

**Students:** Nadav Spitzer, 302228275. Lior Frenkel, 204728315

## **Part 1**

### **POS task –**

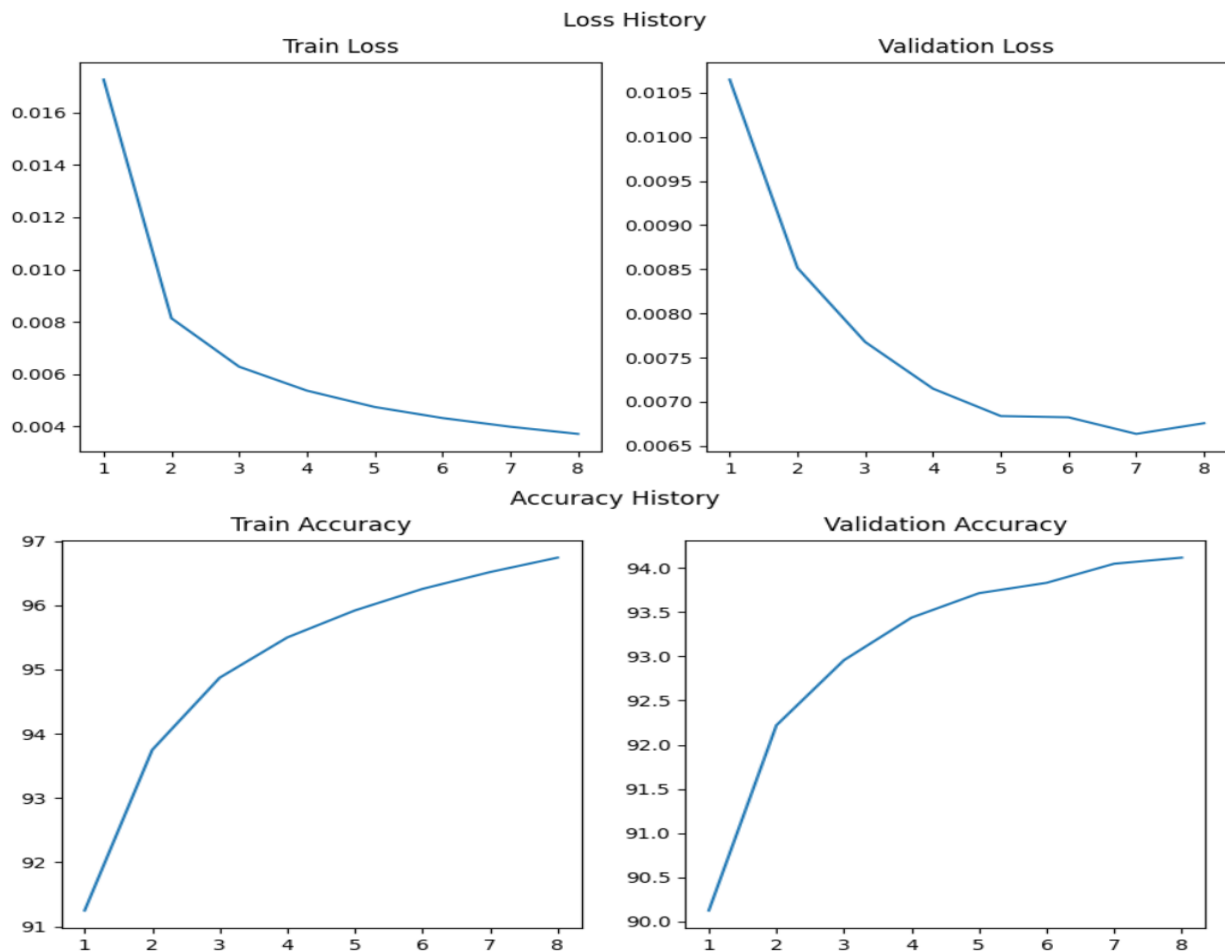
#### **1. Parameters:**

- Number of epochs: 8
- Learning rate:  $1e-3$
- Batch size: 32
- Hidden layer size: 150
- Optimizer: Adam
- Dropout with 0.5 probability

#### **2. Results:**

- Train loss: 0.0037138
- Train accuracy: 96.742%
- Validation loss: 0.006758
- Validation accuracy: 94.116%

#### **3. Graphs:**



## NER task –

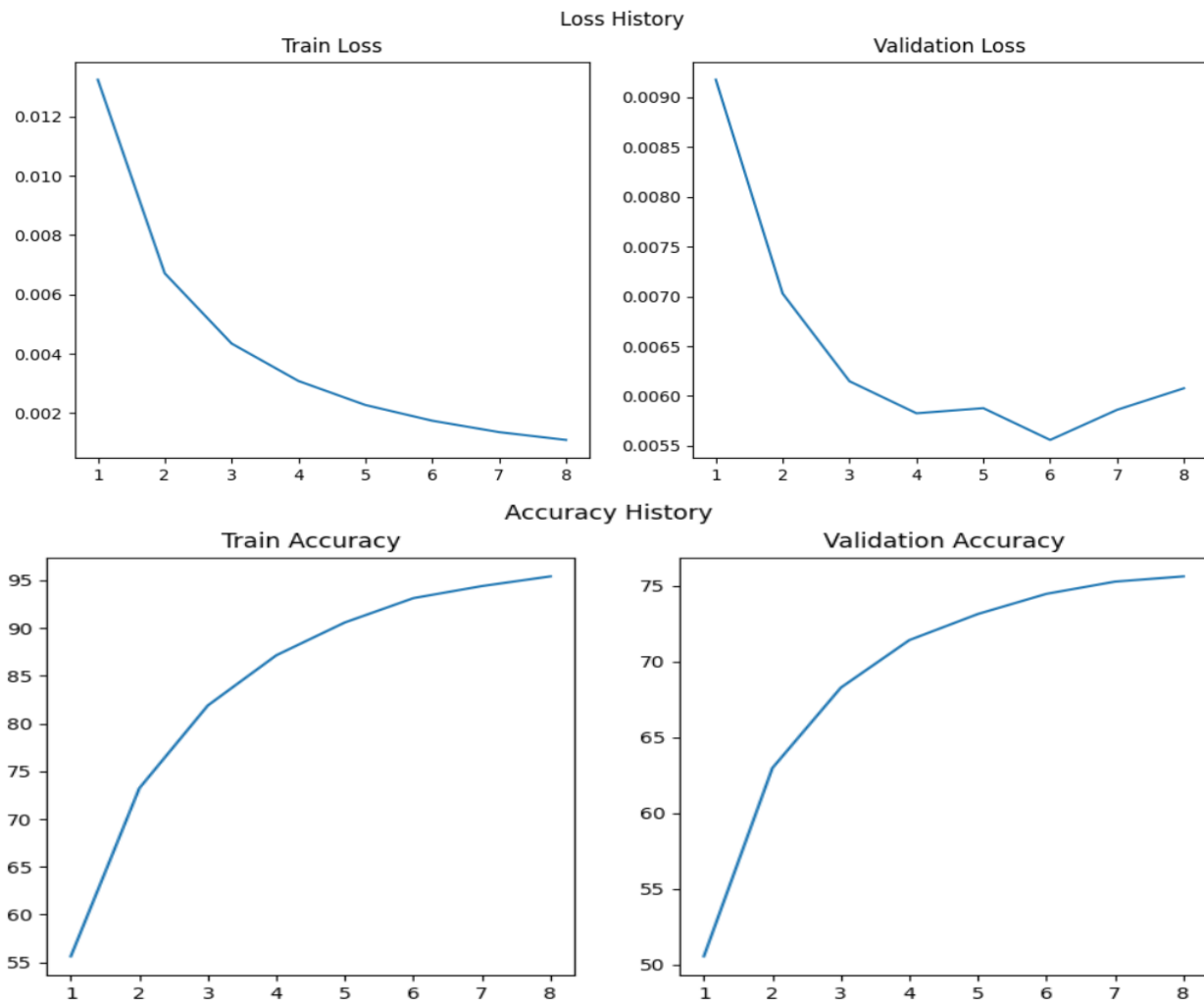
### 1. Parameters:

- Number of epochs: 8
- Learning rate:  $1e-3$
- Train Batch size: 32
- Dev Batch size: 32
- Hidden layer size: 150
- Optimizer: Adam
- Dropout with 0.5 probability
- Weight decay:  $1e-4$

### 2. Results:

- Train loss: 0.00109529
- Train accuracy: 95.401%
- Validation loss: 0.006078
- Validation accuracy: 75.63%

### 3. Graphs:



**Considerations:** Each sentence in the data (sequence of words between blank rows) were padded with special words for start and end ('<s>', '</s>') in order to achieve window of size 5 when the required word is in the middle.

Words that were seen in the DEV or TEST files but not in the TRAIN were given a special word of 'UUUNKKK' and a tag of UNSEEN.