also proposes that scanning may only be done by a European centre of expertise (yet to be established), in cooperation with national hotlines, and thus not by and in cooperation with the US NGO that Google is now working with: NCMEC.

The data protection authorities distinguish between the scanning of ‘known’ material with hashes, and detection of as yet unknown material. The data protection authorities warn that the infringement of scanning for unknown material on the fundamental rights of data subjects (such as protection of privacy, confidentiality of communications and freedom of expression) is neither necessary nor proportionate. “As regards the necessity and proportionality of the envisaged detection measures, the EDPB and EDPS are particularly concerned when it comes to measures envisaged for the detection of unknown child sexual abuse material (‘CSAM’) and solicitation of children (‘grooming’) in interpersonal communication services. Due to their intrusiveness, their probabilistic nature and the error rates associated with such technologies, the EDPB and EDPS consider that the interference created by these measures goes beyond what is necessary and proportionate [emphasis added by Privacy Company].

The EDPB and EDPS call on the Commission to explicitly prohibit voluntary scanning in the new Regulation: “Given that the detection obligations introduced by the Proposal would apply only to recipients of detection orders, it would be important to make clear in the text of the proposed Regulation that the voluntary use of technologies for the detection of CSAM and solicitation of children remains permitted only inasmuch as it is allowed under the e-Privacy Directive and GDPR. This would entail, for instance, that providers of number-independent interpersonal communications services would be prevented from using such technologies on a voluntary basis, unless this would be permitted under the national laws transposing the e-Privacy Directive, in accordance with Article 15(1) of the e-Privacy Directive and the Charter.” The DPAs also indicate that they see no room for scanning on a voluntary basis after the Regulation comes into force: “For instance, the EDPB and

---

25 Idem, Summary, p. 5.
26 Idem, par. 20.
EDPS recommend making clear that the proposed Regulation would not provide for a lawful basis for the processing of personal data for the sole purpose of detecting online child sexual abuse on a voluntary basis.”

In light of this explanation from the DPAs, SURF and SIVON asked Google to explain that it only scans for known CSAM, and does not deploy machine learning (not itself, but also not indirectly through a partner like NCMEC) to discover new cases. Google unequivocally confirmed that it does not scan for unknown CSAM.

**Conclusion: High risk mitigated by statement Google**

Google’s statement that it only scans for known CSAM, and that it does not (indirectly) use machine learning to discover new materials mitigate the potential high risk of scanning for unknown materials.

---

27 Idem, par. 21.