# The project

This project idea came to be because I was copying WWII letters on my computer for a website of a specific battalion that served during the war. They have a page with all the letters from the battalion for the family and others to read those letters so they won’t disappear over time. For this website I had to read and type every word separately. The problem with this is that this takes way too long and is quite frankly a very boring task to do if you have a lot of letters.

The objective of this project is to create a platform where you can upload a photograph of a letter or even a shopping list that you want digitalised, the AI will recognise handwritings and rewrite the letter for you to save in a text file.

# Societal impact

This project has an impact on society so in this part of the research I explain the good impact it can have, but also look into the worst case scenario of creating such an AI model.

## Good impact

* It can digitalise everything you write down
* If you write down something on paper you can lose it, if it is digitalised and stored safely you can’t as easily
* Digitalising things like archives or letters gets done way faster
* This could impact companies who still struggle with things like an financial archive on paper which they have to keep for several years by law

## Worst case scenario

* The worst thing I can think of is that when the AI gets so good at reading things it can be used on things like camera’s to spy on people or something similar
* If everything is digitalised but not stored safely, everyone can see it

# Dataset

## The EMNIST Dataset

The EMNIST dataset is a set of handwritten character digits derived from the NIST Special Database 19 and converted to a 28x28 pixel image format.

There are six different splits provided in this dataset. A short summary of the dataset is provided below:

* EMNIST ByClass: 814,255 characters. 62 unbalanced classes.
* EMNIST ByMerge: 814,255 characters. 47 unbalanced classes.
* EMNIST Balanced: 131,600 characters. 47 balanced classes.
* EMNIST Letters: 145,600 characters. 26 balanced classes.
* EMNIST Digits: 280,000 characters. 10 balanced classes.
* EMNIST MNIST: 70,000 characters. 10 balanced classes.