# Topic 2: Personal Data Identification

## With the rapid digitization of documents, emails, and online communication, protecting personal data has become more important than ever. Organizations face increasing pressure to comply with data privacy regulations like GDPR, CCPA, and HIPAA, which mandate the identification and management of personally identifiable information (PII). One of the most common and sensitive types of PII is the **name of a person**. Your challenge is to find all personal data inside of databases and if wished would mask/log them.

## Key Objectives

* Finds all the names of people in one or multiple databases and inserts the result into one of the results tables fitting your team. Each team will have 1 table to put their results into.

* One can also provide a specific name, like “Klaus Müller”, to look for in the database. The result is also inserted into the table
* The AI should ask the User for each entry if it should just log it, or if it should be masked. If the user selects mask, the AI should encrypt the data with a specific key and save that key to the table as well. It should be possible to decrypt the data with that key again. If the user selects log, then it just gets written to the table.

## Scope and Constraints

* For your solution the use of any AI is allowed, open- and close-sourced
* You may modify the results table to include more metadata as you see fit
* If a name is similar enough to the one, I search for, it should also be found. For example, “Max Mustermann” & “Max Musterman”.

## Provided

* You will receive a 150, - CHF voucher for the use of any AI service of your choice. If you want to make use of this, provide [Diellore Halitaj (mailto: diellore.halitaj@dm-international.com)](mailto: diellore.halitaj@dm-international.com) with the bill for your AI Service and we will reimburse you afterwards.
* You will receive access to a database, reachable with:
  + Server: sql-lakeside-server.database.windows.net
  + Port: 1433
  + TEAM CREDENTIALS:
    - Team Alpha: hackathon\_alpha
    - Team Beta: hackathon\_beta
    - Team Gamma: hackathon\_gamma
    - Team Delta: hackathon\_delta
    - Team Epsilon: hackathon\_epsilon
  + Password: will be provided individually per Team by OTS
* On this AZURE SQL Server you will find 4 Databases:
  + ECC60jkl\_HACK
  + ORACLE\_EBS\_HACK
  + AdventureWorks2019
  + Jde920\_demo
* Use these databases to search for personal data.
* On this Database server you will also see the “Results” database, here create the table for your results

## Expected Result:

* If you search for the name of “Paul Jonas” you should get at least 3 Results
* If you search for the name of “Paula Erickson” you should get at least 3 Results
* You may use these to verify your results.

## Evaluation Criteria

* Creativity and innovation in the AI-powered solution.
* Accuracy of the results

## How to submit your results

On Friday you will receive a USB flash drive from **Diellore Halitaj** where you must transfer all your code, presentation and results to for grading. Either **Heinrich Krupp** or **Diellore Halitaj** will collect these USB flash drives at 13:00 o’clock.

How to get help

If you have any questions regarding the topics or the agenda of this Hackathon our on-side staff **Heinrich Krupp** will be happy to provide you with any help you might need. Feel free to contact him.