# Project Description for Degree Projects Department of Computer Science

|  |  |  |
| --- | --- | --- |
| Student: | Felix Böhlin | fb222ur@student.lnu.se |
| Supervisor: | Your supervisor at the university (if you already have been appointed a supervisor) | |
| External company: | TietoEvry | |
| External supervisor: | mattias.holmgren@tietoevry.com | |

# Background

Data privacy and protection is every year becoming more and more important for individuals and organizations. Individuals seek data privacy to protect their personal information from unauthorized access, identity theft, and maintain control over their digital footprint. For organizations, data privacy is essential not only for compliance with laws like GDPR but also for maintaining customer trust and safeguarding against data breaches, which can lead to significant financial and reputational damage. The General Data Protection Regulation (GDPR) goes through how to correctly handle personal information in a secure manner. However, manually going though data and erasing any sensitive information is impractical and resource heavy. Therefor is this project aimed to make that possible with the help of machine learning to create a model that can accurately anonymize sensitive data in unstructured text.

# Problem formulation

The problem for this project is to investigate the anonymization of sensitive information within free-text inputs. To come up with an AI algorithm that in a cost-effective way removes any sensitive data which could identify individuals.

# Expected result

To create a model that can with high success rate detect and encrypt sensitive information in free-text inputs. Achieving a balance of protecting individuals’ data while preserving the readability of the text.

# References

György Szarvas, Richárd Farkas, Róbert Busa-Fekete, “State-of-the-art Anonymization of Medical Records Using an Iterative Machine Learning Framework”, Journal of the American Medical Informatics Association, 2007, url: <https://academic.oup.com/jamia/article/14/5/574/720728?login=true>

Payne, Brad. “Privacy protection with AI: Survey of data-anonymization techniques.", 2020, url: <https://bradpayne.ca/privacy-protection-with-ai-survey-of-data-anonymization-techniques/>