# Personal Identifiable Information Protection: Sensitive Data Discovery test

The test is commissioned by Kaspersky and performed by AV-TEST GmbH.
All rights to the test results and the report belong to Kaspersky.
Date of the report: August, 9th , 2021

## Executive Summary

In June 2021 AV-TEST performed a Sensitive Data Discovery Test of the two solutions: Kaspersky Endpoint Security Cloud and Microsoft Office 365 E5. The test aimed to reveal the solutions' capabilities to detect German Personal Identifiable Information (PII) in documents uploaded to an Office365 account and inform the system admin about the potential security issues.

The set of test cases was created in the test lab and included six different types of German PII (identity card number, passport number, social insurance number (SIN), tax identification number (TIN), residence permit number, driving license number) plus three different types of Banking Cards. These numbers were generated in accordance with related legal and technical regulations and were represented in several valid styles before being integrated into readable texts, and saved in ten popular office document types with valid and invalid file extensions. Additionally, False Positives levels of the solutions were measured.

During the test, the products were expected to detect PII and prevent false positive detections on non-PII numbers. The best results were achieved by Kaspersky with a 100% detection rate for all seven tested German PII types while avoiding any false positive detections. Regardless of any appropriate changes to the allowed PII delimiters or customization of file extension, the product managed to detect all files with variations.

The build-in Office 365 protection managed to detect just under 25% of the sensitive data shared in three out of the seven tested PII. Additionally, several non-sensitive data files were falsely detected. Only 40% of the previously detected files were detected once the file extensions were changed. When changing the allowed PII delimiter the detection rate of the previously detected files dropped to one in five.

For more detailed information please refer to section 'Test Results' of the report.

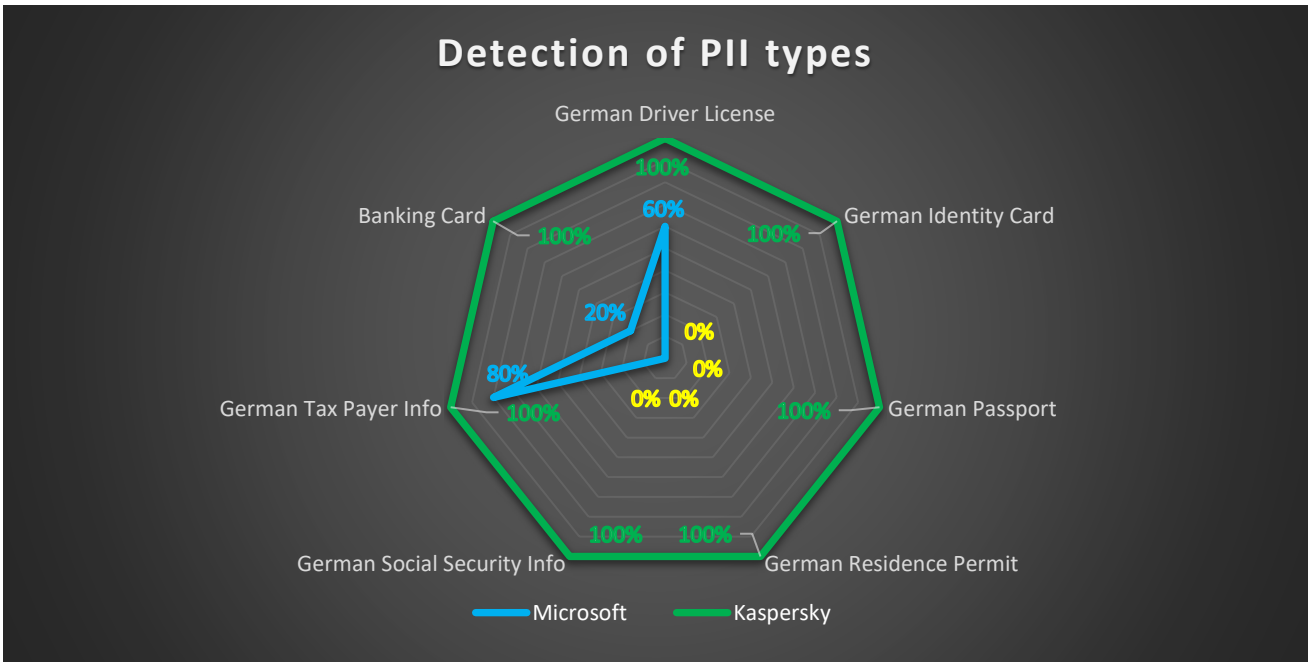



Figure 1 Overall results detection of Sensitive Data for Office 365 and Kaspersky Endpoint Security Cloud.

# Content

# 1. Introduction

For the last decade, methods of accessing company resources have increased steadily. People use their personal devices to access company networks including mobile phones, tablets and laptops and they use them from various locations, for example their private homes or summer residences, public transport, cafes or parks. This has become possible and is accelerated with the ever-increasing growth and availability of cloud services. The manifold capabilities of these cloud services also come with different types of risks. Companies embracing this technological advancement face different risks such as the intentional or unintentional sharing of critical information. Cloud services allow storage of- and collaboration on- critical information such as sensitive customer and employer data which may need protection on a higher level. Organizations can find and use different means to address this threat, but they often require specific knowledge to detect this kind of undesired sharing.

One of the most well-known and most widely used cloud services is Microsoft Office 365. Office 365 comes with its own set of security and privacy settings and tools. It also allows the extension of the cloud service through third-party apps such as Kaspersky Endpoint Security Cloud.

This test is focused on testing the first step of every security operation – identifying personal data worth protecting and comparing the Office 365 internal data protection solution with the third party app by Kaspersky.

The test was performed in May-June, 2021. The report was created on August, 9th , 2021

# 2. Test Methodology

- The test has been carried out as described in this document.
- The report contains all results initially requested for testing. No data was excluded from the report.
- The results were independently reached.

## 2.1. Tested Products

The tested products are listed below. Details on product configurations are aailable in Exhibit1.

| Product Name | Version | Product configuration in the test |
|---|---|---|
| **Kaspersky Endpoint Security Cloud** | 11.0.1.90 | Kaspersky Endpoint Security Cloud license with Data Discovery feature enabled, the solution is connected to Office365 account without Office 365 DLP policies set up. |
| **Microsoft Office 365 E5** | 4.12.17007.18022 | E5 license with the default configuration and DLP policies, which enable detection of all supported sensitive data categories |

## 2.2. Test scenarios

The testing includes a True Positive test (Test scenario 1) which evaluates how well the products detect the potential leak of sensitive information. The false positive test (Test scenario 2) examines if, for the sake of detecting of all potential leaks, any unaffected files are also tagged to include sensitive data. The final test scenarios (Test scenario 3.1 and 3.2) address the capabilities of the solutions to detect variations in the documents, introduced either through different kinds and positions of delimiters or through different file extensions.

### 2.2.1. Test scenario #1: True Positives, German PII without modifications

This scenario reviews a collection of 70 True Positive samples, which are expected to be detected by the DLP solutions.

The set is the basis to create modified sample sets for scenarios 3.1 and 3.2.

The following criteria are applied to create the collection:

- Six types of German PII and one general PII type are selected for the assessment:
  - German identity card (ID) number
  - German passport number (general format without control digit)
  - German social insurance number (SIN)
  - German tax identification number (TIN)
  - German residence permit number
  - German driving license number
  - Credit/debit card numbers systems (American Express, Mastercard, Visa)
- In accordance with official legal and technical regulations about the formats of these PII types, a set of numbers was generated for each mentioned PII type, and then one number per PII type was randomly selected
- Surrounded with human-readable text, each selected PII identifier was inserted into one of 10 files types of the most popular document formats: TSV, CSV, PDF, RTF, MS Office (xlsx, docx, ppts), OpenOffice (odt, odp, ods).

### 2.2.2. Test scenario #2: False Positives

This scenario reviews the collection of 70 False Positive samples, which are expected NOT to be detected by the DLP solutions.

This collection consists of:

- 30 samples taken are generated in the same manner as described in test scenario 1.
- Generated four numbers per German PII type and two numbers per each of three Banking Cards
- The PII numbers are modified so that they do not pass the validity check
- Each number was surrounded with human-readable text and inserted into files of the same file type previously used for the TP samples
- 50 randomly chosen files with German text but without valid German PII numbers.

### 2.2.3. Test scenario #3: German PII with legitimate modifications

The purpose of this scenario is to verify that samples from test scenario 1 which are detected by the both DLP solutions are still detected after they have been legitimately modified.

This set consists of two subsets described below.

#### *2.2.3.1. Test scenario #3.1: use of legitimate modifications*

The initial plan to create this subset was as follows:

- Select PII types, which do legitimately support modifications: these are the five types: **German Identity Card**, **German Passport**, **German Residence Permit**, **German Tax Payer Info** and **Banking Cards**:
- Select samples of these types from test scenario 1, which are detected by both solutions to check if each solution still detects them after modification.
- Legitimately modify the samples by one of each supported delimiters: Space delimiter, No-Break delimiter, Tab delimiter.

Execution of Test scenario 1 showed that the solution from Microsoft did detect only three types of PII out of seven: **German Driver License**, **German Tax Payer Info** and **Banking Cards**, and did not detect the remaining four types: **German Identity Card**, **German Passport** (without control digit), **German Residence Permit** and **German Social Security Info**.

The lack of detection by Office 365 in Test scenario 1 made it not possible to include all required PII in this test scenario as previously described in the test methodology. To be able to validate Kaspersky's' capabilities despite the Office 365 misses, it was decided to add two samples of each

of these three PII types (**German Identity Card**, **German Passport**, **German Residence Permit**), by random selection from the collection of scenario #1.

The final collection for Test scenario #3.1 consists of:

- 16 variations of **German Tax Payer Info** + 2 variations of **Banking Cards**, which were detected by both Kaspersky and Microsoft in test scenario 1
- Additionally: 2 samples of **German Identity Card** + 2 samples of **German Passport** + 2 samples of **German Residence Permit** to check Kaspersky's capabilities.

#### *2.2.3.2. Test scenario #3.2: use of modified filename extensions*

The initial plan to create this subset was as follows:

- Select samples of these types from the Test scenario 1, which are detected by both solutions, to check if each solution still detects them after extension modification.
- Modify filename extension of the selected samples to one of four variants: alternate extension, no extension, extension "TXT", extension "EX_".

Similarly to the situation described in Test scenario 3.1, usage of the planned methodology turned out non-possible due to the lack of detections by Office 365 of four PII types. To address this situation, it was decided to randomly select one sample per each of these four missed PII types (**German Identity Card**, **German Passport**, **German Residence Permit, German Social Security Info**) from the collection of Test scenario 1, to use for the Test scenario #3.2

The final collection for Test scenario #3.2 consists of:

- 6 samples of **German Driver License** + 8 samples of **German Tax Payer Info +** 2 samples of **Banking Cards**, which were detected by both Kaspersky and Microsoft in test scenario 1 True positive.
- 1 German Driver License with the original xlsx extension was not tested for an alternative extension because there is no extension that would provide compatibility with the format.
- Additionally: 1 sample of **German Identity Card** + 1 sample of **German Passport** + 1 sample of **German Residence Permit +** 1 sample of **German Social Security Info** to check Kaspersky's capabilities.

## 2.3. Test Execution

Test execution consisted of several steps:

- An organization account was set up in Office365 with one employee account inside. This is done for two separate organization accounts, to assess the capabilities of the solution separately.
- The collection of test samples described in Test Cases was uploaded to Office365 OneDrive of the employee's account for both organizations.
- All the test samples were shared outside the organization by creating a "share link".
- A minimum of 24 hours was given to the solutions to analyze the content of the uploaded files.
- Results of detection were scrutinized the following way:
  i. For Microsoft Office 365 - in the policy Tips in OneDrive interface.
  ii. For Kaspersky Endpoint Security Cloud - in the product interface.

## 2.4. Scoring

When a file with sensitive data inside is detected by a product, one point is awarded to the product. If a file with sensitive data inside is not detected by a product, no points are awarded to the product.

# 3. Test Results

## 3.1. Test scenario 1: True Positives, German PII without modifications

The following graph shows the detection rate of Microsoft Office 365 and Kaspersky Endpoint Security Cloud.

The Kaspersky product had no issues finding all tested documents which included the different types of sensitive German data. Office 365 had its best detection rate for German Tax payer info for which it achieved a respectable 80%. Second most files detected contained German driver licenses. Only 20% detection was achieved for Banking Card details. Office 365 didn't detect any German- Identity cards, passports, residence permits or social security numbers.

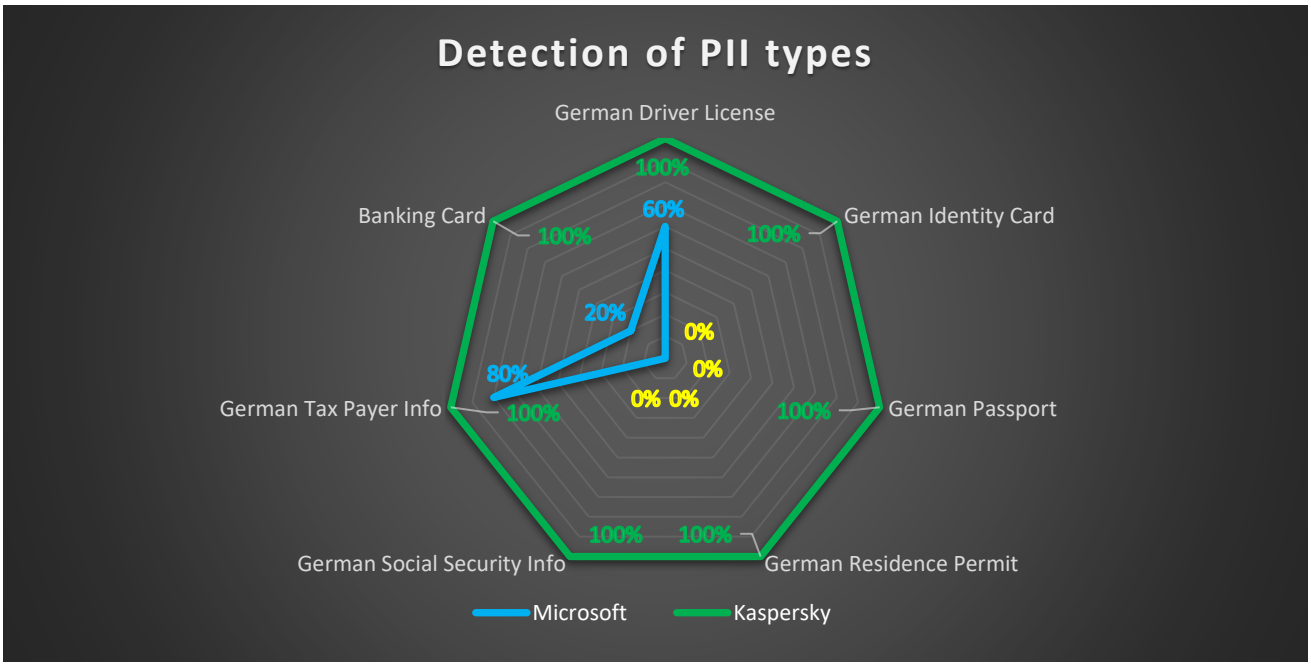



Figure 2 True Positive detection results comparing the onboard Microsoft Office 365 security feature with Kaspersky Endpoint Security Cloud. The more the value is, the better.

## 3.2. Test scenario 2: False Positives

Figure 3 show the number of files falsely detected as the ones with sensitive data. Kaspersky didn't flag any clean files in the test. Office 365 tagged four out of 70 files, which is a 6% detection rate of clean files as potential including sensitive data.

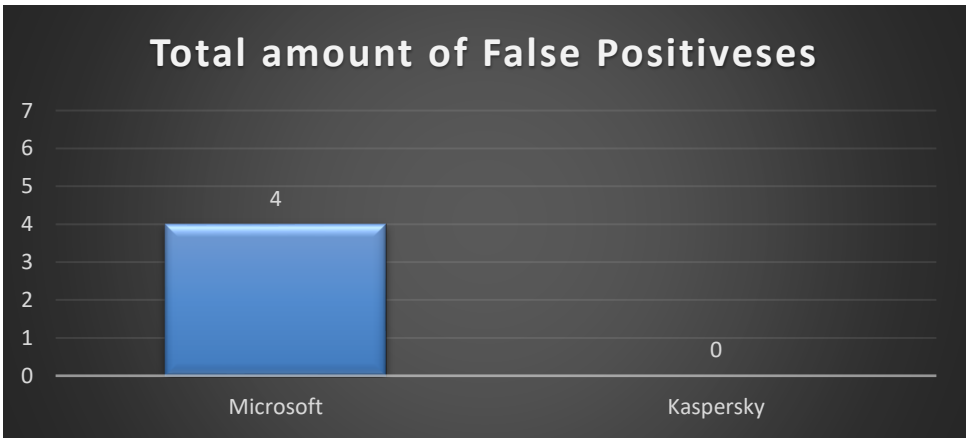
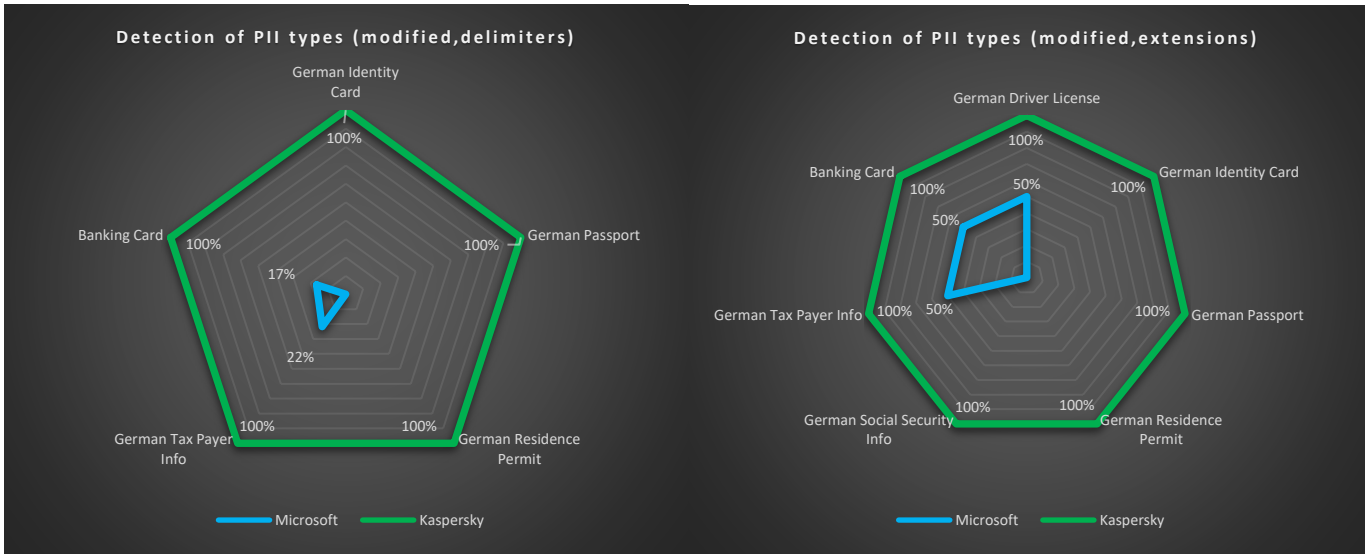



Figure 3 Number of detections for documents without any sensitive data inside. The less the value is, the better.

## 3.3. Test scenario 3: German PII with legitimate modifications

As seen in figures and 4 and 5 below, the Kaspersky product has no issues with detecting PII in files after their modification by delimeters, or delimeter positions, or delimiter types or alternative file extensions. Kaspersky detected all samples as it did in Test scenario 1 regardless of delimiter changes or alternate extensions.

As in Test scenario 1 Office 365 didn't detect any German Identity Card, Passport, Residence Permit and Social Security Info. For the modified delimiters of previously detected documents, the detection rate is only around one in five. When it comes to file extensions Office 365 doesn't support documents without extensions or with alternative "unknown" extensions. The overall detection rate for the documents types which Office 265 detects with a changed extensions is 50%.




Figures 4 and 5 respectively, outlay the detection of the documents with the modified delimiters for the sensitive data and the altered file name extensions, comparing the two tested products. The more the value is, the better.

It should be noted that:

- Initially the solution from Microsoft did detect only three types of PII out of 7: **German Driver License**, **German Tax Payer Info** and **Banking Cards**, and did not detect the other 4 types: **German Identity Card**, **German Passport** (without control digit), **German Residence Permit** and **German Social Security Info**. So the modifications into the numbers for filename extensions did not help to show a non-zero Detection rate.
- PII types **German Driver License** and **German Social Security Info** are excluded from subtest #3.1 since these PII types do not support any types of delimiters

## Summary

The test results outline the real capabilities of the two solutions to discover German Personal Identifiable Information being stored in Microsoft Office365.

For the prevention of accidental leakage of sensitive data, Microsoft Office 365 provides very basic protection only. A desolate three of the seven tested PIIs types were somehow detected. The average detection rate for those PIIs is just over 50% or 23% when considering all tested PII types in Test scenario 1. Detection of four files with no PII inside gave a 6% of false positive rate. When it comes to changed files the detection further decreases. There are common statutes on where delimiters can be placed in PIIs and what such delimiters might be. Unfortunately, Office 365 doesn't recognize most of them and detects only one in five of these altered office files. The picture was not improved in scenario with alternative filename extensions. If there are no filename extensions or uncommon filename extensions that are often done by security settings are used, there are zero detections. Only for commonly known- and for TXT file- extensions did Office 365 find anything and that only for every other file.

Kaspersky Endpoint Security Cloud on the other hand shows that it is capable of detecting sensitive data in all cases faced. The solution detects all tested 70 samples for all seven PII types including credit card types in Test scenario #1, while producing zero wrong detections in test scenario #2. Kaspersky was completely unimpressed by any changes to delimiter or filename extension changes we confronted it with. A 100% detection rate was achieved for all three types of delimiters and all common delimiter positions in the PII which were tested. A 100% detection was also achieved for all four different types of extension changes applied.

Basing on the test results, seal "APPROVED Data Loss Prevention" was awarded to Kaspersky.