

HAAG NLP Summarization Week 11

Michael Bock

October 2024

1 Slack Questions

What did you accomplish this week?

- Trained a model which achieved 90% accuracy on the DNO issues dataset, but only a .22 F1-Score.
- Explored Class imbalance in the DNO dataset.

What are you planning on working on next?

- Nathan sent me a link to multi-label classification which isn't a problem I'd heard of before. He sent code of huggingface `AutoForSequenceClassification` models running on this problem so I think I will change the dataset to use that.
- A class imbalance is harming the F1 score of the model. There are many ways to fight this and it doesn't work for all issues. I can apply data augmentation techniques to get more data on the data I have. I can also use sample weighting or oversampling to try to improve the class balance, but this may cause data to leak between the train and validation sets.
- Nathan and I's long term plan is to remove the 51 classifiers and train with only one output layer. It should produce similar results, but will allow us to just use the huggingface automodel instead of needing to make our own. One issue here is that the latent spaces for each class are together. To separate them out, we had the idea of using a Class Activation Mapping to find the section of the latent space which corresponds to each class. I also have the idea of keeping only specific training clusters and training on those.

What is blocking you from progressing?

- Due to the class imbalance, I will need to download more data from UPenn, but I don't have S3 Credentials. I've emailed the database administrator asking for credentials, but I haven't heard back yet.

2 Abstract

An important paradigm of natural language processing consists of large-scale pre-training on general domain data and adaptation to particular tasks or domains. As we pre-train larger models, full

fine-tuning, which retrains all model parameters, becomes less feasible. Using GPT-3 175B as an example – deploying independent instances of fine-tuned models, each with 175B parameters, is prohibitively expensive. We propose Low-Rank Adaptation, or LoRA, which freezes the pre-trained model weights and injects trainable rank decomposition matrices into each layer of the Transformer architecture, greatly reducing the number of trainable parameters for downstream tasks. Compared to GPT-3 175B fine-tuned with Adam, LoRA can reduce the number of trainable parameters by 10,000 times and the GPU memory requirement by 3 times. LoRA performs on-par or better than fine-tuning in model quality on RoBERTa, DeBERTa, GPT-2, and GPT-3, despite having fewer trainable parameters, a higher training throughput, and, unlike adapters, no additional inference latency. We also provide an empirical investigation into rank-deficiency in language model adaptation, which sheds light on the efficacy of LoRA. We release a package that facilitates the integration of LoRA with PyTorch models and provide our implementations and model checkpoints for RoBERTa, DeBERTa, and GPT-2 at <https://github.com/microsoft/LoRA>.

2.1 Brief Analysis

I've discussed the fine tuning paradigm a few weeks ago when I introduced BERT. Many models use supervised pretraining to generally understand language before they are fine tuned to specialize on a task. BERT's fine tuning isn't really fine tuning is the way I usually talk about it. Where I come from (Computer Vision), fine tuning is when you freeze all the layers of a model except the last layer and then you train with a low learning rate. BERT's fine tuning is literally resuming training with a low learning rate and old weights. It's more like restoring the model from a checkpoint and changing its output. But with LLMs like LLaMa, it can be very expensive to fine tune models this way. If we could fine tune models like we do in computer vision with freezing the whole model and training only one layer, that would make fine tuning LLMs cheaper.

I'm a little surprised the author's got away with given the LoRA algorithm an actual name, because LoRA is really just normal fine tuning with 2 common tricks applied. The first is what I was talking about with freezing the layers. LoRA freezes almost all of the parameters of LLaMa and then I can train an output layer to accomplish whatever task I want. Using this, I was able to make my training pipeline only train about 75% of the parameters of LLaMa 3.2. This makes it feasible for me to fine tune LLaMa on Pace with an H-100. Meta needed to use 100,000 H-100s to train LLaMa and I only needed 1 to fine tune. The other trick LoRA does is similar to pruning. Pruning a model involves removing layers until performance is no longer good. LoRA does this by reducing the rank of linear layers in transformers. This also saves parameters from training because the new low rank layers have less parameters than the original LLaMa model.

I used LoRA in my training script below. It allowed me to fine tune with much less resources than what LLaMa was initially trained with.

3 Scripts and Code Blocks

models.py

```
1 from datasets import load_dataset, Dataset
2 from transformers import AutoTokenizer
3 from transformers import AutoModelForSequenceClassification, AutoModel
4 from transformers import TrainingArguments, Trainer, BitsAndBytesConfig
5 import numpy as np
```

```

6 import evaluate
7 import os
8 import pandas as pd
9 from torch import nn
10 from torch.utils.tensorboard import SummaryWriter
11 from labels import DNO_ISSUES
12 from peft import prepare_model_for_kbit_training, AutoPeftModel
13 from peft import LoraConfig, get_peft_model
14 import datetime
15 import os
16
17 model_name = "meta-llama/Llama-3.2-1B"
18
19 def print_trainable_parameters(model):
20     """
21     Prints the number of trainable parameters in the model.
22     """
23     trainable_params = 0
24     all_param = 0
25     for _, param in model.named_parameters():
26         all_param += param.numel()
27         if param.requires_grad:
28             trainable_params += param.numel()
29     print(
30         f"trainable params: {trainable_params} || all params: {all_param} ||
31         trainable%: {100 * trainable_params / all_param}"
32     )
33
34 #model_name = "allenai/led-base-16384"
35 class LongFormerClassification(nn.Module):
36     def __init__(self, n_issues = 51, device = 'cuda'):
37         super().__init__()
38         self.heads = []
39         bnb_config = BitsAndBytesConfig(load_in_4bit=True, bnb_4bit_use_double_quant
40 =True, bnb_4bit_quant_type="nf4", bnb_4bit_compute_dtype=torch.bfloat16)
41         self.backbone = AutoModel.from_pretrained(model_name, device_map={"":0},
42 quantization_config=bnb_config, token = os.environ['HF_TOKEN'])
43         self.backbone.gradient_checkpointing_enable()
44         self.backbone = prepare_model_for_kbit_training(self.backbone)
45         config = LoraConfig(r=8, lora_alpha=32, target_modules="all-linear",
46 lora_dropout=0.2, bias="none", task_type="FEATURE_EXTRACTION")
47         self.backbone = get_peft_model(self.backbone, config)
48         print_trainable_parameters(self.backbone)
49         for _ in range(n_issues):
50             self.heads.append(nn.Sequential(nn.Linear(self.backbone.config.
51 hidden_size, 1), nn.Sigmoid()))
52             self.heads[-1].to(device)
53
54     def forward(self, input_ids, attention_mask):
55         x = self.backbone(input_ids = input_ids, attention_mask = attention_mask,
56 output_attentions = False, output_hidden_states = False, return_dict = False)[0]
57         sequence_lengths = torch.eq(input_ids, input_ids[-1]).int().argmax(-1) - 1
58         sequence_lengths = sequence_lengths % input_ids.shape[-1]
59         sequence_lengths = sequence_lengths.to(x.device)
60         x = x[torch.arange(input_ids.shape[0], device=x.device), sequence_lengths]
61         outputs = []
62         for head in self.heads:
63             outputs.append(head(x).squeeze(-1))

```

```

58     return tuple(outputs)
59     #return x
60
61 metric = evaluate.load("accuracy")
62
63 def compute_metrics(eval_pred):
64     logits, labels = eval_pred
65     predictions = np.argmax(logits, axis=-1)
66     return metric.compute(predictions=predictions, references=labels)
67
68 training_args = TrainingArguments(output_dir="/home/hicel/mbock9/scratch/
    tutorial_runs/", eval_strategy = 'epoch', save_total_limit = 5,
    load_best_model_at_end = True, save_strategy = "epoch", num_train_epochs = 50)
69
70 df = pd.read_csv("dno_labels.csv", sep=",").dropna()
71
72 dataset = Dataset.from_pandas(df).train_test_split(test_size=0.2)
73
74 tokenizer = AutoTokenizer.from_pretrained(model_name, token = os.environ['HF_TOKEN'
    ])
75 #tokenizer = AutoTokenizer.from_pretrained("FacebookAI/xlm-roberta-base")
76 #tokenizer.add_special_tokens({'pad_token': '[PAD]'})
77 tokenizer.pad_token = tokenizer.eos_token
78
79 max_seq_length = tokenizer.model_max_length # or model.config.
    max_position_embeddings
80 print(f"Model's maximum sequence length: {max_seq_length}")
81
82 def tokenize_function(examples):
83     examples['Text'] = [e.replace('\n', ' ') for e in examples["Text"]]
84     return tokenizer(examples['Text'], padding="max_length", max_length=
    max_seq_length//10, truncation=True, return_tensors='pt')
85
86 tokenized_dataset = dataset.map(tokenize_function, batched=True)
87 tokenized_dataset = tokenized_dataset.remove_columns(["Text"])
88 tokenized_dataset = tokenized_dataset.remove_columns(["Name"])
89 #tokenized_dataset = tokenized_dataset.rename_column("label", "labels")
90 tokenized_dataset.set_format("torch")
91
92 #small_train_dataset = tokenized_dataset["train"].shuffle(seed=42).select(range
    (1000))
93 #small_eval_dataset = tokenized_dataset["test"].shuffle(seed=42).select(range(1000))
94 #trainer = Trainer(model = model, args = training_args, train_dataset=
    small_train_dataset, eval_dataset=small_eval_dataset, compute_metrics =
    compute_metrics)
95 #trainer.train()
96 from torch.utils.data import DataLoader
97
98 #small_train_dataset = tokenized_dataset["train"].shuffle(seed=42).select(range(1))
99 #small_eval_dataset = tokenized_dataset["test"].shuffle(seed=42).select(range(1))
100 train_dataloader = DataLoader(tokenized_dataset["train"], shuffle=True, batch_size
    =8)
101 eval_dataloader = DataLoader(tokenized_dataset["test"], batch_size=8)
102 from torch.optim import AdamW
103
104 import torch
105
106 device = torch.device("cuda") if torch.cuda.is_available() else torch.device("cpu")

```

```

107
108 model = LongFormerClassification(device = device)
109
110 optimizer = AdamW(model.parameters(), lr=5e-5)
111
112 from transformers import get_scheduler
113
114 num_epochs = 10
115 num_training_steps = num_epochs * len(train_dataloader)
116 lr_scheduler = get_scheduler(
117     name="linear", optimizer=optimizer, num_warmup_steps=0, num_training_steps=
118     num_training_steps
119 )
120 from tqdm.auto import tqdm
121
122 criterion = nn.BCELoss()
123
124 now = datetime.datetime.now()
125 logdir = now.strftime('/home/hice1/mbock9/scratch/runs/tensorboard/%Y%m%d_%H%M%S')
126 savedir = now.strftime('/home/hice1/mbock9/scratch/runs/checkpoints/%Y%m%d_%H%M%S')
127 writer = SummaryWriter(logdir, flush_secs = 1)
128 os.makedirs(savedir)
129
130 max_saves = 5
131
132 for epoch in range(num_epochs):
133     model.train()
134     train_losses = {k: 0 for k in DNO_ISSUES}
135     train_total_losses = 0
136     train_n_correct = {k: 0 for k in DNO_ISSUES}
137     train_n_total = {k: 0 for k in DNO_ISSUES}
138     train_cm = {k: np.zeros((2, 2)) for k in DNO_ISSUES}
139
140     for batch in tqdm(train_dataloader, total = len(train_dataloader)):
141         batch = {k: v for k, v in batch.items()}
142         inputs = {"input_ids" : batch["input_ids"].to(device), "attention_mask":
143         batch["attention_mask"].to(device)}
144         outputs = model(**inputs)
145         labels = {k: batch[k].to(device) for k in DNO_ISSUES}
146         losses = {k: criterion(output, labels[k]) for output, k in zip(outputs,
147         DNO_ISSUES)}
148         total_loss = sum(losses.values())
149         #loss = outputs.loss
150         total_loss.backward()
151
152         optimizer.step()
153         lr_scheduler.step()
154         optimizer.zero_grad()
155
156         for loss, issue in zip(losses.values(), DNO_ISSUES):
157             train_losses[issue]+=loss.item()
158             train_total_losses += total_loss.item()
159
160         for output, issue, issue_present in zip(outputs, labels, labels.values()):
161             train_n_total[issue]+=output.shape[0]
162             #print((output.cpu().detach().numpy() > 0.5)==issue_present.cpu().numpy
163             ())

```

```

160         train_n_correct[issue]+=np.count_nonzero((output.cpu().detach().numpy()
161 > 0.5)==issue_present.cpu().numpy())
162         preds = (output.cpu().detach().numpy() > 0.5).astype(np.int32)
163         trues = issue_present.cpu().numpy().astype(np.int32)
164         for pred, true in zip(preds, trues):
165             train_cm[issue][true, pred] += 1
166
167     val_total_losses = 0
168     val_losses = {k: 0 for k in DNO_ISSUES}
169     val_n_correct = {k: 0 for k in DNO_ISSUES}
170     val_n_total = {k: 0 for k in DNO_ISSUES}
171     val_cm = {k: np.zeros((2, 2)) for k in DNO_ISSUES}
172
173     for batch in tqdm(eval_dataloader, total = len(eval_dataloader)):
174         with torch.no_grad():
175             batch = {k: v for k, v in batch.items()}
176             inputs = {"input_ids" : batch["input_ids"].to(device), "attention_mask":
177 batch["attention_mask"].to(device)}
178             outputs = model(**inputs)
179             labels = {k: batch[k].to(device) for k in DNO_ISSUES}
180             losses = {k: criterion(output, labels[k]) for output, k in zip(outputs,
181 DNO_ISSUES)}
182             total_loss = sum(losses.values())
183             val_total_losses += total_loss.item()
184
185             for loss, issue in zip(losses.values(), DNO_ISSUES):
186                 val_losses[issue]+=loss.item()
187             for loss, issue in zip(losses.values(), DNO_ISSUES):
188                 val_losses[issue]+=loss.item()
189
190             for output, issue, issue_present in zip(outputs, labels, labels.values()):
191                 val_n_total[issue]+=output.shape[0]
192                 val_n_correct[issue]+=np.count_nonzero((output.cpu().detach().numpy
193 () > 0.5)==issue_present.cpu().numpy())
194                 preds = (output.cpu().detach().numpy() > 0.5).astype(np.int32)
195                 trues = issue_present.cpu().numpy().astype(np.int32)
196                 for pred, true in zip(preds, trues):
197                     val_cm[issue][true, pred] += 1
198
199     torch.save({
200         'epoch': epoch,
201         'model_state_dict': model.state_dict(),
202         'optimizer_state_dict': optimizer.state_dict(),
203     }, os.path.join(savedir, f'{epoch}.pt'))
204     torch.save(model.state_dict(), os.path.join(savedir, f'{epoch}_fullmodel.pt'))
205
206     if epoch > max_saves:
207         os.remove(os.path.join(savedir, f"{epoch-max_saves}.pt"))
208         os.remove(os.path.join(savedir, f"{epoch-max_saves}_fullmodel.pt"))
209
210     writer.add_scalar("Loss/train", train_total_losses/len(train_dataloader), epoch)
211     writer.add_scalar("Loss/val", val_total_losses/len(eval_dataloader), epoch)
212     total_train_correct = 0
213     total_val_correct = 0
214     total_train_tp = 0
215     total_train_fp = 0

```

```

213 total_train_fn = 0
214 total_val_tp = 0
215 total_val_fp = 0
216 total_val_fn = 0
217 total_train = 0
218 total_val = 0
219 for issue in DNO_ISSUES:
220     writer.add_scalar(f"Accuracy/train/{issue}", train_n_correct[issue]/
train_n_total[issue], epoch)
221     writer.add_scalar(f"Accuracy/val/{issue}", val_n_correct[issue]/val_n_total[
issue], epoch)
222     writer.add_scalar(f"F1/train/{issue}", train_cm[issue][1, 1]/(train_cm[issue
][1, 1] + .5 * (train_cm[issue][0, 1] + train_cm[issue][1, 0])), epoch)
223     writer.add_scalar(f"F1/val/{issue}", val_cm[issue][1, 1]/(val_cm[issue][1,
1] + .5 * (val_cm[issue][0, 1] + val_cm[issue][1, 0])), epoch)
224     total_train_correct += train_n_correct[issue]
225     total_val_correct += val_n_correct[issue]
226     total_train += train_n_total[issue]
227     total_val += val_n_total[issue]
228     total_train_tp += train_cm[issue][1, 1]
229     total_train_fp += train_cm[issue][0, 1]
230     total_train_fn += train_cm[issue][1, 0]
231
232     total_val_tp += val_cm[issue][1, 1]
233     total_val_fp += val_cm[issue][0, 1]
234     total_val_fn += val_cm[issue][1, 0]
235     writer.add_scalar(f"TotalAccuracy/train/", total_train_correct/total_train,
epoch)
236     writer.add_scalar(f"TotalAccuracy/val/", total_val_correct/total_val, epoch)
237     writer.add_scalar(f"F1/train", total_train_tp/(total_train_tp + 0.5 * (
total_train_fp + total_train_fn)), epoch)
238     writer.add_scalar(f"F1/val", total_val_tp/(total_val_tp + 0.5 * (total_val_fp +
total_val_fn)), epoch)
239
240     print(f'Epoch: {epoch}')
241     print(f'Train Loss: {train_total_losses/len(train_dataloader)}')
242     print(f'Val Loss: {val_total_losses/len(eval_dataloader)}')
243     for issue in DNO_ISSUES:
244         print(f"Accuracy/train/{issue}:", train_n_correct[issue]/train_n_total[issue
])
245         print(f"Accuracy/val/{issue}: ", val_n_correct[issue]/val_n_total[issue])
246     for issue in DNO_ISSUES:
247         print(f"F1/train/{issue}:", train_cm[issue][1, 1]/(train_cm[issue][1, 1] +
.5 * (train_cm[issue][0, 1] + train_cm[issue][1, 0])))
248         print(f"F1/val/{issue}: ", val_cm[issue][1, 1]/(val_cm[issue][1, 1] + .5 * (
val_cm[issue][0, 1] + val_cm[issue][1, 0])))
249
250     print(f'Total Train Accuracy: {total_train_correct/total_train}')
251     print(f'Total Val Accuracy: {total_val_correct/total_val}')
252     print(f"F1 Train", total_train_tp/(total_train_tp + 0.5 * (total_train_fp +
total_train_fn)))
253     print(f"F1 Val", total_val_tp/(total_val_tp + 0.5 * (total_val_fp + total_val_fn
)))

```

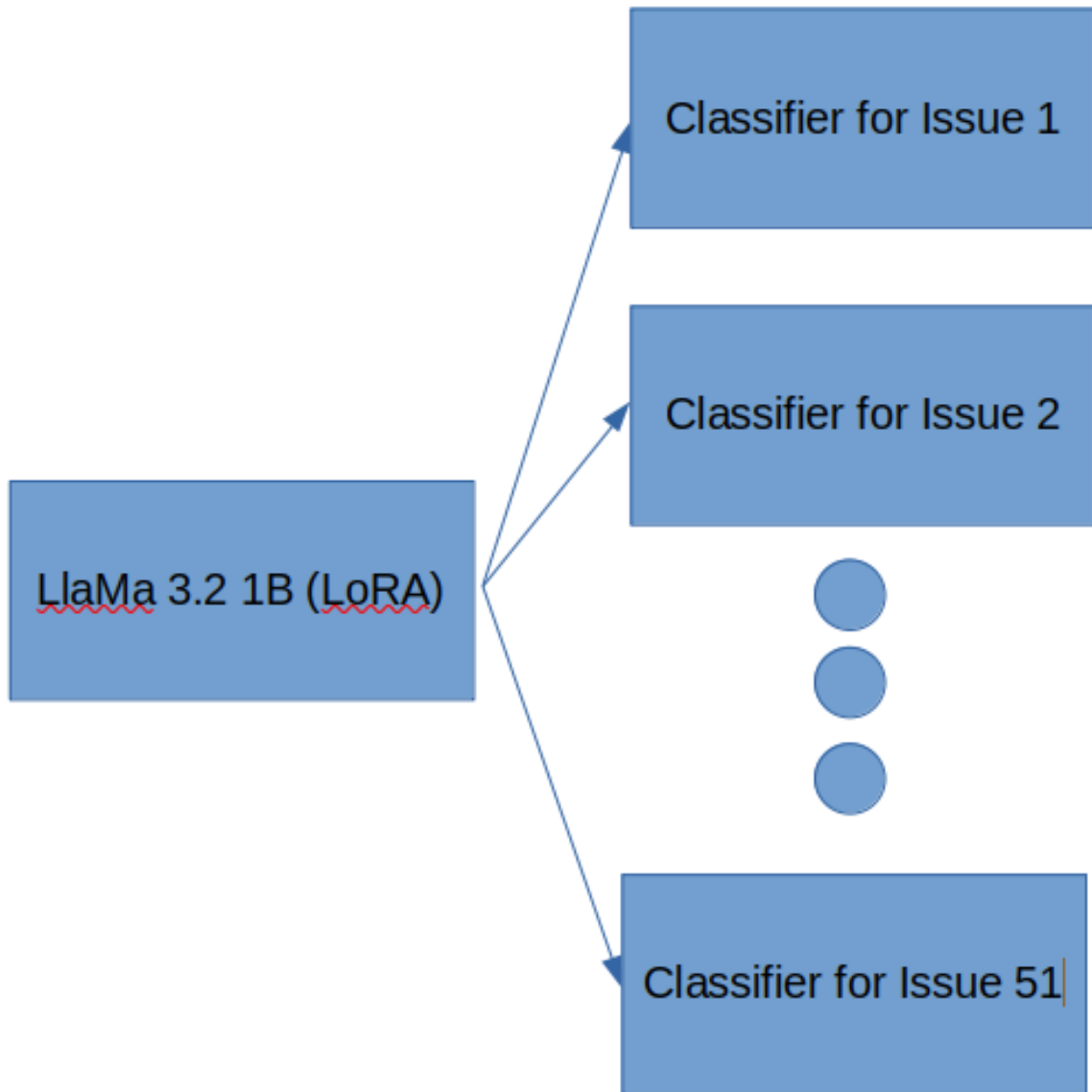


Figure 1: Flowchart of how this model works. It uses LLaMa 3.2 as a base and then feeds that into 51 binary classifiers; 1 for each issue. This model achieves an accuracy of 90% on the validation set, but these results are not reliable because the F1 score is only .2, indicating the accuracy is only high due to class imbalance.

4 Documentation

I now understand the DNO classification problem as a multi-label classification problem. The model may identify several different DNO classes as being present in a case. In contrast to multi-class classification, which only allows one label to be selected per case, this model can accept many labels per class. The accuracy of the model is 90% on the validation set, as shown in Figure 3. However, the F1 Score remains low at only 0.22 on the validation set, as shown in Figure 4. This is due to class imbalance. Figure 5 show how many true positives each issue has. Certain issues have many true positives, but many have very few true positives. Some even have no true positives. I can improve this using oversampling or data augmentation, but the best solution is to add more cases to the dataset. Some classes, such as Cyber, have no true positives at all. With this dataset, it is not possible to create a model for the Cyber class because we have no examples of cases that contain the Cyber issue.

Figure 6 shows how often each issue is present in the dataset. Notice how accuracy rises as the rate of positives goes down. This is because the model learns that some issues happen very rarely, so it simply says the issue is not present every time, which achieves high accuracy but isn't what we want. The green line is at $y = 0.2$. 0.2 is the percent of examples which go to the validation set. Given that split, any class whose blue bar is below the line will statistically have no positives in the validation set. If there are no true positives in the validation set, the F1 Score can only be 0, which will occur when the validation accuracy is less than 100%, or NaN, which occurs in classes that have a validation accuracy of 100%. To see why this is, realize that the numerator of F1 Score (Figure 2) is true positives. When the accuracy is 100% and there are no true positives, there will also be no false positives or false negatives. Therefore $TP, FP, FN = 0$, $\frac{TP}{TP + \frac{1}{2}(FP + FN)} = 0/0$. If the accuracy is not 100% there is at least one false positive or false negative, so $FP + FN > 0$, and F1 score is 0. In Table 1 and Table 2. In these tables, those issue which are above the green $y = 0.2$ line have reliable validation F1 scores, but all the other classes have F1 scores of near 0 or NaN because few or no true positives made it into the validation set.

$$\text{F1 Score} = \frac{TP}{TP + \frac{1}{2}(FP + FN)}$$

Figure 2: F1 Score formula

5 Script Validation(Optional)

<https://youtu.be/uY1JMHIVAEY>

```
1 -----
2 Begin Slurm Prolog: Oct-30-2024 20:43:14
3 Job ID:      898362
4 User ID:    mbock9
5 Account:    coc
6 Job name:   HGX_H100_Example
7 Partition:  coe-gpu
8 -----
9 Lmod has detected the following error: The following module(s) are unknown:
10 "python/3.6"
11
12 Please check the spelling or version number. Also try "module spider ..."
13 It is also possible your cache file is out-of-date; it may help to try:
14 $ module --ignore_cache load "python/3.6"
15
16 Also make sure that all modulefiles written in TCL start with the string
17 #Module
18
19
20
21 /var/lib/slurm/slurmd/job898362/slurm_script: line 11: pip: command not found
22 /usr/bin/python: No module named pip
23 Model's maximum sequence length: 131072
24 Map: 100%|| 660/660 [00:09<00:00, 73.21 examples/s]
25 Map: 100%|| 166/166 [00:02<00:00, 61.02 examples/s]
26 trainable params: 5636096 || all params: 754911232 || trainable%: 0.746590560729662
27 0%|          | 0/83 [00:00<?, ?it/s]'use_cache=True' is incompatible with gradient
   checkpointing. Setting 'use_cache=False'.
28 /home/hice1/mbock9/.local/lib/python3.9/site-packages/torch/utils/checkpoint.py:460:
   UserWarning: torch.utils.checkpoint: please pass in use_reentrant=True or
   use_reentrant=False explicitly. The default value of use_reentrant will be
   updated to be False in the future. To maintain current behavior, pass
   use_reentrant=True. It is recommended that you use use_reentrant=False. Refer to
   docs for more details on the differences between the two variants.
29 warnings.warn(
30 100%|| 83/83 [57:44<00:00, 41.74s/it]
31 0%|          | 0/21 [00:00<?, ?it/s]/home/hice1/mbock9/.local/lib/python3.9/site-
   packages/torch/utils/checkpoint.py:90: UserWarning: None of the inputs have
   requires_grad=True. Gradients will be None
32 warnings.warn(
33 100%|| 21/21 [03:57<00:00, 11.31s/it]
34 /storage/ice1/5/9/mbock9/law-data-design-vip/ner/llamas/model.py:223: RuntimeWarning
   : invalid value encountered in scalar divide
35 writer.add_scalar(f"F1/val/{issue}", val_cm[issue][1, 1]/(val_cm[issue][1, 1] + .5
   * (val_cm[issue][0, 1] + val_cm[issue][1, 0])), epoch)
36 /storage/ice1/5/9/mbock9/law-data-design-vip/ner/llamas/model.py:248: RuntimeWarning
   : invalid value encountered in scalar divide
37 print(f"F1/val/{issue}: ", val_cm[issue][1, 1]/(val_cm[issue][1, 1] + .5 * (val_cm
   [issue][0, 1] + val_cm[issue][1, 0])))
38 Epoch: 0
39 Train Loss: 14.930663648858127
40 Val Loss: 14.006731986999512
41 Accuracy/train/Bodily injury: 0.9363636363636364
42 Accuracy/val/Bodily injury: 0.9156626506024096
```

43 Accuracy/train/Didn't settle when should have: 0.8833333333333333
44 Accuracy/val/Didn't settle when should have: 0.9578313253012049
45 Accuracy/train/Fraud/criminal/illegal conduct: 0.8303030303030303
46 Accuracy/val/Fraud/criminal/illegal conduct: 0.8132530120481928
47 Accuracy/train/Restitution/ disgorgement is not "Loss": 0.906060606060606
48 Accuracy/val/Restitution/ disgorgement is not "Loss": 0.9156626506024096
49 Accuracy/train/Settlement amount unreasonable: 0.9742424242424242
50 Accuracy/val/Settlement amount unreasonable: 0.9939759036144579
51 Accuracy/train/What is a "Claim"?: 0.8545454545454545
52 Accuracy/val/What is a "Claim"?: 0.7771084337349398
53 Accuracy/train/"Insured" v "Insured": 0.8575757575757575
54 Accuracy/val/"Insured" v "Insured": 0.8734939759036144
55 Accuracy/train/Failed to get insurer consent: 0.9303030303030303
56 Accuracy/val/Failed to get insurer consent: 0.9698795180722891
57 Accuracy/train/Other "Loss" issues: 0.8742424242424243
58 Accuracy/val/Other "Loss" issues: 0.9156626506024096
59 Accuracy/train/Other issues arising from insurer settlement conduct:
0.9121212121212121
60 Accuracy/val/Other issues arising from insurer settlement conduct:
0.8855421686746988
61 Accuracy/train/Property damage: 0.9378787878787879
62 Accuracy/val/Property damage: 0.9457831325301205
63 Accuracy/train/What is a "Securities Claim"?: 0.8939393939393939
64 Accuracy/val/What is a "Securities Claim"?: 0.9457831325301205
65 Accuracy/train/Libel or slander: 0.9681818181818181
66 Accuracy/val/Libel or slander: 0.9879518072289156
67 Accuracy/train/Other issues arising from PH settlement conduct: 0.9833333333333333
68 Accuracy/val/Other issues arising from PH settlement conduct: 0.9939759036144579
69 Accuracy/train/Unjust enrichment (profits not entitled): 0.8833333333333333
70 Accuracy/val/Unjust enrichment (profits not entitled): 0.8433734939759037
71 Accuracy/train/What is a "Related Claim" / "Interrelated Wrongful Act"?:
0.8560606060606061
72 Accuracy/val/What is a "Related Claim" / "Interrelated Wrongful Act"?:
0.8072289156626506
73 Accuracy/train/Employment practices: 0.9348484848484848
74 Accuracy/val/Employment practices: 0.9939759036144579
75 Accuracy/train/Prior claim / notice: 0.8833333333333333
76 Accuracy/val/Prior claim / notice: 0.8734939759036144
77 Accuracy/train/What counts as "Loss"?: 0.7848484848484848
78 Accuracy/val/What counts as "Loss"?: 0.7530120481927711
79 Accuracy/train/Prior acts: 0.9393939393939394
80 Accuracy/val/Prior acts: 0.9457831325301205
81 Accuracy/train/Professional services: 0.8787878787878788
82 Accuracy/val/Professional services: 0.927710843373494
83 Accuracy/train/Who is an "Insured"?: 0.8803030303030303
84 Accuracy/val/Who is an "Insured"?: 0.8554216867469879
85 Accuracy/train/Fiduciary liability: 0.9621212121212122
86 Accuracy/val/Fiduciary liability: 0.9939759036144579
87 Accuracy/train/Prior or pending litigation / proceeding: 0.8712121212121212
88 Accuracy/val/Prior or pending litigation / proceeding: 0.9156626506024096
89 Accuracy/train/Wrongful act not in capacity as a director or officer of the insured
(includes cases involving an "other capacity" exclusion): 0.9242424242424242
90 Accuracy/val/Wrongful act not in capacity as a director or officer of the insured (
includes cases involving an "other capacity" exclusion): 0.8855421686746988
91 Accuracy/train/Cyber: 0.9939393939393939
92 Accuracy/val/Cyber: 1.0
93 Accuracy/train/Late notice or reporting issue: 0.7893939393939394
94 Accuracy/val/Late notice or reporting issue: 0.7951807228915663

```

95 Accuracy/train/Prior knowledge: 0.9242424242424242
96 Accuracy/val/Prior knowledge: 0.9518072289156626
97 Accuracy/train/Bump up: 0.943939393939394
98 Accuracy/val/Bump up: 0.9759036144578314
99 Accuracy/train/Retro date issue: 0.95
100 Accuracy/val/Retro date issue: 0.9698795180722891
101 Accuracy/train/Misrepresentation/Rescission: 0.8575757575757575
102 Accuracy/val/Misrepresentation/Rescission: 0.9156626506024096
103 Accuracy/train/Regulatory: 0.9651515151515152
104 Accuracy/val/Regulatory: 1.0
105 Accuracy/train/Insolvency: 0.9878787878787879
106 Accuracy/val/Insolvency: 0.9939759036144579
107 Accuracy/train/Market segmentation exclusion issues: 0.9015151515151515
108 Accuracy/val/Market segmentation exclusion issues: 0.8975903614457831
109 Accuracy/train/Contract: 0.8409090909090909
110 Accuracy/val/Contract: 0.8674698795180723
111 Accuracy/train/Exclusion issues: 0.5681818181818182
112 Accuracy/val/Exclusion issues: 0.6445783132530121
113 Accuracy/train/Antitrust/restraint of trade/unfair business practice: 0.95
114 Accuracy/val/Antitrust/restraint of trade/unfair business practice:
    0.9819277108433735
115 Accuracy/train/Severability: 0.9742424242424242
116 Accuracy/val/Severability: 0.9939759036144579
117 Accuracy/train/Insurer refused to pay defense (be sure to check the reasons why):
    0.5348484848484848
118 Accuracy/val/Insurer refused to pay defense (be sure to check the reasons why):
    0.5963855421686747
119 Accuracy/train/Privacy/IP: 0.9606060606060606
120 Accuracy/val/Privacy/IP: 0.9698795180722891
121 Accuracy/train/Laser exclusion: 0.9378787878787879
122 Accuracy/val/Laser exclusion: 0.9698795180722891
123 Accuracy/train/PH failed to cooperate: 0.9681818181818181
124 Accuracy/val/PH failed to cooperate: 0.963855421686747
125 Accuracy/train/PH settlement conduct: 0.8939393939393939
126 Accuracy/val/PH settlement conduct: 0.9397590361445783
127 Accuracy/train/What counts as "Final Adjudication": 0.9560606060606060
128 Accuracy/val/What counts as "Final Adjudication": 0.9939759036144579
129 Accuracy/train/Insurer settlement conduct: 0.8787878787878788
130 Accuracy/val/Insurer settlement conduct: 0.8493975903614458
131 Accuracy/train/Other exclusion issues: 0.7848484848484848
132 Accuracy/val/Other exclusion issues: 0.7590361445783133
133 Accuracy/train/Other insurance (i.e. which insurance has to pay first):
    0.896969696969697
134 Accuracy/val/Other insurance (i.e. which insurance has to pay first):
    0.891566265060241
135 Accuracy/train/Allocation: 0.8818181818181818
136 Accuracy/val/Allocation: 0.9337349397590361
137 Accuracy/train/Arbitration: 0.9515151515151515
138 Accuracy/val/Arbitration: 0.9939759036144579
139 Accuracy/train/Bad faith: 0.6545454545454545
140 Accuracy/val/Bad faith: 0.7289156626506024
141 Accuracy/train/Other Coverage Issues: 0.8348484848484848
142 Accuracy/val/Other Coverage Issues: 0.8373493975903614
143 F1/train/Bodily injury: 0.0
144 F1/val/Bodily injury: 0.0
145 F1/train/Didn't settle when should have: 0.18947368421052632
146 F1/val/Didn't settle when should have: 0.0
147 F1/train/Fraud/criminal/illegal conduct: 0.034482758620689655

```

148 F1/val/Fraud/criminal/illegal conduct: 0.0
149 F1/train/Restitution/ disgorgement is not "Loss": 0.06060606060606061
150 F1/val/Restitution/ disgorgement is not "Loss": 0.0
151 F1/train/Settlement amount unreasonable: 0.0
152 F1/val/Settlement amount unreasonable: 0.0
153 F1/train/What is a "Claim"?: 0.12727272727272726
154 F1/val/What is a "Claim"?: 0.0
155 F1/train/"Insured" v "Insured": 0.06
156 F1/val/"Insured" v "Insured": 0.0
157 F1/train/Failed to get insurer consent: 0.08
158 F1/val/Failed to get insurer consent: 0.0
159 F1/train/Other "Loss" issues: 0.04597701149425287
160 F1/val/Other "Loss" issues: 0.0
161 F1/train/Other issues arising from insurer settlement conduct: 0.09375
162 F1/val/Other issues arising from insurer settlement conduct: 0.0
163 F1/train/Property damage: 0.046511627906976744
164 F1/val/Property damage: 0.18181818181818182
165 F1/train/What is a "Securities Claim"?: 0.02777777777777776
166 F1/val/What is a "Securities Claim"?: 0.0
167 F1/train/Libel or slander: 0.0
168 F1/val/Libel or slander: 0.0
169 F1/train/Other issues arising from PH settlement conduct: 0.0
170 F1/val/Other issues arising from PH settlement conduct: 0.0
171 F1/train/Unjust enrichment (profits not entitled): 0.04938271604938271
172 F1/val/Unjust enrichment (profits not entitled): 0.0
173 F1/train/What is a "Related Claim" / "Interrelated Wrongful Act"?:
0.020618556701030927
174 F1/val/What is a "Related Claim" / "Interrelated Wrongful Act"?: 0.0
175 F1/train/Employment practices: 0.044444444444444446
176 F1/val/Employment practices: 0.0
177 F1/train/Prior claim / notice: 0.0
178 F1/val/Prior claim / notice: 0.0
179 F1/train/What counts as "Loss"?: 0.05333333333333334
180 F1/val/What counts as "Loss"?: 0.0
181 F1/train/Prior acts: 0.047619047619047616
182 F1/val/Prior acts: 0.0
183 F1/train/Professional services: 0.11111111111111111
184 F1/val/Professional services: 0.0
185 F1/train/Who is an "Insured"?: 0.04819277108433735
186 F1/val/Who is an "Insured"?: 0.0
187 F1/train/Fiduciary liability: 0.0
188 F1/val/Fiduciary liability: 0.0
189 F1/train/Prior or pending litigation / proceeding: 0.08602150537634409
190 F1/val/Prior or pending litigation / proceeding: 0.0
191 F1/train/Wrongful act not in capacity as a director or officer of the insured (
includes cases involving an "other capacity" exclusion): 0.0
192 F1/val/Wrongful act not in capacity as a director or officer of the insured (
includes cases involving an "other capacity" exclusion): 0.0
193 F1/train/Cyber: 0.0
194 F1/val/Cyber: nan
195 F1/train/Late notice or reporting issue: 0.15757575757575756
196 F1/val/Late notice or reporting issue: 0.05555555555555555
197 F1/train/Prior knowledge: 0.0
198 F1/val/Prior knowledge: 0.0
199 F1/train/Bump up: 0.0
200 F1/val/Bump up: 0.0
201 F1/train/Retro date issue: 0.10810810810810811
202 F1/val/Retro date issue: 0.0

```

203 F1/train/Misrepresentation/Rescission: 0.09615384615384616
204 F1/val/Misrepresentation/Rescission: 0.125
205 F1/train/Regulatory: 0.0
206 F1/val/Regulatory: nan
207 F1/train/Insolvency: 0.0
208 F1/val/Insolvency: 0.0
209 F1/train/Market segmentation exclusion issues: 0.029850746268656716
210 F1/val/Market segmentation exclusion issues: 0.0
211 F1/train/Contract: 0.07079646017699115
212 F1/val/Contract: 0.0
213 F1/train/Exclusion issues: 0.29280397022332505
214 F1/val/Exclusion issues: 0.1917808219178082
215 F1/train/Antitrust/restraint of trade/unfair business practice: 0.15384615384615385
216 F1/val/Antitrust/restraint of trade/unfair business practice: 0.0
217 F1/train/Severability: 0.0
218 F1/val/Severability: 0.0
219 F1/train/Insurer refused to pay defense (be sure to check the reasons why):
    0.6442641946697567
220 F1/val/Insurer refused to pay defense (be sure to check the reasons why):
    0.7330677290836654
221 F1/train/Privacy/IP: 0.0
222 F1/val/Privacy/IP: 0.0
223 F1/train/Laser exclusion: 0.1276595744680851
224 F1/val/Laser exclusion: 0.0
225 F1/train/PH failed to cooperate: 0.0
226 F1/val/PH failed to cooperate: 0.0
227 F1/train/PH settlement conduct: 0.07894736842105263
228 F1/val/PH settlement conduct: 0.0
229 F1/train/What counts as "Final Adjudication": 0.0
230 F1/val/What counts as "Final Adjudication": 0.0
231 F1/train/Insurer settlement conduct: 0.06976744186046512
232 F1/val/Insurer settlement conduct: 0.0
233 F1/train/Other exclusion issues: 0.07792207792207792
234 F1/val/Other exclusion issues: 0.0
235 F1/train/Other insurance (i.e. which insurance has to pay first):
    0.02857142857142857
236 F1/val/Other insurance (i.e. which insurance has to pay first): 0.0
237 F1/train/Allocation: 0.11363636363636363
238 F1/val/Allocation: 0.0
239 F1/train/Arbitration: 0.0
240 F1/val/Arbitration: 0.0
241 F1/train/Bad faith: 0.19718309859154928
242 F1/val/Bad faith: 0.0425531914893617
243 F1/train/Other Coverage Issues: 0.05217391304347826
244 F1/val/Other Coverage Issues: 0.0
245 Total Train Accuracy: 0.888680926916221
246 Total Val Accuracy: 0.9039688164422396
247 F1 Train 0.20021344717182496
248 F1 Val 0.20215897939156036
249 100%|| 83/83 [57:43<00:00, 41.73s/it]
250 100%|| 21/21 [03:57<00:00, 11.31s/it]
251 /storage/ice1/5/9/mbock9/law-data-design-vip/ner/llamas/model.py:222: RuntimeWarning
    : invalid value encountered in scalar divide
252     writer.add_scalar(f"F1/train/{issue}", train_cm[issue][1, 1]/(train_cm[issue][1,
    1] + .5 * (train_cm[issue][0, 1] + train_cm[issue][1, 0])), epoch)
253 /storage/ice1/5/9/mbock9/law-data-design-vip/ner/llamas/model.py:247: RuntimeWarning
    : invalid value encountered in scalar divide

```

```
254 print(f"F1/train/{issue}:", train_cm[issue][1, 1]/(train_cm[issue][1, 1] + .5 * (
    train_cm[issue][0, 1] + train_cm[issue][1, 0]))
255 Epoch: 1
256 Train Loss: 12.10582906654082
257 Val Loss: 13.469373430524554
258 Accuracy/train/Bodily injury: 0.9666666666666667
259 Accuracy/val/Bodily injury: 0.9216867469879518
260 Accuracy/train/Didn't settle when should have: 0.9272727272727272
261 Accuracy/val/Didn't settle when should have: 0.9578313253012049
262 Accuracy/train/Fraud/criminal/illegal conduct: 0.8651515151515151
263 Accuracy/val/Fraud/criminal/illegal conduct: 0.8192771084337349
264 Accuracy/train/Restitution/ disgorgement is not "Loss": 0.9515151515151515
265 Accuracy/val/Restitution/ disgorgement is not "Loss": 0.9337349397590361
266 Accuracy/train/Settlement amount unreasonable: 0.996969696969697
267 Accuracy/val/Settlement amount unreasonable: 1.0
268 Accuracy/train/What is a "Claim"?: 0.8772727272727273
269 Accuracy/val/What is a "Claim"?: 0.7891566265060241
270 Accuracy/train/"Insured" v "Insured": 0.9090909090909091
271 Accuracy/val/"Insured" v "Insured": 0.891566265060241
272 Accuracy/train/Failed to get insurer consent: 0.9727272727272728
273 Accuracy/val/Failed to get insurer consent: 0.9759036144578314
274 Accuracy/train/Other "Loss" issues: 0.9136363636363637
275 Accuracy/val/Other "Loss" issues: 0.9156626506024096
276 Accuracy/train/Other issues arising from insurer settlement conduct:
    0.9484848484848485
277 Accuracy/val/Other issues arising from insurer settlement conduct:
    0.891566265060241
278 Accuracy/train/Property damage: 0.9757575757575757
279 Accuracy/val/Property damage: 0.9397590361445783
280 Accuracy/train/What is a "Securities Claim"?: 0.9666666666666667
281 Accuracy/val/What is a "Securities Claim"?: 0.9578313253012049
282 Accuracy/train/Label or slander: 0.996969696969697
283 Accuracy/val/Label or slander: 0.9939759036144579
284 Accuracy/train/Other issues arising from PH settlement conduct: 0.9878787878787879
285 Accuracy/val/Other issues arising from PH settlement conduct: 0.9939759036144579
286 Accuracy/train/Unjust enrichment (profits not entitled): 0.9015151515151515
287 Accuracy/val/Unjust enrichment (profits not entitled): 0.8614457831325302
288 Accuracy/train/What is a "Related Claim" / "Interrelated Wrongful Act"?:
    0.8590909090909091
289 Accuracy/val/What is a "Related Claim" / "Interrelated Wrongful Act"?:
    0.8132530120481928
290 Accuracy/train/Employment practices: 0.9833333333333333
291 Accuracy/val/Employment practices: 0.9939759036144579
292 Accuracy/train/Prior claim / notice: 0.9166666666666666
293 Accuracy/val/Prior claim / notice: 0.8855421686746988
294 Accuracy/train/What counts as "Loss"?: 0.8136363636363636
295 Accuracy/val/What counts as "Loss"?: 0.7530120481927711
296 Accuracy/train/Prior acts: 0.95
297 Accuracy/val/Prior acts: 0.9457831325301205
298 Accuracy/train/Professional services: 0.9242424242424242
299 Accuracy/val/Professional services: 0.9337349397590361
300 Accuracy/train/Who is an "Insured"?: 0.8939393939393939
301 Accuracy/val/Who is an "Insured"?: 0.8554216867469879
302 Accuracy/train/Fiduciary liability: 0.996969696969697
303 Accuracy/val/Fiduciary liability: 1.0
304 Accuracy/train/Prior or pending litigation / proceeding: 0.9181818181818182
305 Accuracy/val/Prior or pending litigation / proceeding: 0.9397590361445783
```


306 Accuracy/train/Wrongful act not in capacity as a director or officer of the insured
(includes cases involving an "other capacity" exclusion): 0.9287878787878788
307 Accuracy/val/Wrongful act not in capacity as a director or officer of the insured (
includes cases involving an "other capacity" exclusion): 0.8855421686746988
308 Accuracy/train/Cyber: 1.0
309 Accuracy/val/Cyber: 1.0
310 Accuracy/train/Late notice or reporting issue: 0.8651515151515151
311 Accuracy/val/Late notice or reporting issue: 0.8072289156626506
312 Accuracy/train/Prior knowledge: 0.9424242424242424
313 Accuracy/val/Prior knowledge: 0.963855421686747
314 Accuracy/train/Bump up: 0.9863636363636363
315 Accuracy/val/Bump up: 0.9759036144578314
316 Accuracy/train/Retro date issue: 0.9727272727272728
317 Accuracy/val/Retro date issue: 0.9759036144578314
318 Accuracy/train/Misrepresentation/Rescission: 0.9030303030303031
319 Accuracy/val/Misrepresentation/Rescission: 0.927710843373494
320 Accuracy/train/Regulatory: 0.9924242424242424
321 Accuracy/val/Regulatory: 1.0
322 Accuracy/train/Insolvency: 0.9924242424242424
323 Accuracy/val/Insolvency: 1.0
324 Accuracy/train/Market segmentation exclusion issues: 0.9075757575757576
325 Accuracy/val/Market segmentation exclusion issues: 0.8975903614457831
326 Accuracy/train/Contract: 0.8924242424242425
327 Accuracy/val/Contract: 0.8734939759036144
328 Accuracy/train/Exclusion issues: 0.6166666666666667
329 Accuracy/val/Exclusion issues: 0.6265060240963856
330 Accuracy/train/Antitrust/restraint of trade/unfair business practice:
0.9772727272727273
331 Accuracy/val/Antitrust/restraint of trade/unfair business practice:
0.9819277108433735
332 Accuracy/train/Severability: 0.9984848484848485
333 Accuracy/val/Severability: 0.9939759036144579
334 Accuracy/train/Insurer refused to pay defense (be sure to check the reasons why):
0.5424242424242425
335 Accuracy/val/Insurer refused to pay defense (be sure to check the reasons why):
0.6024096385542169
336 Accuracy/train/Privacy/IP: 0.9878787878787879
337 Accuracy/val/Privacy/IP: 0.9698795180722891
338 Accuracy/train/Laser exclusion: 0.9727272727272728
339 Accuracy/val/Laser exclusion: 0.9759036144578314
340 Accuracy/train/PH failed to cooperate: 0.9772727272727273
341 Accuracy/val/PH failed to cooperate: 0.9698795180722891
342 Accuracy/train/PH settlement conduct: 0.9621212121212122
343 Accuracy/val/PH settlement conduct: 0.9698795180722891
344 Accuracy/train/What counts as "Final Adjudication": 0.9757575757575757
345 Accuracy/val/What counts as "Final Adjudication": 0.9939759036144579
346 Accuracy/train/Insurer settlement conduct: 0.9
347 Accuracy/val/Insurer settlement conduct: 0.8554216867469879
348 Accuracy/train/Other exclusion issues: 0.8106060606060606
349 Accuracy/val/Other exclusion issues: 0.8012048192771084
350 Accuracy/train/Other insurance (i.e. which insurance has to pay first):
0.9318181818181818
351 Accuracy/val/Other insurance (i.e. which insurance has to pay first):
0.8975903614457831
352 Accuracy/train/Allocation: 0.9242424242424242
353 Accuracy/val/Allocation: 0.9397590361445783
354 Accuracy/train/Arbitration: 0.9939393939393939
355 Accuracy/val/Arbitration: 0.9939759036144579


```

356 Accuracy/train/Bad faith: 0.6909090909090909
357 Accuracy/val/Bad faith: 0.6987951807228916
358 Accuracy/train/Other Coverage Issues: 0.8590909090909091
359 Accuracy/val/Other Coverage Issues: 0.8614457831325302
360 F1/train/Bodily injury: 0.0
361 F1/val/Bodily injury: 0.0
362 F1/train/Didn't settle when should have: 0.0
363 F1/val/Didn't settle when should have: 0.0
364 F1/train/Fraud/criminal/illegal conduct: 0.02197802197802198
365 F1/val/Fraud/criminal/illegal conduct: 0.0
366 F1/train/Restitution/ disgorgement is not "Loss": 0.0
367 F1/val/Restitution/ disgorgement is not "Loss": 0.0
368 F1/train/Settlement amount unreasonable: 0.0
369 F1/val/Settlement amount unreasonable: nan
370 F1/train/What is a "Claim"?: 0.0
371 F1/val/What is a "Claim"?: 0.0
372 F1/train/"Insured" v "Insured": 0.0
373 F1/val/"Insured" v "Insured": 0.0
374 F1/train/Failed to get insurer consent: 0.0
375 F1/val/Failed to get insurer consent: 0.0
376 F1/train/Other "Loss" issues: 0.0
377 F1/val/Other "Loss" issues: 0.0
378 F1/train/Other issues arising from insurer settlement conduct: 0.0
379 F1/val/Other issues arising from insurer settlement conduct: 0.0
380 F1/train/Property damage: 0.0
381 F1/val/Property damage: 0.0
382 F1/train/What is a "Securities Claim"?: 0.0
383 F1/val/What is a "Securities Claim"?: 0.0
384 F1/train/Libel or slander: 0.0
385 F1/val/Libel or slander: 0.0
386 F1/train/Other issues arising from PH settlement conduct: 0.0
387 F1/val/Other issues arising from PH settlement conduct: 0.0
388 F1/train/Unjust enrichment (profits not entitled): 0.029850746268656716
389 F1/val/Unjust enrichment (profits not entitled): 0.0
390 F1/train/What is a "Related Claim" / "Interrelated Wrongful Act"?:
    0.021052631578947368
391 F1/val/What is a "Related Claim" / "Interrelated Wrongful Act"?: 0.0
392 F1/train/Employment practices: 0.0
393 F1/val/Employment practices: 0.0
394 F1/train/Prior claim / notice: 0.0
395 F1/val/Prior claim / notice: 0.0
396 F1/train/What counts as "Loss"?: 0.0
397 F1/val/What counts as "Loss"?: 0.0
398 F1/train/Prior acts: 0.0
399 F1/val/Prior acts: 0.0
400 F1/train/Professional services: 0.0
401 F1/val/Professional services: 0.0
402 F1/train/Who is an "Insured"?: 0.0
403 F1/val/Who is an "Insured"?: 0.0
404 F1/train/Fiduciary liability: 0.0
405 F1/val/Fiduciary liability: nan
406 F1/train/Prior or pending litigation / proceeding: 0.0
407 F1/val/Prior or pending litigation / proceeding: 0.0
408 F1/train/Wrongful act not in capacity as a director or officer of the insured (
    includes cases involving an "other capacity" exclusion): 0.0
409 F1/val/Wrongful act not in capacity as a director or officer of the insured (
    includes cases involving an "other capacity" exclusion): 0.0
410 F1/train/Cyber: nan

```

```
411 F1/val/Cyber: nan
412 F1/train/Late notice or reporting issue: 0.043010752688172046
413 F1/val/Late notice or reporting issue: 0.0
414 F1/train/Prior knowledge: 0.0
415 F1/val/Prior knowledge: 0.0
416 F1/train/Bump up: 0.0
417 F1/val/Bump up: 0.0
418 F1/train/Retro date issue: 0.0
419 F1/val/Retro date issue: 0.0
420 F1/train/Misrepresentation/Rescission: 0.0
421 F1/val/Misrepresentation/Rescission: 0.0
422 F1/train/Regulatory: 0.0
423 F1/val/Regulatory: nan
424 F1/train/Insolvency: 0.0
425 F1/val/Insolvency: nan
426 F1/train/Market segmentation exclusion issues: 0.031746031746031744
427 F1/val/Market segmentation exclusion issues: 0.0
428 F1/train/Contract: 0.0
429 F1/val/Contract: 0.0
430 F1/train/Exclusion issues: 0.29526462395543174
431 F1/val/Exclusion issues: 0.29545454545454547
432 F1/train/Antitrust/restraint of trade/unfair business practice: 0.0
433 F1/val/Antitrust/restraint of trade/unfair business practice: 0.0
434 F1/train/Severability: 0.0
435 F1/val/Severability: 0.0
436 F1/train/Insurer refused to pay defense (be sure to check the reasons why):
    0.6504629629629629
437 F1/val/Insurer refused to pay defense (be sure to check the reasons why):
    0.7317073170731707
438 F1/train/Privacy/IP: 0.0
439 F1/val/Privacy/IP: 0.0
440 F1/train/Laser exclusion: 0.0
441 F1/val/Laser exclusion: 0.0
442 F1/train/PH failed to cooperate: 0.0
443 F1/val/PH failed to cooperate: 0.0
444 F1/train/PH settlement conduct: 0.0
445 F1/val/PH settlement conduct: 0.0
446 F1/train/What counts as "Final Adjudication": 0.0
447 F1/val/What counts as "Final Adjudication": 0.0
448 F1/train/Insurer settlement conduct: 0.0
449 F1/val/Insurer settlement conduct: 0.0
450 F1/train/Other exclusion issues: 0.04580152671755725
451 F1/val/Other exclusion issues: 0.0
452 F1/train/Other insurance (i.e. which insurance has to pay first): 0.0
453 F1/val/Other insurance (i.e. which insurance has to pay first): 0.0
454 F1/train/Allocation: 0.0
455 F1/val/Allocation: 0.0
456 F1/train/Arbitration: 0.0
457 F1/val/Arbitration: 0.0
458 F1/train/Bad faith: 0.1774193548387097
459 F1/val/Bad faith: 0.375
460 F1/train/Other Coverage Issues: 0.06060606060606061
461 F1/val/Other Coverage Issues: 0.0
462 Total Train Accuracy: 0.919964349376114
463 Total Val Accuracy: 0.9098747932908103
464 F1 Train 0.21457725947521866
465 F1 Val 0.23623623623623624
466 100%|| 83/83 [57:49<00:00, 41.80s/it]
```

```
467 100%|| 21/21 [03:58<00:00, 11.33s/it]
468 Epoch: 2
469 Train Loss: 11.713588220527374
470 Val Loss: 13.351588975815545
471 Accuracy/train/Bodily injury: 0.9681818181818181
472 Accuracy/val/Bodily injury: 0.9216867469879518
473 Accuracy/train/Didn't settle when should have: 0.9272727272727272
474 Accuracy/val/Didn't settle when should have: 0.9578313253012049
475 Accuracy/train/Fraud/criminal/illegal conduct: 0.8681818181818182
476 Accuracy/val/Fraud/criminal/illegal conduct: 0.8132530120481928
477 Accuracy/train/Restitution/ disgorgement is not "Loss": 0.953030303030303
478 Accuracy/val/Restitution/ disgorgement is not "Loss": 0.9337349397590361
479 Accuracy/train/Settlement amount unreasonable: 0.996969696969697
480 Accuracy/val/Settlement amount unreasonable: 1.0
481 Accuracy/train/What is a "Claim"?: 0.8787878787878788
482 Accuracy/val/What is a "Claim"?: 0.7951807228915663
483 Accuracy/train/"Insured" v "Insured": 0.9090909090909091
484 Accuracy/val/"Insured" v "Insured": 0.891566265060241
485 Accuracy/train/Failed to get insurer consent: 0.9757575757575757
486 Accuracy/val/Failed to get insurer consent: 0.9759036144578314
487 Accuracy/train/Other "Loss" issues: 0.9136363636363637
488 Accuracy/val/Other "Loss" issues: 0.9156626506024096
489 Accuracy/train/Other issues arising from insurer settlement conduct:
    0.953030303030303
490 Accuracy/val/Other issues arising from insurer settlement conduct:
    0.891566265060241
491 Accuracy/train/Property damage: 0.9772727272727273
492 Accuracy/val/Property damage: 0.9397590361445783
493 Accuracy/train/What is a "Securities Claim"?: 0.9696969696969697
494 Accuracy/val/What is a "Securities Claim"?: 0.9578313253012049
495 Accuracy/train/Libel or slander: 0.996969696969697
496 Accuracy/val/Libel or slander: 0.9939759036144579
497 Accuracy/train/Other issues arising from PH settlement conduct: 0.9878787878787879
498 Accuracy/val/Other issues arising from PH settlement conduct: 0.9939759036144579
499 Accuracy/train/Unjust enrichment (profits not entitled): 0.9015151515151515
500 Accuracy/val/Unjust enrichment (profits not entitled): 0.8614457831325302
501 Accuracy/train/What is a "Related Claim" / "Interrelated Wrongful Act"?:
    0.8575757575757575
502 Accuracy/val/What is a "Related Claim" / "Interrelated Wrongful Act"?:
    0.8192771084337349
503 Accuracy/train/Employment practices: 0.9833333333333333
504 Accuracy/val/Employment practices: 0.9939759036144579
505 Accuracy/train/Prior claim / notice: 0.9166666666666666
506 Accuracy/val/Prior claim / notice: 0.8855421686746988
507 Accuracy/train/What counts as "Loss"?: 0.8166666666666667
508 Accuracy/val/What counts as "Loss"?: 0.7530120481927711
509 Accuracy/train/Prior acts: 0.9515151515151515
510 Accuracy/val/Prior acts: 0.9457831325301205
511 Accuracy/train/Professional services: 0.9242424242424242
512 Accuracy/val/Professional services: 0.9337349397590361
513 Accuracy/train/Who is an "Insured"?: 0.8939393939393939
514 Accuracy/val/Who is an "Insured"?: 0.8554216867469879
515 Accuracy/train/Fiduciary liability: 0.996969696969697
516 Accuracy/val/Fiduciary liability: 1.0
517 Accuracy/train/Prior or pending litigation / proceeding: 0.9212121212121213
518 Accuracy/val/Prior or pending litigation / proceeding: 0.9397590361445783
519 Accuracy/train/Wrongful act not in capacity as a director or officer of the insured
    (includes cases involving an "other capacity" exclusion): 0.9303030303030303
```

520 Accuracy/val/Wrongful act not in capacity as a director or officer of the insured (includes cases involving an "other capacity" exclusion): 0.8855421686746988
521 Accuracy/train/Cyber: 1.0
522 Accuracy/val/Cyber: 1.0
523 Accuracy/train/Late notice or reporting issue: 0.8621212121212121
524 Accuracy/val/Late notice or reporting issue: 0.8072289156626506
525 Accuracy/train/Prior knowledge: 0.943939393939394
526 Accuracy/val/Prior knowledge: 0.963855421686747
527 Accuracy/train/Bump up: 0.9878787878787879
528 Accuracy/val/Bump up: 0.9759036144578314
529 Accuracy/train/Retro date issue: 0.9727272727272728
530 Accuracy/val/Retro date issue: 0.9759036144578314
531 Accuracy/train/Misrepresentation/Rescission: 0.906060606060606
532 Accuracy/val/Misrepresentation/Rescission: 0.927710843373494
533 Accuracy/train/Regulatory: 0.9939393939393939
534 Accuracy/val/Regulatory: 1.0
535 Accuracy/train/Insolvency: 0.9924242424242424
536 Accuracy/val/Insolvency: 1.0
537 Accuracy/train/Market segmentation exclusion issues: 0.9090909090909091
538 Accuracy/val/Market segmentation exclusion issues: 0.8975903614457831
539 Accuracy/train/Contract: 0.8954545454545455
540 Accuracy/val/Contract: 0.8734939759036144
541 Accuracy/train/Exclusion issues: 0.6227272727272727
542 Accuracy/val/Exclusion issues: 0.6566265060240963
543 Accuracy/train/Antitrust/restraint of trade/unfair business practice: 0.9772727272727273
544 Accuracy/val/Antitrust/restraint of trade/unfair business practice: 0.9819277108433735
545 Accuracy/train/Severability: 0.9984848484848485
546 Accuracy/val/Severability: 0.9939759036144579
547 Accuracy/train/Insurer refused to pay defense (be sure to check the reasons why): 0.6090909090909091
548 Accuracy/val/Insurer refused to pay defense (be sure to check the reasons why): 0.5903614457831325
549 Accuracy/train/Privacy/IP: 0.9878787878787879
550 Accuracy/val/Privacy/IP: 0.9698795180722891
551 Accuracy/train/Laser exclusion: 0.9727272727272728
552 Accuracy/val/Laser exclusion: 0.9759036144578314
553 Accuracy/train/PH failed to cooperate: 0.9787878787878788
554 Accuracy/val/PH failed to cooperate: 0.9698795180722891
555 Accuracy/train/PH settlement conduct: 0.9666666666666667
556 Accuracy/val/PH settlement conduct: 0.9698795180722891
557 Accuracy/train/What counts as "Final Adjudication": 0.9757575757575757
558 Accuracy/val/What counts as "Final Adjudication": 0.9939759036144579
559 Accuracy/train/Insurer settlement conduct: 0.9
560 Accuracy/val/Insurer settlement conduct: 0.8493975903614458
561 Accuracy/train/Other exclusion issues: 0.8136363636363636
562 Accuracy/val/Other exclusion issues: 0.8012048192771084
563 Accuracy/train/Other insurance (i.e. which insurance has to pay first): 0.9318181818181818
564 Accuracy/val/Other insurance (i.e. which insurance has to pay first): 0.8975903614457831
565 Accuracy/train/Allocation: 0.9257575757575758
566 Accuracy/val/Allocation: 0.9397590361445783
567 Accuracy/train/Arbitration: 0.9939393939393939
568 Accuracy/val/Arbitration: 0.9939759036144579
569 Accuracy/train/Bad faith: 0.6757575757575758
570 Accuracy/val/Bad faith: 0.7289156626506024

571 Accuracy/train/Other Coverage Issues: 0.8545454545454545
572 Accuracy/val/Other Coverage Issues: 0.8433734939759037
573 F1/train/Bodily injury: 0.0
574 F1/val/Bodily injury: 0.0
575 F1/train/Didn't settle when should have: 0.0
576 F1/val/Didn't settle when should have: 0.0
577 F1/train/Fraud/criminal/illegal conduct: 0.02247191011235955
578 F1/val/Fraud/criminal/illegal conduct: 0.0
579 F1/train/Restitution/ disgorgement is not "Loss": 0.0
580 F1/val/Restitution/ disgorgement is not "Loss": 0.0
581 F1/train/Settlement amount unreasonable: 0.0
582 F1/val/Settlement amount unreasonable: nan
583 F1/train/What is a "Claim"?: 0.024390243902439025
584 F1/val/What is a "Claim"?: 0.0
585 F1/train/"Insured" v "Insured": 0.0
586 F1/val/"Insured" v "Insured": 0.0
587 F1/train/Failed to get insurer consent: 0.0
588 F1/val/Failed to get insurer consent: 0.0
589 F1/train/Other "Loss" issues: 0.0
590 F1/val/Other "Loss" issues: 0.0
591 F1/train/Other issues arising from insurer settlement conduct: 0.0
592 F1/val/Other issues arising from insurer settlement conduct: 0.0
593 F1/train/Property damage: 0.0
594 F1/val/Property damage: 0.0
595 F1/train/What is a "Securities Claim"?: 0.0
596 F1/val/What is a "Securities Claim"?: 0.0
597 F1/train/Label or slander: 0.0
598 F1/val/Label or slander: 0.0
599 F1/train/Other issues arising from PH settlement conduct: 0.0
600 F1/val/Other issues arising from PH settlement conduct: 0.0
601 F1/train/Unjust enrichment (profits not entitled): 0.0
602 F1/val/Unjust enrichment (profits not entitled): 0.0
603 F1/train/What is a "Related Claim" / "Interrelated Wrongful Act"?: 0.0
604 F1/val/What is a "Related Claim" / "Interrelated Wrongful Act"?: 0.0625
605 F1/train/Employment practices: 0.0
606 F1/val/Employment practices: 0.0
607 F1/train/Prior claim / notice: 0.0
608 F1/val/Prior claim / notice: 0.0
609 F1/train/What counts as "Loss"?: 0.0
610 F1/val/What counts as "Loss"?: 0.0
611 F1/train/Prior acts: 0.0
612 F1/val/Prior acts: 0.0
613 F1/train/Professional services: 0.0
614 F1/val/Professional services: 0.0
615 F1/train/Who is an "Insured"?: 0.0
616 F1/val/Who is an "Insured"?: 0.0
617 F1/train/Fiduciary liability: 0.0
618 F1/val/Fiduciary liability: nan
619 F1/train/Prior or pending litigation / proceeding: 0.0
620 F1/val/Prior or pending litigation / proceeding: 0.0
621 F1/train/Wrongful act not in capacity as a director or officer of the insured (includes cases involving an "other capacity" exclusion): 0.0
622 F1/val/Wrongful act not in capacity as a director or officer of the insured (includes cases involving an "other capacity" exclusion): 0.0
623 F1/train/Cyber: nan
624 F1/val/Cyber: nan
625 F1/train/Late notice or reporting issue: 0.0
626 F1/val/Late notice or reporting issue: 0.058823529411764705

```
627 F1/train/Prior knowledge: 0.0
628 F1/val/Prior knowledge: 0.0
629 F1/train/Bump up: 0.0
630 F1/val/Bump up: 0.0
631 F1/train/Retro date issue: 0.0
632 F1/val/Retro date issue: 0.0
633 F1/train/Misrepresentation/Rescission: 0.0
634 F1/val/Misrepresentation/Rescission: 0.0
635 F1/train/Regulatory: 0.0
636 F1/val/Regulatory: nan
637 F1/train/Insolvency: 0.0
638 F1/val/Insolvency: nan
639 F1/train/Market segmentation exclusion issues: 0.03225806451612903
640 F1/val/Market segmentation exclusion issues: 0.0
641 F1/train/Contract: 0.0
642 F1/val/Contract: 0.0
643 F1/train/Exclusion issues: 0.35989717223650386
644 F1/val/Exclusion issues: 0.2191780821917808
645 F1/train/Antitrust/restraint of trade/unfair business practice: 0.0
646 F1/val/Antitrust/restraint of trade/unfair business practice: 0.0
647 F1/train/Severability: 0.0
648 F1/val/Severability: 0.0
649 F1/train/Insurer refused to pay defense (be sure to check the reasons why):
    0.7041284403669725
650 F1/val/Insurer refused to pay defense (be sure to check the reasons why):
    0.7235772357723578
651 F1/train/Privacy/IP: 0.0
652 F1/val/Privacy/IP: 0.0
653 F1/train/Laser exclusion: 0.0
654 F1/val/Laser exclusion: 0.0
655 F1/train/PH failed to cooperate: 0.0
656 F1/val/PH failed to cooperate: 0.0
657 F1/train/PH settlement conduct: 0.0
658 F1/val/PH settlement conduct: 0.0
659 F1/train/What counts as "Final Adjudication": 0.0
660 F1/val/What counts as "Final Adjudication": 0.0
661 F1/train/Insurer settlement conduct: 0.0
662 F1/val/Insurer settlement conduct: 0.0
663 F1/train/Other exclusion issues: 0.0
664 F1/val/Other exclusion issues: 0.0
665 F1/train/Other insurance (i.e. which insurance has to pay first): 0.0
666 F1/val/Other insurance (i.e. which insurance has to pay first): 0.0
667 F1/train/Allocation: 0.0
668 F1/val/Allocation: 0.0
669 F1/train/Arbitration: 0.0
670 F1/val/Arbitration: 0.0
671 F1/train/Bad faith: 0.15079365079365079
672 F1/val/Bad faith: 0.0425531914893617
673 F1/train/Other Coverage Issues: 0.0
674 F1/val/Other Coverage Issues: 0.0
675 Total Train Accuracy: 0.9219251336898395
676 Total Val Accuracy: 0.9104653909756674
677 F1 Train 0.2329246935201401
678 F1 Val 0.20876826722338204
679 100%|| 83/83 [57:51<00:00, 41.82s/it]
680 100%|| 21/21 [03:58<00:00, 11.34s/it]
681 Epoch: 3
682 Train Loss: 11.474507326103119
```

683 Val Loss: 13.242114339556013
684 Accuracy/train/Bodily injury: 0.9681818181818181
685 Accuracy/val/Bodily injury: 0.9216867469879518
686 Accuracy/train/Didn't settle when should have: 0.9287878787878788
687 Accuracy/val/Didn't settle when should have: 0.9578313253012049
688 Accuracy/train/Fraud/criminal/illegal conduct: 0.8666666666666667
689 Accuracy/val/Fraud/criminal/illegal conduct: 0.8132530120481928
690 Accuracy/train/Restitution/ disgorgement is not "Loss": 0.9530303030303030
691 Accuracy/val/Restitution/ disgorgement is not "Loss": 0.9337349397590361
692 Accuracy/train/Settlement amount unreasonable: 0.996969696969697
693 Accuracy/val/Settlement amount unreasonable: 1.0
694 Accuracy/train/What is a "Claim"?: 0.8787878787878788
695 Accuracy/val/What is a "Claim"?: 0.7771084337349398
696 Accuracy/train/"Insured" v "Insured" v "Insured": 0.9090909090909091
697 Accuracy/val/"Insured" v "Insured": 0.891566265060241
698 Accuracy/train/Failed to get insurer consent: 0.9757575757575757
699 Accuracy/val/Failed to get insurer consent: 0.9759036144578314
700 Accuracy/train/Other "Loss" issues: 0.9136363636363637
701 Accuracy/val/Other "Loss" issues: 0.9156626506024096
702 Accuracy/train/Other issues arising from insurer settlement conduct:
0.9530303030303030
703 Accuracy/val/Other issues arising from insurer settlement conduct:
0.891566265060241
704 Accuracy/train/Property damage: 0.9772727272727273
705 Accuracy/val/Property damage: 0.9397590361445783
706 Accuracy/train/What is a "Securities Claim"?: 0.9696969696969697
707 Accuracy/val/What is a "Securities Claim"?: 0.9578313253012049
708 Accuracy/train/Libel or slander: 0.996969696969697
709 Accuracy/val/Libel or slander: 0.9939759036144579
710 Accuracy/train/Other issues arising from PH settlement conduct: 0.9878787878787879
711 Accuracy/val/Other issues arising from PH settlement conduct: 0.9939759036144579
712 Accuracy/train/Unjust enrichment (profits not entitled): 0.9015151515151515
713 Accuracy/val/Unjust enrichment (profits not entitled): 0.8554216867469879
714 Accuracy/train/What is a "Related Claim" / "Interrelated Wrongful Act"?:
0.8606060606060606
715 Accuracy/val/What is a "Related Claim" / "Interrelated Wrongful Act"?:
0.8192771084337349
716 Accuracy/train/Employment practices: 0.9833333333333333
717 Accuracy/val/Employment practices: 0.9939759036144579
718 Accuracy/train/Prior claim / notice: 0.9166666666666666
719 Accuracy/val/Prior claim / notice: 0.8855421686746988
720 Accuracy/train/What counts as "Loss"?: 0.8151515151515152
721 Accuracy/val/What counts as "Loss"?: 0.7530120481927711
722 Accuracy/train/Prior acts: 0.9515151515151515
723 Accuracy/val/Prior acts: 0.9457831325301205
724 Accuracy/train/Professional services: 0.9257575757575758
725 Accuracy/val/Professional services: 0.9337349397590361
726 Accuracy/train/Who is an "Insured"?: 0.8939393939393939
727 Accuracy/val/Who is an "Insured"?: 0.8554216867469879
728 Accuracy/train/Fiduciary liability: 0.996969696969697
729 Accuracy/val/Fiduciary liability: 1.0
730 Accuracy/train/Prior or pending litigation / proceeding: 0.9212121212121213
731 Accuracy/val/Prior or pending litigation / proceeding: 0.9397590361445783
732 Accuracy/train/Wrongful act not in capacity as a director or officer of the insured
(includes cases involving an "other capacity" exclusion): 0.9303030303030303
733 Accuracy/val/Wrongful act not in capacity as a director or officer of the insured (
includes cases involving an "other capacity" exclusion): 0.8855421686746988
734 Accuracy/train/Cyber: 1.0

735 Accuracy/val/Cyber: 1.0
736 Accuracy/train/Late notice or reporting issue: 0.8636363636363636
737 Accuracy/val/Late notice or reporting issue: 0.7951807228915663
738 Accuracy/train/Prior knowledge: 0.9454545454545454
739 Accuracy/val/Prior knowledge: 0.963855421686747
740 Accuracy/train/Bump up: 0.9878787878787879
741 Accuracy/val/Bump up: 0.9759036144578314
742 Accuracy/train/Retro date issue: 0.9727272727272728
743 Accuracy/val/Retro date issue: 0.9759036144578314
744 Accuracy/train/Misrepresentation/Rescission: 0.906060606060606
745 Accuracy/val/Misrepresentation/Rescission: 0.927710843373494
746 Accuracy/train/Regulatory: 0.9939393939393939
747 Accuracy/val/Regulatory: 1.0
748 Accuracy/train/Insolvency: 0.9924242424242424
749 Accuracy/val/Insolvency: 1.0
750 Accuracy/train/Market segmentation exclusion issues: 0.9090909090909091
751 Accuracy/val/Market segmentation exclusion issues: 0.8975903614457831
752 Accuracy/train/Contract: 0.8954545454545455
753 Accuracy/val/Contract: 0.8734939759036144
754 Accuracy/train/Exclusion issues: 0.6363636363636364
755 Accuracy/val/Exclusion issues: 0.536144578313253
756 Accuracy/train/Antitrust/restraint of trade/unfair business practice:
0.9772727272727273
757 Accuracy/val/Antitrust/restraint of trade/unfair business practice:
0.9819277108433735
758 Accuracy/train/Severability: 0.9984848484848485
759 Accuracy/val/Severability: 0.9939759036144579
760 Accuracy/train/Insurer refused to pay defense (be sure to check the reasons why):
0.6242424242424243
761 Accuracy/val/Insurer refused to pay defense (be sure to check the reasons why):
0.5542168674698795
762 Accuracy/train/Privacy/IP: 0.9878787878787879
763 Accuracy/val/Privacy/IP: 0.9698795180722891
764 Accuracy/train/Laser exclusion: 0.9727272727272728
765 Accuracy/val/Laser exclusion: 0.9759036144578314
766 Accuracy/train/PH failed to cooperate: 0.9787878787878788
767 Accuracy/val/PH failed to cooperate: 0.9698795180722891
768 Accuracy/train/PH settlement conduct: 0.9666666666666667
769 Accuracy/val/PH settlement conduct: 0.9698795180722891
770 Accuracy/train/What counts as "Final Adjudication": 0.9757575757575757
771 Accuracy/val/What counts as "Final Adjudication": 0.9939759036144579
772 Accuracy/train/Insurer settlement conduct: 0.9
773 Accuracy/val/Insurer settlement conduct: 0.8554216867469879
774 Accuracy/train/Other exclusion issues: 0.8166666666666667
775 Accuracy/val/Other exclusion issues: 0.7951807228915663
776 Accuracy/train/Other insurance (i.e. which insurance has to pay first):
0.9318181818181818
777 Accuracy/val/Other insurance (i.e. which insurance has to pay first):
0.891566265060241
778 Accuracy/train/Allocation: 0.9257575757575758
779 Accuracy/val/Allocation: 0.9397590361445783
780 Accuracy/train/Arbitration: 0.9939393939393939
781 Accuracy/val/Arbitration: 0.9939759036144579
782 Accuracy/train/Bad faith: 0.7
783 Accuracy/val/Bad faith: 0.7289156626506024
784 Accuracy/train/Other Coverage Issues: 0.8545454545454545
785 Accuracy/val/Other Coverage Issues: 0.8433734939759037
786 F1/train/Bodily injury: 0.0


```

787 F1/val/Bodily injury: 0.0
788 F1/train/Didn't settle when should have: 0.04081632653061224
789 F1/val/Didn't settle when should have: 0.0
790 F1/train/Fraud/criminal/illegal conduct: 0.02222222222222223
791 F1/val/Fraud/criminal/illegal conduct: 0.0
792 F1/train/Restitution/ disgorgement is not "Loss": 0.0
793 F1/val/Restitution/ disgorgement is not "Loss": 0.0
794 F1/train/Settlement amount unreasonable: 0.0
795 F1/val/Settlement amount unreasonable: nan
796 F1/train/What is a "Claim"?: 0.024390243902439025
797 F1/val/What is a "Claim"?: 0.0
798 F1/train/"Insured" v "Insured": 0.0
799 F1/val/"Insured" v "Insured": 0.0
800 F1/train/Failed to get insurer consent: 0.0
801 F1/val/Failed to get insurer consent: 0.0
802 F1/train/Other "Loss" issues: 0.0
803 F1/val/Other "Loss" issues: 0.0
804 F1/train/Other issues arising from insurer settlement conduct: 0.0
805 F1/val/Other issues arising from insurer settlement conduct: 0.0
806 F1/train/Property damage: 0.0
807 F1/val/Property damage: 0.0
808 F1/train/What is a "Securities Claim"?: 0.0
809 F1/val/What is a "Securities Claim"?: 0.0
810 F1/train/Libel or slander: 0.0
811 F1/val/Libel or slander: 0.0
812 F1/train/Other issues arising from PH settlement conduct: 0.0
813 F1/val/Other issues arising from PH settlement conduct: 0.0
814 F1/train/Unjust enrichment (profits not entitled): 0.0
815 F1/val/Unjust enrichment (profits not entitled): 0.0
816 F1/train/What is a "Related Claim" / "Interrelated Wrongful Act"?:
    0.04166666666666664
817 F1/val/What is a "Related Claim" / "Interrelated Wrongful Act"?: 0.0625
818 F1/train/Employment practices: 0.0
819 F1/val/Employment practices: 0.0
820 F1/train/Prior claim / notice: 0.0
821 F1/val/Prior claim / notice: 0.0
822 F1/train/What counts as "Loss"?: 0.0
823 F1/val/What counts as "Loss"?: 0.0
824 F1/train/Prior acts: 0.0
825 F1/val/Prior acts: 0.0
826 F1/train/Professional services: 0.0392156862745098
827 F1/val/Professional services: 0.0
828 F1/train/Who is an "Insured"?: 0.0
829 F1/val/Who is an "Insured"?: 0.0
830 F1/train/Fiduciary liability: 0.0
831 F1/val/Fiduciary liability: nan
832 F1/train/Prior or pending litigation / proceeding: 0.0
833 F1/val/Prior or pending litigation / proceeding: 0.0
834 F1/train/Wrongful act not in capacity as a director or officer of the insured (
    includes cases involving an "other capacity" exclusion): 0.0
835 F1/val/Wrongful act not in capacity as a director or officer of the insured (
    includes cases involving an "other capacity" exclusion): 0.0
836 F1/train/Cyber: nan
837 F1/val/Cyber: nan
838 F1/train/Late notice or reporting issue: 0.0
839 F1/val/Late notice or reporting issue: 0.05555555555555555
840 F1/train/Prior knowledge: 0.0
841 F1/val/Prior knowledge: 0.0

```

```
842 F1/train/Bump up: 0.0
843 F1/val/Bump up: 0.0
844 F1/train/Retro date issue: 0.0
845 F1/val/Retro date issue: 0.0
846 F1/train/Misrepresentation/Rescission: 0.0
847 F1/val/Misrepresentation/Rescission: 0.14285714285714285
848 F1/train/Regulatory: 0.0
849 F1/val/Regulatory: nan
850 F1/train/Insolvency: 0.0
851 F1/val/Insolvency: nan
852 F1/train/Market segmentation exclusion issues: 0.03225806451612903
853 F1/val/Market segmentation exclusion issues: 0.0
854 F1/train/Contract: 0.0
855 F1/val/Contract: 0.0
856 F1/train/Exclusion issues: 0.4117647058823529
857 F1/val/Exclusion issues: 0.47619047619047616
858 F1/train/Antitrust/restraint of trade/unfair business practice: 0.0
859 F1/val/Antitrust/restraint of trade/unfair business practice: 0.0
860 F1/train/Severability: 0.0
861 F1/val/Severability: 0.0
862 F1/train/Insurer refused to pay defense (be sure to check the reasons why):
    0.7333333333333333
863 F1/val/Insurer refused to pay defense (be sure to check the reasons why):
    0.6542056074766355
864 F1/train/Privacy/IP: 0.0
865 F1/val/Privacy/IP: 0.0
866 F1/train/Laser exclusion: 0.0
867 F1/val/Laser exclusion: 0.0
868 F1/train/PH failed to cooperate: 0.0
869 F1/val/PH failed to cooperate: 0.0
870 F1/train/PH settlement conduct: 0.0
871 F1/val/PH settlement conduct: 0.0
872 F1/train/What counts as "Final Adjudication": 0.0
873 F1/val/What counts as "Final Adjudication": 0.0
874 F1/train/Insurer settlement conduct: 0.0
875 F1/val/Insurer settlement conduct: 0.0
876 F1/train/Other exclusion issues: 0.032
877 F1/val/Other exclusion issues: 0.0
878 F1/train/Other insurance (i.e. which insurance has to pay first): 0.0
879 F1/val/Other insurance (i.e. which insurance has to pay first): 0.0
880 F1/train/Allocation: 0.0
881 F1/val/Allocation: 0.0
882 F1/train/Arbitration: 0.0
883 F1/val/Arbitration: 0.0
884 F1/train/Bad faith: 0.1391304347826087
885 F1/val/Bad faith: 0.2857142857142857
886 F1/train/Other Coverage Issues: 0.0
887 F1/val/Other Coverage Issues: 0.0
888 Total Train Accuracy: 0.9231431966726085
889 Total Val Accuracy: 0.9065674462556107
890 F1 Train 0.2581015199311729
891 F1 Val 0.22829268292682928
892 100%|| 83/83 [57:52<00:00, 41.83s/it]
893 100%|| 21/21 [03:58<00:00, 11.34s/it]
894 Epoch: 4
895 Train Loss: 11.139838844896799
896 Val Loss: 13.154514585222516
897 Accuracy/train/Bodily injury: 0.9681818181818181
```

898 Accuracy/val/Bodily injury: 0.9216867469879518
899 Accuracy/train/Didn't settle when should have: 0.9303030303030303
900 Accuracy/val/Didn't settle when should have: 0.9578313253012049
901 Accuracy/train/Fraud/criminal/illegal conduct: 0.8681818181818182
902 Accuracy/val/Fraud/criminal/illegal conduct: 0.8132530120481928
903 Accuracy/train/Restitution/ disgorgement is not "Loss": 0.953030303030303
904 Accuracy/val/Restitution/ disgorgement is not "Loss": 0.9337349397590361
905 Accuracy/train/Settlement amount unreasonable: 0.996969696969697
906 Accuracy/val/Settlement amount unreasonable: 1.0
907 Accuracy/train/What is a "Claim"?: 0.8787878787878788
908 Accuracy/val/What is a "Claim"?: 0.7951807228915663
909 Accuracy/train/"Insured" v "Insured": 0.9090909090909091
910 Accuracy/val/"Insured" v "Insured": 0.891566265060241
911 Accuracy/train/Failed to get insurer consent: 0.9757575757575757
912 Accuracy/val/Failed to get insurer consent: 0.9759036144578314
913 Accuracy/train/Other "Loss" issues: 0.9136363636363637
914 Accuracy/val/Other "Loss" issues: 0.9156626506024096
915 Accuracy/train/Other issues arising from insurer settlement conduct:
0.953030303030303
916 Accuracy/val/Other issues arising from insurer settlement conduct:
0.891566265060241
917 Accuracy/train/Property damage: 0.9772727272727273
918 Accuracy/val/Property damage: 0.9397590361445783
919 Accuracy/train/What is a "Securities Claim"?: 0.9712121212121212
920 Accuracy/val/What is a "Securities Claim"?: 0.9578313253012049
921 Accuracy/train/Libel or slander: 0.996969696969697
922 Accuracy/val/Libel or slander: 0.9939759036144579
923 Accuracy/train/Other issues arising from PH settlement conduct: 0.9878787878787879
924 Accuracy/val/Other issues arising from PH settlement conduct: 0.9939759036144579
925 Accuracy/train/Unjust enrichment (profits not entitled): 0.9030303030303031
926 Accuracy/val/Unjust enrichment (profits not entitled): 0.8614457831325302
927 Accuracy/train/What is a "Related Claim" / "Interrelated Wrongful Act"?:
0.8621212121212121
928 Accuracy/val/What is a "Related Claim" / "Interrelated Wrongful Act"?:
0.8192771084337349
929 Accuracy/train/Employment practices: 0.9833333333333333
930 Accuracy/val/Employment practices: 0.9939759036144579
931 Accuracy/train/Prior claim / notice: 0.9166666666666666
932 Accuracy/val/Prior claim / notice: 0.8855421686746988
933 Accuracy/train/What counts as "Loss"?: 0.8181818181818182
934 Accuracy/val/What counts as "Loss"?: 0.7469879518072289
935 Accuracy/train/Prior acts: 0.9515151515151515
936 Accuracy/val/Prior acts: 0.9457831325301205
937 Accuracy/train/Professional services: 0.9242424242424242
938 Accuracy/val/Professional services: 0.9337349397590361
939 Accuracy/train/Who is an "Insured"?: 0.8939393939393939
940 Accuracy/val/Who is an "Insured"?: 0.8554216867469879
941 Accuracy/train/Fiduciary liability: 0.996969696969697
942 Accuracy/val/Fiduciary liability: 1.0
943 Accuracy/train/Prior or pending litigation / proceeding: 0.9212121212121213
944 Accuracy/val/Prior or pending litigation / proceeding: 0.9397590361445783
945 Accuracy/train/Wrongful act not in capacity as a director or officer of the insured
(includes cases involving an "other capacity" exclusion): 0.9303030303030303
946 Accuracy/val/Wrongful act not in capacity as a director or officer of the insured (
includes cases involving an "other capacity" exclusion): 0.8855421686746988
947 Accuracy/train/Cyber: 1.0
948 Accuracy/val/Cyber: 1.0
949 Accuracy/train/Late notice or reporting issue: 0.8712121212121212

950 Accuracy/val/Late notice or reporting issue: 0.8012048192771084
951 Accuracy/train/Prior knowledge: 0.9454545454545454
952 Accuracy/val/Prior knowledge: 0.9698795180722891
953 Accuracy/train/Bump up: 0.9878787878787879
954 Accuracy/val/Bump up: 0.9759036144578314
955 Accuracy/train/Retro date issue: 0.9727272727272728
956 Accuracy/val/Retro date issue: 0.9759036144578314
957 Accuracy/train/Misrepresentation/Rescission: 0.9075757575757576
958 Accuracy/val/Misrepresentation/Rescission: 0.927710843373494
959 Accuracy/train/Regulatory: 0.9939393939393939
960 Accuracy/val/Regulatory: 1.0
961 Accuracy/train/Insolvency: 0.9924242424242424
962 Accuracy/val/Insolvency: 1.0
963 Accuracy/train/Market segmentation exclusion issues: 0.9106060606060606
964 Accuracy/val/Market segmentation exclusion issues: 0.8975903614457831
965 Accuracy/train/Contract: 0.8954545454545455
966 Accuracy/val/Contract: 0.8795180722891566
967 Accuracy/train/Exclusion issues: 0.6545454545454545
968 Accuracy/val/Exclusion issues: 0.6987951807228916
969 Accuracy/train/Antitrust/restraint of trade/unfair business practice:
0.9772727272727273
970 Accuracy/val/Antitrust/restraint of trade/unfair business practice:
0.9819277108433735
971 Accuracy/train/Severability: 0.9984848484848485
972 Accuracy/val/Severability: 0.9939759036144579
973 Accuracy/train/Insurer refused to pay defense (be sure to check the reasons why):
0.6363636363636364
974 Accuracy/val/Insurer refused to pay defense (be sure to check the reasons why):
0.6385542168674698
975 Accuracy/train/Privacy/IP: 0.9878787878787879
976 Accuracy/val/Privacy/IP: 0.9698795180722891
977 Accuracy/train/Laser exclusion: 0.9727272727272728
978 Accuracy/val/Laser exclusion: 0.9759036144578314
979 Accuracy/train/PH failed to cooperate: 0.9787878787878788
980 Accuracy/val/PH failed to cooperate: 0.9698795180722891
981 Accuracy/train/PH settlement conduct: 0.9666666666666667
982 Accuracy/val/PH settlement conduct: 0.9698795180722891
983 Accuracy/train/What counts as "Final Adjudication": 0.9757575757575757
984 Accuracy/val/What counts as "Final Adjudication": 0.9939759036144579
985 Accuracy/train/Insurer settlement conduct: 0.9
986 Accuracy/val/Insurer settlement conduct: 0.8614457831325302
987 Accuracy/train/Other exclusion issues: 0.8166666666666667
988 Accuracy/val/Other exclusion issues: 0.7951807228915663
989 Accuracy/train/Other insurance (i.e. which insurance has to pay first):
0.9318181818181818
990 Accuracy/val/Other insurance (i.e. which insurance has to pay first):
0.8975903614457831
991 Accuracy/train/Allocation: 0.9257575757575758
992 Accuracy/val/Allocation: 0.9397590361445783
993 Accuracy/train/Arbitration: 0.9939393939393939
994 Accuracy/val/Arbitration: 0.9939759036144579
995 Accuracy/train/Bad faith: 0.7151515151515152
996 Accuracy/val/Bad faith: 0.7048192771084337
997 Accuracy/train/Other Coverage Issues: 0.8575757575757575
998 Accuracy/val/Other Coverage Issues: 0.8433734939759037
999 F1/train/Bodily injury: 0.0
1000 F1/val/Bodily injury: 0.0
1001 F1/train/Didn't settle when should have: 0.08

1002 F1/val/Didn't settle when should have: 0.0
1003 F1/train/Fraud/criminal/illegal conduct: 0.02247191011235955
1004 F1/val/Fraud/criminal/illegal conduct: 0.0
1005 F1/train/Restitution/ disgorgement is not "Loss": 0.0
1006 F1/val/Restitution/ disgorgement is not "Loss": 0.0
1007 F1/train/Settlement amount unreasonable: 0.0
1008 F1/val/Settlement amount unreasonable: nan
1009 F1/train/What is a "Claim"?: 0.024390243902439025
1010 F1/val/What is a "Claim"?: 0.0
1011 F1/train/"Insured" v "Insured": 0.0
1012 F1/val/"Insured" v "Insured": 0.0
1013 F1/train/Failed to get insurer consent: 0.0
1014 F1/val/Failed to get insurer consent: 0.0
1015 F1/train/Other "Loss" issues: 0.0
1016 F1/val/Other "Loss" issues: 0.0
1017 F1/train/Other issues arising from insurer settlement conduct: 0.0
1018 F1/val/Other issues arising from insurer settlement conduct: 0.0
1019 F1/train/Property damage: 0.0
1020 F1/val/Property damage: 0.0
1021 F1/train/What is a "Securities Claim"?: 0.09523809523809523
1022 F1/val/What is a "Securities Claim"?: 0.0
1023 F1/train/Libel or slander: 0.0
1024 F1/val/Libel or slander: 0.0
1025 F1/train/Other issues arising from PH settlement conduct: 0.0
1026 F1/val/Other issues arising from PH settlement conduct: 0.0
1027 F1/train/Unjust enrichment (profits not entitled): 0.030303030303030304
1028 F1/val/Unjust enrichment (profits not entitled): 0.0
1029 F1/train/What is a "Related Claim" / "Interrelated Wrongful Act"?:
0.09900990099009901
1030 F1/val/What is a "Related Claim" / "Interrelated Wrongful Act"?: 0.0625
1031 F1/train/Employment practices: 0.0
1032 F1/val/Employment practices: 0.0
1033 F1/train/Prior claim / notice: 0.0
1034 F1/val/Prior claim / notice: 0.0
1035 F1/train/What counts as "Loss"?: 0.03225806451612903
1036 F1/val/What counts as "Loss"?: 0.0
1037 F1/train/Prior acts: 0.0
1038 F1/val/Prior acts: 0.0
1039 F1/train/Professional services: 0.0
1040 F1/val/Professional services: 0.0
1041 F1/train/Who is an "Insured"?: 0.0
1042 F1/val/Who is an "Insured"?: 0.0
1043 F1/train/Fiduciary liability: 0.0
1044 F1/val/Fiduciary liability: nan
1045 F1/train/Prior or pending litigation / proceeding: 0.0
1046 F1/val/Prior or pending litigation / proceeding: 0.0
1047 F1/train/Wrongful act not in capacity as a director or officer of the insured (
includes cases involving an "other capacity" exclusion): 0.041666666666666664
1048 F1/val/Wrongful act not in capacity as a director or officer of the insured (
includes cases involving an "other capacity" exclusion): 0.0
1049 F1/train/Cyber: nan
1050 F1/val/Cyber: nan
1051 F1/train/Late notice or reporting issue: 0.10526315789473684
1052 F1/val/Late notice or reporting issue: 0.05714285714285714
1053 F1/train/Prior knowledge: 0.0
1054 F1/val/Prior knowledge: 0.0
1055 F1/train/Bump up: 0.0
1056 F1/val/Bump up: 0.0

```
1057 F1/train/Retro date issue: 0.0
1058 F1/val/Retro date issue: 0.0
1059 F1/train/Misrepresentation/Rescission: 0.031746031746031744
1060 F1/val/Misrepresentation/Rescission: 0.0
1061 F1/train/Regulatory: 0.0
1062 F1/val/Regulatory: nan
1063 F1/train/Insolvency: 0.0
1064 F1/val/Insolvency: nan
1065 F1/train/Market segmentation exclusion issues: 0.06349206349206349
1066 F1/val/Market segmentation exclusion issues: 0.0
1067 F1/train/Contract: 0.0
1068 F1/val/Contract: 0.0
1069 F1/train/Exclusion issues: 0.5043478260869565
1070 F1/val/Exclusion issues: 0.4444444444444444
1071 F1/train/Antitrust/restraint of trade/unfair business practice: 0.0
1072 F1/val/Antitrust/restraint of trade/unfair business practice: 0.0
1073 F1/train/Severability: 0.0
1074 F1/val/Severability: 0.0
1075 F1/train/Insurer refused to pay defense (be sure to check the reasons why):
      0.7350993377483444
1076 F1/val/Insurer refused to pay defense (be sure to check the reasons why):
      0.7391304347826086
1077 F1/train/Privacy/IP: 0.0
1078 F1/val/Privacy/IP: 0.0
1079 F1/train/Laser exclusion: 0.0
1080 F1/val/Laser exclusion: 0.0
1081 F1/train/PH failed to cooperate: 0.0
1082 F1/val/PH failed to cooperate: 0.0
1083 F1/train/PH settlement conduct: 0.0
1084 F1/val/PH settlement conduct: 0.0
1085 F1/train/What counts as "Final Adjudication": 0.0
1086 F1/val/What counts as "Final Adjudication": 0.0
1087 F1/train/Insurer settlement conduct: 0.0
1088 F1/val/Insurer settlement conduct: 0.0
1089 F1/train/Other exclusion issues: 0.06201550387596899
1090 F1/val/Other exclusion issues: 0.0
1091 F1/train/Other insurance (i.e. which insurance has to pay first): 0.0
1092 F1/val/Other insurance (i.e. which insurance has to pay first): 0.0
1093 F1/train/Allocation: 0.0
1094 F1/val/Allocation: 0.0
1095 F1/train/Arbitration: 0.0
1096 F1/val/Arbitration: 0.0
1097 F1/train/Bad faith: 0.27692307692307694
1098 F1/val/Bad faith: 0.36363636363636365
1099 F1/train/Other Coverage Issues: 0.04081632653061224
1100 F1/val/Other Coverage Issues: 0.0
1101 Total Train Accuracy: 0.9244800950683304
1102 Total Val Accuracy: 0.9118828254193243
1103 F1 Train 0.2875560538116592
1104 F1 Val 0.24493927125506074
1105 100%|| 83/83 [57:52<00:00, 41.84s/it]
1106 100%|| 21/21 [03:58<00:00, 11.34s/it]
1107 Epoch: 5
1108 Train Loss: 10.773490130183209
1109 Val Loss: 13.041338602701822
1110 Accuracy/train/Bodily injury: 0.9681818181818181
1111 Accuracy/val/Bodily injury: 0.9216867469879518
1112 Accuracy/train/Didn't settle when should have: 0.9333333333333333
```

1113 Accuracy/val/Didn't settle when should have: 0.9578313253012049
1114 Accuracy/train/Fraud/criminal/illegal conduct: 0.8666666666666667
1115 Accuracy/val/Fraud/criminal/illegal conduct: 0.8132530120481928
1116 Accuracy/train/Restitution/ disgorgement is not "Loss": 0.9545454545454546
1117 Accuracy/val/Restitution/ disgorgement is not "Loss": 0.9337349397590361
1118 Accuracy/train/Settlement amount unreasonable: 0.996969696969697
1119 Accuracy/val/Settlement amount unreasonable: 1.0
1120 Accuracy/train/What is a "Claim"?: 0.8787878787878788
1121 Accuracy/val/What is a "Claim"?: 0.7891566265060241
1122 Accuracy/train/"Insured" v "Insured": 0.9090909090909091
1123 Accuracy/val/"Insured" v "Insured": 0.891566265060241
1124 Accuracy/train/Failed to get insurer consent: 0.9757575757575757
1125 Accuracy/val/Failed to get insurer consent: 0.9759036144578314
1126 Accuracy/train/Other "Loss" issues: 0.9136363636363637
1127 Accuracy/val/Other "Loss" issues: 0.9156626506024096
1128 Accuracy/train/Other issues arising from insurer settlement conduct:
0.9530303030303030
1129 Accuracy/val/Other issues arising from insurer settlement conduct:
0.891566265060241
1130 Accuracy/train/Property damage: 0.9787878787878788
1131 Accuracy/val/Property damage: 0.9397590361445783
1132 Accuracy/train/What is a "Securities Claim"?: 0.9712121212121212
1133 Accuracy/val/What is a "Securities Claim"?: 0.9578313253012049
1134 Accuracy/train/Libel or slander: 0.996969696969697
1135 Accuracy/val/Libel or slander: 0.9939759036144579
1136 Accuracy/train/Other issues arising from PH settlement conduct: 0.9878787878787879
1137 Accuracy/val/Other issues arising from PH settlement conduct: 0.9939759036144579
1138 Accuracy/train/Unjust enrichment (profits not entitled): 0.9030303030303031
1139 Accuracy/val/Unjust enrichment (profits not entitled): 0.8614457831325302
1140 Accuracy/train/What is a "Related Claim" / "Interrelated Wrongful Act"?:
0.8666666666666667
1141 Accuracy/val/What is a "Related Claim" / "Interrelated Wrongful Act"?:
0.8132530120481928
1142 Accuracy/train/Employment practices: 0.9833333333333333
1143 Accuracy/val/Employment practices: 0.9939759036144579
1144 Accuracy/train/Prior claim / notice: 0.9181818181818182
1145 Accuracy/val/Prior claim / notice: 0.8855421686746988
1146 Accuracy/train/What counts as "Loss"?: 0.8151515151515152
1147 Accuracy/val/What counts as "Loss"?: 0.7530120481927711
1148 Accuracy/train/Prior acts: 0.9515151515151515
1149 Accuracy/val/Prior acts: 0.9457831325301205
1150 Accuracy/train/Professional services: 0.9242424242424242
1151 Accuracy/val/Professional services: 0.927710843373494
1152 Accuracy/train/Who is an "Insured"?: 0.896969696969697
1153 Accuracy/val/Who is an "Insured"?: 0.8554216867469879
1154 Accuracy/train/Fiduciary liability: 0.996969696969697
1155 Accuracy/val/Fiduciary liability: 1.0
1156 Accuracy/train/Prior or pending litigation / proceeding: 0.9227272727272727
1157 Accuracy/val/Prior or pending litigation / proceeding: 0.927710843373494
1158 Accuracy/train/Wrongful act not in capacity as a director or officer of the insured
(includes cases involving an "other capacity" exclusion): 0.9333333333333333
1159 Accuracy/val/Wrongful act not in capacity as a director or officer of the insured (
includes cases involving an "other capacity" exclusion): 0.8795180722891566
1160 Accuracy/train/Cyber: 1.0
1161 Accuracy/val/Cyber: 1.0
1162 Accuracy/train/Late notice or reporting issue: 0.8681818181818182
1163 Accuracy/val/Late notice or reporting issue: 0.8012048192771084
1164 Accuracy/train/Prior knowledge: 0.9454545454545454

1165 Accuracy/val/Prior knowledge: 0.9698795180722891
1166 Accuracy/train/Bump up: 0.9878787878787879
1167 Accuracy/val/Bump up: 0.9759036144578314
1168 Accuracy/train/Retro date issue: 0.9727272727272728
1169 Accuracy/val/Retro date issue: 0.9759036144578314
1170 Accuracy/train/Misrepresentation/Rescission: 0.9075757575757576
1171 Accuracy/val/Misrepresentation/Rescission: 0.9216867469879518
1172 Accuracy/train/Regulatory: 0.9939393939393939
1173 Accuracy/val/Regulatory: 1.0
1174 Accuracy/train/Insolvency: 0.9924242424242424
1175 Accuracy/val/Insolvency: 1.0
1176 Accuracy/train/Market segmentation exclusion issues: 0.9121212121212121
1177 Accuracy/val/Market segmentation exclusion issues: 0.8975903614457831
1178 Accuracy/train/Contract: 0.896969696969697
1179 Accuracy/val/Contract: 0.8734939759036144
1180 Accuracy/train/Exclusion issues: 0.6606060606060606
1181 Accuracy/val/Exclusion issues: 0.6686746987951807
1182 Accuracy/train/Antitrust/restraint of trade/unfair business practice:
0.9772727272727273
1183 Accuracy/val/Antitrust/restraint of trade/unfair business practice:
0.9819277108433735
1184 Accuracy/train/Severability: 0.9984848484848485
1185 Accuracy/val/Severability: 0.9939759036144579
1186 Accuracy/train/Insurer refused to pay defense (be sure to check the reasons why):
0.6621212121212121
1187 Accuracy/val/Insurer refused to pay defense (be sure to check the reasons why):
0.5903614457831325
1188 Accuracy/train/Privacy/IP: 0.9878787878787879
1189 Accuracy/val/Privacy/IP: 0.9698795180722891
1190 Accuracy/train/Laser exclusion: 0.9727272727272728
1191 Accuracy/val/Laser exclusion: 0.9759036144578314
1192 Accuracy/train/PH failed to cooperate: 0.9787878787878788
1193 Accuracy/val/PH failed to cooperate: 0.9698795180722891
1194 Accuracy/train/PH settlement conduct: 0.9666666666666667
1195 Accuracy/val/PH settlement conduct: 0.9698795180722891
1196 Accuracy/train/What counts as "Final Adjudication": 0.9757575757575757
1197 Accuracy/val/What counts as "Final Adjudication": 0.9939759036144579
1198 Accuracy/train/Insurer settlement conduct: 0.9015151515151515
1199 Accuracy/val/Insurer settlement conduct: 0.8493975903614458
1200 Accuracy/train/Other exclusion issues: 0.8287878787878787
1201 Accuracy/val/Other exclusion issues: 0.7771084337349398
1202 Accuracy/train/Other insurance (i.e. which insurance has to pay first):
0.9333333333333333
1203 Accuracy/val/Other insurance (i.e. which insurance has to pay first):
0.891566265060241
1204 Accuracy/train/Allocation: 0.9242424242424242
1205 Accuracy/val/Allocation: 0.9397590361445783
1206 Accuracy/train/Arbitration: 0.9939393939393939
1207 Accuracy/val/Arbitration: 0.9939759036144579
1208 Accuracy/train/Bad faith: 0.7257575757575757
1209 Accuracy/val/Bad faith: 0.7168674698795181
1210 Accuracy/train/Other Coverage Issues: 0.8666666666666667
1211 Accuracy/val/Other Coverage Issues: 0.8433734939759037
1212 F1/train/Bodily injury: 0.0
1213 F1/val/Bodily injury: 0.0
1214 F1/train/Didn't settle when should have: 0.15384615384615385
1215 F1/val/Didn't settle when should have: 0.0
1216 F1/train/Fraud/criminal/illegal conduct: 0.043478260869565216


```

1217 F1/val/Fraud/criminal/illegal conduct: 0.06060606060606061
1218 F1/train/Restitution/ disgorgement is not "Loss": 0.0625
1219 F1/val/Restitution/ disgorgement is not "Loss": 0.0
1220 F1/train/Settlement amount unreasonable: 0.0
1221 F1/val/Settlement amount unreasonable: nan
1222 F1/train/What is a "Claim"?: 0.047619047619047616
1223 F1/val/What is a "Claim"?: 0.0
1224 F1/train/"Insured" v "Insured": 0.0
1225 F1/val/"Insured" v "Insured": 0.0
1226 F1/train/Failed to get insurer consent: 0.0
1227 F1/val/Failed to get insurer consent: 0.0
1228 F1/train/Other "Loss" issues: 0.0
1229 F1/val/Other "Loss" issues: 0.0
1230 F1/train/Other issues arising from insurer settlement conduct: 0.0
1231 F1/val/Other issues arising from insurer settlement conduct: 0.0
1232 F1/train/Property damage: 0.125
1233 F1/val/Property damage: 0.0
1234 F1/train/What is a "Securities Claim"?: 0.09523809523809523
1235 F1/val/What is a "Securities Claim"?: 0.0
1236 F1/train/Libel or slander: 0.0
1237 F1/val/Libel or slander: 0.0
1238 F1/train/Other issues arising from PH settlement conduct: 0.0
1239 F1/val/Other issues arising from PH settlement conduct: 0.0
1240 F1/train/Unjust enrichment (profits not entitled): 0.030303030303030304
1241 F1/val/Unjust enrichment (profits not entitled): 0.0
1242 F1/train/What is a "Related Claim" / "Interrelated Wrongful Act"?: 0.12
1243 F1/val/What is a "Related Claim" / "Interrelated Wrongful Act"?:
    0.06060606060606061
1244 F1/train/Employment practices: 0.0
1245 F1/val/Employment practices: 0.0
1246 F1/train/Prior claim / notice: 0.03571428571428571
1247 F1/val/Prior claim / notice: 0.0
1248 F1/train/What counts as "Loss"?: 0.016129032258064516
1249 F1/val/What counts as "Loss"?: 0.046511627906976744
1250 F1/train/Prior acts: 0.0
1251 F1/val/Prior acts: 0.0
1252 F1/train/Professional services: 0.0
1253 F1/val/Professional services: 0.0
1254 F1/train/Who is an "Insured"?: 0.08108108108108109
1255 F1/val/Who is an "Insured"?: 0.0
1256 F1/train/Fiduciary liability: 0.0
1257 F1/val/Fiduciary liability: nan
1258 F1/train/Prior or pending litigation / proceeding: 0.03773584905660377
1259 F1/val/Prior or pending litigation / proceeding: 0.0
1260 F1/train/Wrongful act not in capacity as a director or officer of the insured (
    includes cases involving an "other capacity" exclusion): 0.08333333333333333
1261 F1/val/Wrongful act not in capacity as a director or officer of the insured (
    includes cases involving an "other capacity" exclusion): 0.0
1262 F1/train/Cyber: nan
1263 F1/val/Cyber: nan
1264 F1/train/Late notice or reporting issue: 0.06451612903225806
1265 F1/val/Late notice or reporting issue: 0.05714285714285714
1266 F1/train/Prior knowledge: 0.0
1267 F1/val/Prior knowledge: 0.0
1268 F1/train/Bump up: 0.0
1269 F1/val/Bump up: 0.0
1270 F1/train/Retro date issue: 0.0
1271 F1/val/Retro date issue: 0.0

```

```

1272 F1/train/Misrepresentation/Rescission: 0.031746031746031744
1273 F1/val/Misrepresentation/Rescission: 0.13333333333333333
1274 F1/train/Regulatory: 0.0
1275 F1/val/Regulatory: nan
1276 F1/train/Insolvency: 0.0
1277 F1/val/Insolvency: nan
1278 F1/train/Market segmentation exclusion issues: 0.09375
1279 F1/val/Market segmentation exclusion issues: 0.0
1280 F1/train/Contract: 0.02857142857142857
1281 F1/val/Contract: 0.0
1282 F1/train/Exclusion issues: 0.5172413793103449
1283 F1/val/Exclusion issues: 0.38202247191011235
1284 F1/train/Antitrust/restraint of trade/unfair business practice: 0.0
1285 F1/val/Antitrust/restraint of trade/unfair business practice: 0.0
1286 F1/train/Severability: 0.0
1287 F1/val/Severability: 0.0
1288 F1/train/Insurer refused to pay defense (be sure to check the reasons why):
      0.7480225988700565
1289 F1/val/Insurer refused to pay defense (be sure to check the reasons why): 0.575
1290 F1/train/Privacy/IP: 0.0
1291 F1/val/Privacy/IP: 0.0
1292 F1/train/Laser exclusion: 0.0
1293 F1/val/Laser exclusion: 0.0
1294 F1/train/PH failed to cooperate: 0.0
1295 F1/val/PH failed to cooperate: 0.0
1296 F1/train/PH settlement conduct: 0.0
1297 F1/val/PH settlement conduct: 0.0
1298 F1/train/What counts as "Final Adjudication": 0.0
1299 F1/val/What counts as "Final Adjudication": 0.0
1300 F1/train/Insurer settlement conduct: 0.029850746268656716
1301 F1/val/Insurer settlement conduct: 0.0
1302 F1/train/Other exclusion issues: 0.13740458015267176
1303 F1/val/Other exclusion issues: 0.05128205128205128
1304 F1/train/Other insurance (i.e. which insurance has to pay first):
      0.043478260869565216
1305 F1/val/Other insurance (i.e. which insurance has to pay first): 0.0
1306 F1/train/Allocation: 0.0
1307 F1/val/Allocation: 0.0
1308 F1/train/Arbitration: 0.0
1309 F1/val/Arbitration: 0.0
1310 F1/train/Bad faith: 0.3418181818181818
1311 F1/val/Bad faith: 0.0784313725490196
1312 F1/train/Other Coverage Issues: 0.15384615384615385
1313 F1/val/Other Coverage Issues: 0.0
1314 Total Train Accuracy: 0.9260546642899584
1315 Total Val Accuracy: 0.9090479565320104
1316 F1 Train 0.30687830687830686
1317 F1 Val 0.15570175438596492
1318 100%|| 83/83 [57:52<00:00, 41.84s/it]
1319 100%|| 21/21 [03:58<00:00, 11.34s/it]
1320 Epoch: 6
1321 Train Loss: 10.205644843090012
1322 Val Loss: 12.979580061776298
1323 Accuracy/train/Bodily injury: 0.9681818181818181
1324 Accuracy/val/Bodily injury: 0.9216867469879518
1325 Accuracy/train/Didn't settle when should have: 0.9318181818181818
1326 Accuracy/val/Didn't settle when should have: 0.9518072289156626
1327 Accuracy/train/Fraud/criminal/illegal conduct: 0.8818181818181818

```

1328 Accuracy/val/Fraud/criminal/illegal conduct: 0.8192771084337349
1329 Accuracy/train/Restitution/ disgorgement is not "Loss": 0.9545454545454546
1330 Accuracy/val/Restitution/ disgorgement is not "Loss": 0.9337349397590361
1331 Accuracy/train/Settlement amount unreasonable: 0.996969696969697
1332 Accuracy/val/Settlement amount unreasonable: 1.0
1333 Accuracy/train/What is a "Claim"?: 0.8803030303030303
1334 Accuracy/val/What is a "Claim"?: 0.7891566265060241
1335 Accuracy/train/"Insured" v "Insured": 0.9090909090909091
1336 Accuracy/val/"Insured" v "Insured": 0.891566265060241
1337 Accuracy/train/Failed to get insurer consent: 0.9757575757575757
1338 Accuracy/val/Failed to get insurer consent: 0.9759036144578314
1339 Accuracy/train/Other "Loss" issues: 0.9136363636363637
1340 Accuracy/val/Other "Loss" issues: 0.9156626506024096
1341 Accuracy/train/Other issues arising from insurer settlement conduct:
0.95303030303030303
1342 Accuracy/val/Other issues arising from insurer settlement conduct:
0.891566265060241
1343 Accuracy/train/Property damage: 0.9787878787878788
1344 Accuracy/val/Property damage: 0.9397590361445783
1345 Accuracy/train/What is a "Securities Claim"?: 0.9712121212121212
1346 Accuracy/val/What is a "Securities Claim"?: 0.9578313253012049
1347 Accuracy/train/Libel or slander: 0.996969696969697
1348 Accuracy/val/Libel or slander: 0.9939759036144579
1349 Accuracy/train/Other issues arising from PH settlement conduct: 0.9878787878787879
1350 Accuracy/val/Other issues arising from PH settlement conduct: 0.9939759036144579
1351 Accuracy/train/Unjust enrichment (profits not entitled): 0.9045454545454545
1352 Accuracy/val/Unjust enrichment (profits not entitled): 0.8614457831325302
1353 Accuracy/train/What is a "Related Claim" / "Interrelated Wrongful Act"?:
0.8636363636363636
1354 Accuracy/val/What is a "Related Claim" / "Interrelated Wrongful Act"?:
0.8132530120481928
1355 Accuracy/train/Employment practices: 0.9833333333333333
1356 Accuracy/val/Employment practices: 0.9939759036144579
1357 Accuracy/train/Prior claim / notice: 0.9212121212121213
1358 Accuracy/val/Prior claim / notice: 0.8855421686746988
1359 Accuracy/train/What counts as "Loss"?: 0.8212121212121212
1360 Accuracy/val/What counts as "Loss"?: 0.7590361445783133
1361 Accuracy/train/Prior acts: 0.9515151515151515
1362 Accuracy/val/Prior acts: 0.9457831325301205
1363 Accuracy/train/Professional services: 0.9272727272727272
1364 Accuracy/val/Professional services: 0.927710843373494
1365 Accuracy/train/Who is an "Insured"?: 0.8984848484848484
1366 Accuracy/val/Who is an "Insured"?: 0.8554216867469879
1367 Accuracy/train/Fiduciary liability: 0.996969696969697
1368 Accuracy/val/Fiduciary liability: 1.0
1369 Accuracy/train/Prior or pending litigation / proceeding: 0.9227272727272727
1370 Accuracy/val/Prior or pending litigation / proceeding: 0.927710843373494
1371 Accuracy/train/Wrongful act not in capacity as a director or officer of the insured
(includes cases involving an "other capacity" exclusion): 0.9348484848484848
1372 Accuracy/val/Wrongful act not in capacity as a director or officer of the insured (
includes cases involving an "other capacity" exclusion): 0.8795180722891566
1373 Accuracy/train/Cyber: 1.0
1374 Accuracy/val/Cyber: 1.0
1375 Accuracy/train/Late notice or reporting issue: 0.8742424242424243
1376 Accuracy/val/Late notice or reporting issue: 0.8072289156626506
1377 Accuracy/train/Prior knowledge: 0.9454545454545454
1378 Accuracy/val/Prior knowledge: 0.963855421686747
1379 Accuracy/train/Bump up: 0.9878787878787879

1380 Accuracy/val/Bump up: 0.9759036144578314
1381 Accuracy/train/Retro date issue: 0.9727272727272728
1382 Accuracy/val/Retro date issue: 0.9759036144578314
1383 Accuracy/train/Misrepresentation/Rescission: 0.9136363636363637
1384 Accuracy/val/Misrepresentation/Rescission: 0.9216867469879518
1385 Accuracy/train/Regulatory: 0.9939393939393939
1386 Accuracy/val/Regulatory: 1.0
1387 Accuracy/train/Insolvency: 0.9924242424242424
1388 Accuracy/val/Insolvency: 1.0
1389 Accuracy/train/Market segmentation exclusion issues: 0.9106060606060606
1390 Accuracy/val/Market segmentation exclusion issues: 0.8975903614457831
1391 Accuracy/train/Contract: 0.8984848484848484
1392 Accuracy/val/Contract: 0.8734939759036144
1393 Accuracy/train/Exclusion issues: 0.7181818181818181
1394 Accuracy/val/Exclusion issues: 0.6385542168674698
1395 Accuracy/train/Antitrust/restraint of trade/unfair business practice:
0.9787878787878788
1396 Accuracy/val/Antitrust/restraint of trade/unfair business practice:
0.9819277108433735
1397 Accuracy/train/Severability: 0.9984848484848485
1398 Accuracy/val/Severability: 0.9939759036144579
1399 Accuracy/train/Insurer refused to pay defense (be sure to check the reasons why):
0.6863636363636364
1400 Accuracy/val/Insurer refused to pay defense (be sure to check the reasons why):
0.6265060240963856
1401 Accuracy/train/Privacy/IP: 0.9878787878787879
1402 Accuracy/val/Privacy/IP: 0.9698795180722891
1403 Accuracy/train/Laser exclusion: 0.9727272727272728
1404 Accuracy/val/Laser exclusion: 0.9759036144578314
1405 Accuracy/train/PH failed to cooperate: 0.9787878787878788
1406 Accuracy/val/PH failed to cooperate: 0.9698795180722891
1407 Accuracy/train/PH settlement conduct: 0.9666666666666667
1408 Accuracy/val/PH settlement conduct: 0.9698795180722891
1409 Accuracy/train/What counts as "Final Adjudication": 0.9757575757575757
1410 Accuracy/val/What counts as "Final Adjudication": 0.9939759036144579
1411 Accuracy/train/Insurer settlement conduct: 0.9030303030303031
1412 Accuracy/val/Insurer settlement conduct: 0.8493975903614458
1413 Accuracy/train/Other exclusion issues: 0.8378787878787879
1414 Accuracy/val/Other exclusion issues: 0.7710843373493976
1415 Accuracy/train/Other insurance (i.e. which insurance has to pay first):
0.9363636363636364
1416 Accuracy/val/Other insurance (i.e. which insurance has to pay first):
0.891566265060241
1417 Accuracy/train/Allocation: 0.9257575757575758
1418 Accuracy/val/Allocation: 0.9457831325301205
1419 Accuracy/train/Arbitration: 0.9939393939393939
1420 Accuracy/val/Arbitration: 0.9939759036144579
1421 Accuracy/train/Bad faith: 0.7924242424242425
1422 Accuracy/val/Bad faith: 0.7289156626506024
1423 Accuracy/train/Other Coverage Issues: 0.8666666666666667
1424 Accuracy/val/Other Coverage Issues: 0.8433734939759037
1425 F1/train/Bodily injury: 0.0
1426 F1/val/Bodily injury: 0.0
1427 F1/train/Didn't settle when should have: 0.11764705882352941
1428 F1/val/Didn't settle when should have: 0.0
1429 F1/train/Fraud/criminal/illegal conduct: 0.20408163265306123
1430 F1/val/Fraud/criminal/illegal conduct: 0.0625
1431 F1/train/Restitution/ disgorgement is not "Loss": 0.0625

```

1432 F1/val/Restitution/ disgorgement is not "Loss": 0.0
1433 F1/train/Settlement amount unreasonable: 0.0
1434 F1/val/Settlement amount unreasonable: nan
1435 F1/train/What is a "Claim"?: 0.09195402298850575
1436 F1/val/What is a "Claim"?: 0.0
1437 F1/train/"Insured" v "Insured": 0.0
1438 F1/val/"Insured" v "Insured": 0.0
1439 F1/train/Failed to get insurer consent: 0.0
1440 F1/val/Failed to get insurer consent: 0.0
1441 F1/train/Other "Loss" issues: 0.0
1442 F1/val/Other "Loss" issues: 0.0
1443 F1/train/Other issues arising from insurer settlement conduct: 0.0
1444 F1/val/Other issues arising from insurer settlement conduct: 0.0
1445 F1/train/Property damage: 0.125
1446 F1/val/Property damage: 0.0
1447 F1/train/What is a "Securities Claim"?: 0.09523809523809523
1448 F1/val/What is a "Securities Claim"?: 0.0
1449 F1/train/Libel or slander: 0.0
1450 F1/val/Libel or slander: 0.0
1451 F1/train/Other issues arising from PH settlement conduct: 0.0
1452 F1/val/Other issues arising from PH settlement conduct: 0.0
1453 F1/train/Unjust enrichment (profits not entitled): 0.05970149253731343
1454 F1/val/Unjust enrichment (profits not entitled): 0.0
1455 F1/train/What is a "Related Claim" / "Interrelated Wrongful Act"?:
    0.1346153846153846
1456 F1/val/What is a "Related Claim" / "Interrelated Wrongful Act"?:
    0.06060606060606061
1457 F1/train/Employment practices: 0.0
1458 F1/val/Employment practices: 0.0
1459 F1/train/Prior claim / notice: 0.10344827586206896
1460 F1/val/Prior claim / notice: 0.0
1461 F1/train/What counts as "Loss"?: 0.1917808219178082
1462 F1/val/What counts as "Loss"?: 0.16666666666666666
1463 F1/train/Prior acts: 0.0
1464 F1/val/Prior acts: 0.0
1465 F1/train/Professional services: 0.07692307692307693
1466 F1/val/Professional services: 0.0
1467 F1/train/Who is an "Insured"?: 0.10666666666666667
1468 F1/val/Who is an "Insured"?: 0.0
1469 F1/train/Fiduciary liability: 0.0
1470 F1/val/Fiduciary liability: nan
1471 F1/train/Prior or pending litigation / proceeding: 0.07272727272727272
1472 F1/val/Prior or pending litigation / proceeding: 0.0
1473 F1/train/Wrongful act not in capacity as a director or officer of the insured (
    includes cases involving an "other capacity" exclusion): 0.12244897959183673
1474 F1/val/Wrongful act not in capacity as a director or officer of the insured (
    includes cases involving an "other capacity" exclusion): 0.0
1475 F1/train/Cyber: nan
1476 F1/val/Cyber: nan
1477 F1/train/Late notice or reporting issue: 0.14432989690721648
1478 F1/val/Late notice or reporting issue: 0.058823529411764705
1479 F1/train/Prior knowledge: 0.0
1480 F1/val/Prior knowledge: 0.0
1481 F1/train/Bump up: 0.0
1482 F1/val/Bump up: 0.0
1483 F1/train/Retro date issue: 0.0
1484 F1/val/Retro date issue: 0.0
1485 F1/train/Misrepresentation/Rescission: 0.14925373134328357

```

```

1486 F1/val/Misrepresentation/Rescission: 0.0
1487 F1/train/Regulatory: 0.0
1488 F1/val/Regulatory: nan
1489 F1/train/Insolvency: 0.0
1490 F1/val/Insolvency: nan
1491 F1/train/Market segmentation exclusion issues: 0.06349206349206349
1492 F1/val/Market segmentation exclusion issues: 0.0
1493 F1/train/Contract: 0.056338028169014086
1494 F1/val/Contract: 0.0
1495 F1/train/Exclusion issues: 0.5974025974025974
1496 F1/val/Exclusion issues: 0.45454545454545453
1497 F1/train/Antitrust/restraint of trade/unfair business practice: 0.125
1498 F1/val/Antitrust/restraint of trade/unfair business practice: 0.0
1499 F1/train/Severability: 0.0
1500 F1/val/Severability: 0.0
1501 F1/train/Insurer refused to pay defense (be sure to check the reasons why):
      0.7590221187427241
1502 F1/val/Insurer refused to pay defense (be sure to check the reasons why): 0.69
1503 F1/train/Privacy/IP: 0.0
1504 F1/val/Privacy/IP: 0.0
1505 F1/train/Laser exclusion: 0.0
1506 F1/val/Laser exclusion: 0.0
1507 F1/train/PH failed to cooperate: 0.0
1508 F1/val/PH failed to cooperate: 0.0
1509 F1/train/PH settlement conduct: 0.0
1510 F1/val/PH settlement conduct: 0.0
1511 F1/train/What counts as "Final Adjudication": 0.0
1512 F1/val/What counts as "Final Adjudication": 0.0
1513 F1/train/Insurer settlement conduct: 0.058823529411764705
1514 F1/val/Insurer settlement conduct: 0.0
1515 F1/train/Other exclusion issues: 0.24113475177304963
1516 F1/val/Other exclusion issues: 0.0
1517 F1/train/Other insurance (i.e. which insurance has to pay first): 0.125
1518 F1/val/Other insurance (i.e. which insurance has to pay first): 0.0
1519 F1/train/Allocation: 0.0
1520 F1/val/Allocation: 0.0
1521 F1/train/Arbitration: 0.0
1522 F1/val/Arbitration: 0.0
1523 F1/train/Bad faith: 0.5566343042071198
1524 F1/val/Bad faith: 0.2857142857142857
1525 F1/train/Other Coverage Issues: 0.18518518518518517
1526 F1/val/Other Coverage Issues: 0.0
1527 Total Train Accuracy: 0.9300950683303625
1528 Total Val Accuracy: 0.9095204346798961
1529 F1 Train 0.35798090040927694
1530 F1 Val 0.2231237322515213
1531 80%| | 66/83 [46:17<11:56, 42.12s/it]slurmstepd: error: *** JOB 898362 ON at11
      -1-03-011-3-0 CANCELLED AT 2024-10-31T04:43:26 DUE TO TIME LIMIT ***
1532 -----
1533 Begin Slurm Epilog: Oct-31-2024 04:43:27
1534 Job ID: 898362
1535 Array Job ID: _4294967294
1536 User ID: mbock9
1537 Account: coc
1538 Job name: HGX_H100_Example
1539 Resources: cpu=1,gres/gpu:h100=1,mem=224G,node=1
1540 Rsrc Used: cput=08:00:11,vmem=0,walltime=08:00:11,mem=9334076K,energy_used=0
1541 Partition: coe-gpu

```

6 Results Visualization

Issue Name	Train Accuracy	Validation Accuracy
Bodily injury	0.9682	0.9217
Didn't settle when should have	0.9318	0.9518
Fraud/criminal/illegal conduct	0.8818	0.8193
Restitution/ disgorgement is not "Loss"	0.9545	0.9337
Settlement amount unreasonable	0.9970	1.0000
What is a "Claim"?	0.8803	0.7892
"Insured" v "Insured"	0.9091	0.8916
Failed to get insurer consent	0.9758	0.9759
Other "Loss" issues	0.9136	0.9157
Other issues arising from insurer settlement conduct	0.9530	0.8916
Property damage	0.9788	0.9398
What is a "Securities Claim"?	0.9712	0.9578
Libel or slander	0.9970	0.9940
Other issues arising from PH settlement conduct	0.9879	0.9940
Unjust enrichment (profits not entitled)	0.9045	0.8614
What is a "Related Claim" / "Interrelated Wrongful Act"?	0.8636	0.8133
Employment practices	0.9833	0.9940
Prior claim / notice	0.9212	0.8855
What counts as "Loss"?	0.8212	0.7590
Prior acts	0.9515	0.9458
Professional services	0.9273	0.9277
Who is an "Insured"?	0.8985	0.8554
Fiduciary liability	0.9970	1.0000

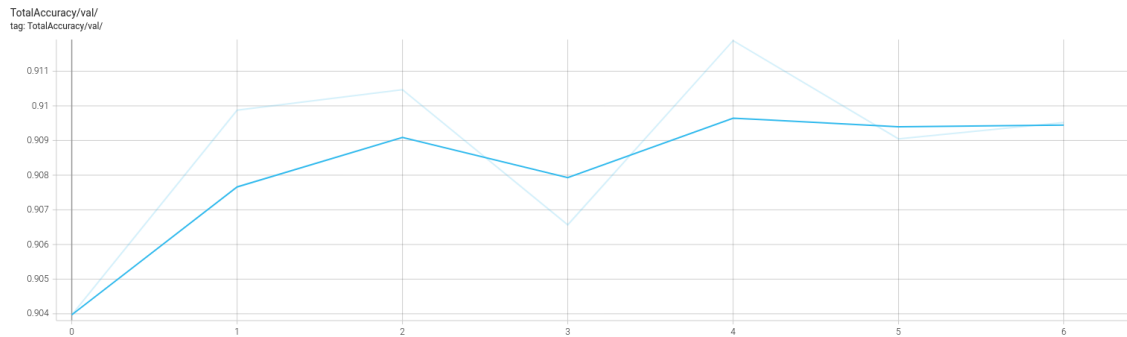


Figure 3: Validation Accuracy

Prior or pending litigation / proceeding	0.9227	0.9277
Wrongful act not in capacity as a director or officer of the insured	0.9348	0.8795
Cyber	1.0000	1.0000
Late notice or reporting issue	0.8742	0.8072
Prior knowledge	0.9455	0.9639
Bump up	0.9879	0.9759
Retro date issue	0.9727	0.9759
Misrepresentation/Rescission	0.9136	0.9217
Regulatory	0.9939	1.0000
Insolvency	0.9924	1.0000
Market segmentation exclusion issues	0.9106	0.8976
Contract	0.8985	0.8735
Exclusion issues	0.7182	0.6386
Antitrust/restraint of trade/unfair business practice	0.9788	0.9819
Severability	0.9985	0.9940
Insurer refused to pay defense	0.6864	0.6265
Privacy/IP	0.9879	0.9699
Laser exclusion	0.9727	0.9759
PH failed to cooperate	0.9788	0.9699
PH settlement conduct	0.9667	0.9699
What counts as "Final Adjudication"	0.9758	0.9940
Insurer settlement conduct	0.9030	0.8494
Other exclusion issues	0.8379	0.7711
Other insurance (which insurance has to pay first)	0.9364	0.8916
Allocation	0.9258	0.9458
Arbitration	0.9939	0.9940
Bad faith	0.7924	0.7289
Other Coverage Issues	0.8667	0.8434

Table 1: List of Issues with Train and Validation Accuracies (Generated by AI with my data)

Issue Name	Train F1 Score	Validation F1 Score
Bodily injury	0.0	0.0
Didn't settle when should have	0.1176	0.0
Fraud/criminal/illegal conduct	0.2041	0.0625
Restitution/ disgorgement is not "Loss"	0.0625	0.0
Settlement amount unreasonable	0.0	NaN
What is a "Claim"?	0.0920	0.0
"Insured" v "Insured"	0.0	0.0
Failed to get insurer consent	0.0	0.0
Other "Loss" issues	0.0	0.0
Other issues arising from insurer settlement conduct	0.0	0.0
Property damage	0.1250	0.0
What is a "Securities Claim"?	0.0952	0.0

Libel or slander	0.0	0.0
Other issues arising from PH settlement conduct	0.0	0.0
Unjust enrichment (profits not entitled)	0.0597	0.0
What is a "Related Claim" / "Interrelated Wrongful Act"?	0.1346	0.0606
Employment practices	0.0	0.0
Prior claim / notice	0.1034	0.0
What counts as "Loss"?	0.1918	0.1667
Prior acts	0.0	0.0
Professional services	0.0769	0.0
Who is an "Insured"?	0.1067	0.0
Fiduciary liability	0.0	NaN
Prior or pending litigation / proceeding	0.0727	0.0
Wrongful act not in capacity as a director or officer of the insured	0.1224	0.0
Cyber	NaN	NaN
Late notice or reporting issue	0.1443	0.0588
Prior knowledge	0.0	0.0
Bump up	0.0	0.0
Retro date issue	0.0	0.0
Misrepresentation/Rescission	0.1493	0.0
Regulatory	0.0	NaN
Insolvency	0.0	NaN
Market segmentation exclusion issues	0.0635	0.0
Contract	0.0563	0.0
Exclusion issues	0.5974	0.4545
Antitrust/restraint of trade/unfair business practice	0.1250	0.0
Severability	0.0	0.0
Insurer refused to pay defense	0.7590	0.6900
Privacy/IP	0.0	0.0
Laser exclusion	0.0	0.0
PH failed to cooperate	0.0	0.0
PH settlement conduct	0.0	0.0
What counts as "Final Adjudication"	0.0	0.0
Insurer settlement conduct	0.0588	0.0
Other exclusion issues	0.2411	0.0
Other insurance	0.1250	0.0
Allocation	0.0	0.0
Arbitration	0.0	0.0
Bad faith	0.5566	0.2857
Other Coverage Issues	0.1852	0.0

Table 2: List of Issues with Train and Validation F1 (Generated by AI with my data)

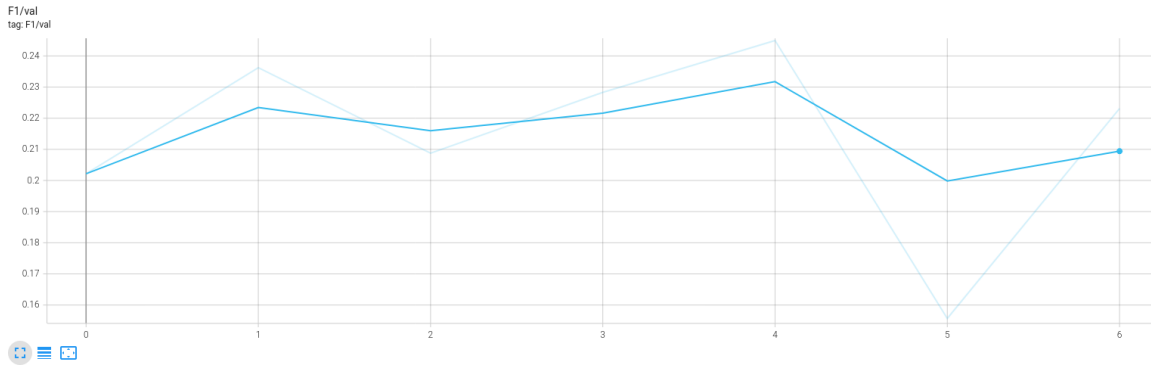


Figure 4: F1 Score

7 Proof of work

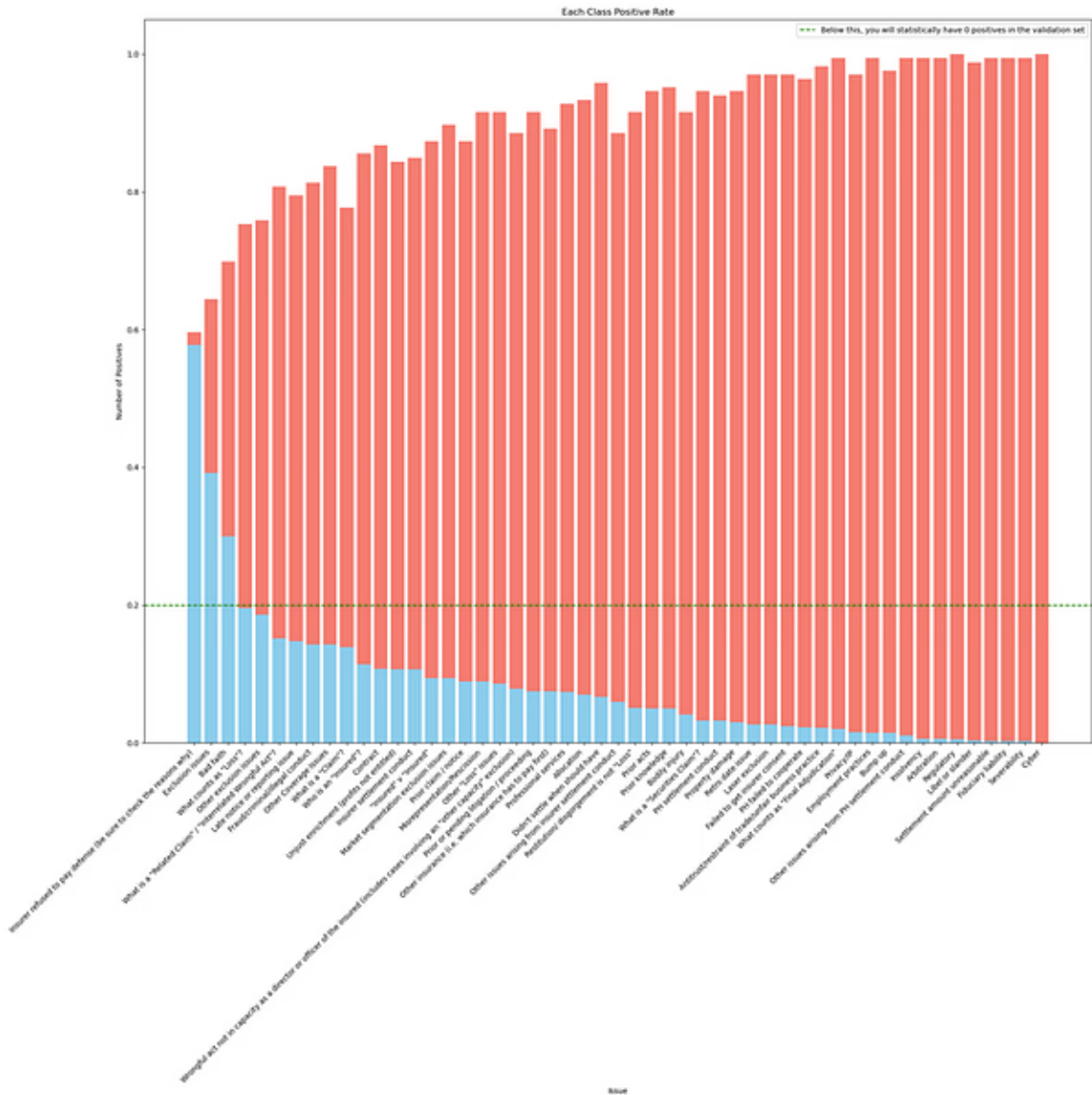


Figure 6: Blue shows the percent of cases which contain each issue. Red shows accuracy on each issue.

8 Next Week's proposal

- We believe that LLaMa had trouble understanding the insurance terminology, so Nathan advised me to fine tune. I have a script set up, but it isn't working because I can't figure out the context length of LLaMa. It says 128,000, but then I pass in that many tokens and it throws an error.

HAAG Research Report

NLP - Sentencias / NLP - Gen Team

Week 11

Víctor C. Fernández
November 2024

1 WEEKLY PROJECT UPDATES

What progress did you make in the last week?

- Adapted model querying by reducing the input text to a window of 2000 characters around the identified date.
- Completed a basic abstract for our paper within the overleaf template.
- Worked on a presentation/PPT for our future call with the judge Miguel.
- Began creating a pipeline to orchestrate the whole process using prefect.
- Researched on creating a simple UI to load and view results using Gradio.

What progress are you making next?

- Keep working on improving context retrieval for the dates.
- Finish preparing slides for call with Judge Miguel.
- Finalize implementation of a simple pipeline to run all processes end to end.
- Keep working on a simple UI to interact with the pipeline and upload and process a single file.
- Keep working on getting more content in for our paper.

Is there anything blocking you from making progress?

Yes, my PACE account has been blocked for exceeding my data usage quote, most likely for using larger models for retrieving date contexts. Already opened a ticket for this to the PACE team, but still awaiting response.

2 ABSTRACTS

1. **Title:** Extraction and Classification of Statute Facets using Few-shot Learning

• **URL:** <https://dl.acm.org/doi/10.1145/3594536.3595134> (may require accessing with GaTech credentials to view actual paper)

• **Abstract:** In this paper, we focus on automatic extraction of statute facets from legal statutes such as Act documents. We define statute facets to be key specific aspects of a statute which can potentially be used in legal arguments. For example, Section 25F of the Industrial Disputes Act (India) contains statute facets such as workman, employer, retrenchment of workmen, continuous service for not less than one year, etc. Such statute facets are often used by lawyers as part of their argumentation and also by judges for deciding on a case. In this paper, we propose a weakly supervised technique for extracting such statute facets from legal text. We use dependency tree structure to extract candidate statute facets and use BM25 ranking function to determine statute-specificity of these candidates. We propose a set of facet types which enable us to realize the definition of statute facets in a more computational way. We use recent deep learning models in a few-shot setting to predict an appropriate facet type for each candidate. Only those candidates with high statute-specificity and for which a facet type is predicted with high confidence, are selected as acceptable statute facets. We evaluate the extracted statute facets through both direct and indirect evaluation as well as conduct a user-study to get validation and feedback from lawyers.

• **Summary:** The paper focuses on automatically extracting key "statute facets" from legal texts, which can serve as useful elements in legal arguments. The authors define statute facets as specific, statute-related concepts or roles, such as offenses, obligations, or conditions, relevant to legal proceedings. The paper proposes a multi-step approach combining dependency parsing, BM25-based ranking to assess specificity, and a few-shot learning model to predict the type of each facet. By using deep learning models like BART and Legal-BERT in a few-shot setup, the authors show how minimal training data can be leveraged to accurately classify and identify critical facets within legal statutes. Evaluations include both direct assessments and user

studies with legal experts, with findings indicating a high acceptance rate of identified facets as meaningful components for legal argumentation.

- **Relevance:** The methodology described, particularly the few-shot learning approach combined with dependency parsing, could be directly adapted for date extraction in our project. Since our documents are rich with unstructured legal text, similar techniques could be applied to accurately identify dates relevant to case timelines. Additionally, the paper’s approach to evaluating the specificity and context of extracted data may help refine our extraction of structured data from sentencias, allowing us to identify case-specific conditions that contribute to judicial bottlenecks. The findings could significantly support the project’s broader objective of informing policy decisions to address court congestion.

2. **Title:** Argumentation Structure Prediction in CJEU Decisions on Fiscal State Aid

- **URL:** <https://dl.acm.org/doi/10.1145/3594536.3595174> (may require accessing with GaTech credentials to view actual paper)
- **Abstract:** Argument structure prediction aims to identify the relations between arguments or between parts of arguments. It is a crucial task in legal argument mining, where it could help identifying motivations behind judgments or even fallacies or inconsistencies. It is also a very challenging task, which is relatively underdeveloped compared to other argument mining tasks, owing to a number of reasons including a low availability of datasets and a high complexity of the reasoning involved. In this work, we address argumentative link prediction in decisions by Court of Justice of the European Union on fiscal state aid. We study how propositions are combined in higher-level structures and how the relations between propositions can be predicted by NLP models. To this end, we present a novel annotation scheme and use it to extend a dataset from literature with an additional annotation layer. We use our new dataset to run an empirical study, where we compare two architectures and explore different combinations of hyperparameters and training regimes. Our results indicate that an ensemble of residual networks yields the best results.

- **Summary:** The paper presents a novel approach for predicting argument structures in legal decisions from the Court of Justice of the European Union (CJEU) regarding fiscal state aid. The authors focus on identifying relationships between propositions—such as support, rebuttal, and undercut relations—in judicial documents. They introduce a unique annotation scheme and apply it to an existing dataset, adding layers that classify types of argumentative links between premises and conclusions. Using this enriched dataset, they conduct experiments to assess different NLP models’ performance in predicting argumentative links, ultimately finding that an ensemble of residual networks yields the best results. The study contributes to computational legal argumentation, a field that seeks to enhance access to legal reasoning by automating the extraction and understanding of complex argument structures.
- **Relevance:** This paper demonstrates methods for parsing and identifying complex relationships in legal text. The annotation approach for linking premises and conclusions in CJEU cases provides a structured method that could inform our project’s extraction framework, especially in identifying the sequence of dates and events relevant to judicial delays. The study’s focus on overcoming the challenges posed by complex legal arguments aligns with our need to parse the often intricate and implicit timelines in Dominican civil cases. Incorporating their strategies could enhance our ability to efficiently structure extracted data, aiding our broader goal of informing policy solutions to address court congestion.

3 SCRIPTS AND CODE BLOCKS

All scripts have been uploaded to the [HAAG NLP Repo](#). Outputs files, processed sentences and any other document that may contain sensitive information is located in the private [NLP-Sentencias Repo](#).

The following code contains the logic and functions I have been working on this week.

1. Created a function to combine all results from all dates into a single file [here](#).

```

def extract_context_around_date(text, date,
                                window_size=1000):
    # text_lower = text.lower()
    date_lower = date.strip()
    start = 0
    start_position = text.find(date_lower, start)
    if start_position == -1:
        print(f"Date '{date}' not found in the text.")
    else:
        start_index = max(start_position - window_size, 0)
        end_index = min(start_position + len(date_lower) +
                        window_size, len(text))
        context = text[start_index:end_index]
    return context

```

Code 1—Function to extract smaller date window of 2000 characters

```

def generate_output(ollama_models: list, query_template:
    ↪ str, input_folder: str, dates_folder: str,
    ↪ clusters_file: str, output_folder: str, repetitions:
    ↪ int = 1, delete_models_after_query: bool = False,
    ↪ model_hyperparameters: dict = {}):
# Instantiate the OllamaModelProcessor
for model in ollama_models:
    # Log the model being processed:
    log_in_color(f"Processing model: {model}", "green")
    # Step 1: Instantiate the OllamaModelProcessor
    processor = OllamaModelProcessor(model,
    ↪ **model_hyperparameters)

    # Step 2: Get the output options from the clusters.json
    ↪ file contained in the options key
    options_file = clusters_file
    with open(options_file, 'r', encoding='utf-8') as f:
        options_content = json.load(f)
    # Now set the options to be the value of the options key
    options = json.dumps(options_content["options"],
    ↪ ensure_ascii=False)

    for filename in os.listdir(input_folder):
        if filename.endswith(".txt"): # Process only txt files
            # Log the file being processed:
            log_in_color(f"Processing file: {filename}",
            ↪ "blue")
            # Remove the extension from the filename
            filename_name = os.path.splitext(filename)[0]
            # We append locate the output folder under a
            ↪ folder with the file name first and then a
            ↪ folder with the model name
            file_output_folder = os.path.join(output_folder,
            ↪ filename_name, model)
            # Create the output folder if it doesn't exist
            os.makedirs(file_output_folder, exist_ok=True)

```

```

# Read the content of the document
document_path = os.path.join(input_folder,
    ↪ filename)
with open(document_path, 'r', encoding='utf-8')
    ↪ as f:
    document_content = f.read()
# Now we query the model replacing the last place
    ↪ holder with each date independently
# Read the dates JSON file with same name as the
    ↪ document
dates_file = os.path.join(dates_folder,
    ↪ f"{filename_name}.json")

with open(dates_file, 'r', encoding='utf-8') as
    ↪ f:
    date_objects = json.load(f)

for i, date_object in enumerate(date_objects):
    date_for_context = date_object["date"]
    date_context = extract_context_around_date(do_
        ↪ cument_content,
        ↪ date_for_context)
    query_without_date =
        ↪ generate_query(query_template,
        ↪ date_context, options)

# Log the date position being processed:
log_in_color(f"Processing date: {i}",
    ↪ "magenta")
expected_output = json.dumps(date_object,
    ↪ ensure_ascii=False)
query = query_without_date.replace("{}MODEL_0_
    ↪ UTPUT_FORMAT}",
    ↪ expected_output)

```

```

# For each query, we generate <repetitions>
→ outputs to ensure output consistency
for repetition in range(repetitions):
    # Log the repetition being processed:
    log_in_color(f"Processing repetition:
→ {repetition + 1}", "yellow")
    output = processor.query_model(query)
    output_path =
→ os.path.join(file_output_folder,
→ f"{filename_name}_{i}_{repetition +
→ 1}.txt")
    with open(output_path, 'w',
→ encoding='utf-8') as f:
        # First write a line with the date
        → object passed to the model
        # f.write(expected_output + "\n\n")
        f.write(output)

# Log the model being deleted:
log_in_color(f"Deleting model: {model}", "red")
# Delete the ollama model to free up space
if delete_models_after_query:
    processor._delete_model()

```

Code 2—Updated function to use smaller date text window for retrieving the context

4 DOCUMENTATION

Similar to what was indicated in past reports, the pipeline/flow we're currently following is the one below, where we first extract and clean the documents. Afterwards, a process takes care of diving the clean documents into smaller pieces that can be then passed as input to a new layer where a [Bert based model](#) in Spanish, that has been fine tuned to better identify dates over legal documents for the Dominican Republic, is used to retrieve the dates from the corpus. Once these dates have been identified, they will be passed on to an additional model

that will then retrieve the context of the date to identify what it is representing. Finally, all dates will be grouped and included in one file, representing the output of all the pieces of the original document being put together.

The following diagram represents this flow:

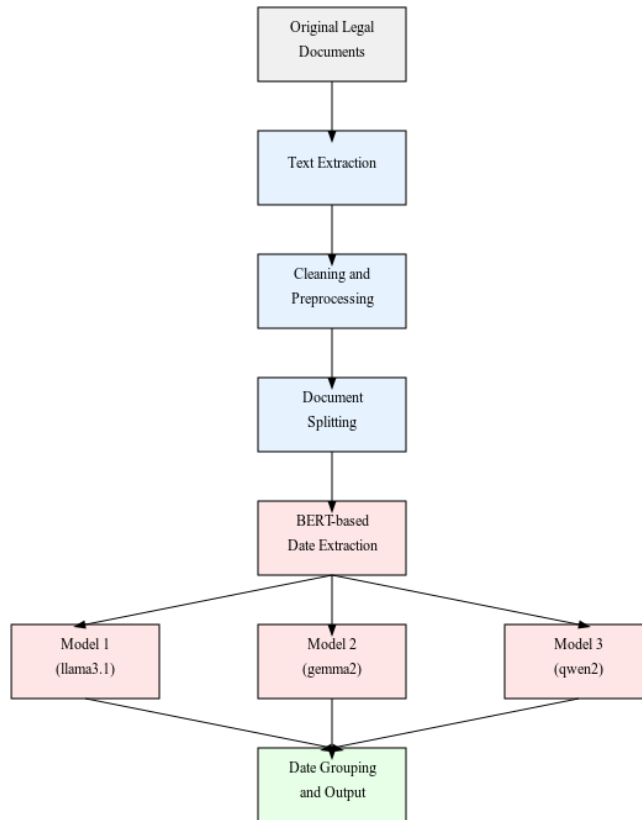


Figure 1—Full date extraction process

This week, my focus has been on the second to last step, with the goal of improving the models results given the low accuracy (0.3) obtained when performing the benchmark. The updated process for each has been the following:

Date context extraction

- Input template generated in txt format to feed the model and retrieve the date context. This template contains placeholders to fill in:
 - Date retrieved by model in previous steps, copied exactly from the original text document.

- 2000 characters window (1000 before and 1000 after the date) from the original text where the date is contained.
- Options/clusters template containing the categories by which to classify the different dates retrieved.

The output of the model will be a single text file containing a JSON object with the input date, a JSON object with the model output and a JSON object containing configuration details for the executed model such as hyperparameters used, model's name and execution time.

5 SCRIPT VALIDATION

The intention is to query the model over a set of 5 files generating 10 outputs for each of the dates contained in the files. Obtaining additionally, performance metrics from the execution. This will be carried out initially only with Llama 3.1 model and compare the results with the ones obtained within the benchmark. It wasn't possible to perform this new check this week given running the models in PACE was not possible. A ticket was already opened regarding this matter.

Execution would be carried out with the following hyperparameters:

- Temperature = 0.0000001,
- Top_k = 5,
- Top_p = 0.5
- Seed = 42

Here is a brief explanation of these hyperparameters:

- **Temperature:** A very low temperature (0.0000001) ensures that the outputs will be highly predictable. This is useful when we are looking for consistency and want results to be stable over time.
- **Top-k:** This limits the choices to only the top 5 probable words. This ensures that the model generates meaningful outputs without straying into highly unlikely predictions. It balances between randomness and relevance.
- **Top-p:** Combined with top-k, this gives fine control over the diversity of model output. A top_p value of 0.5 means the model will only consider words that make up 50% of the total probability distribution, ensuring more relevant results.
- **Seed:** Setting the seed makes the experiments reproducible, helpful for research

purposes. With the same inputs and hyperparameters, in theory, we should get the same outputs every time (but in practice this doesn't always happen).

6 RESULTS VISUALIZATION

The following images provide the results obtained during the previous week when performing the benchmark, and are the base for improvement, focusing initially on the Llama 3.1 model.

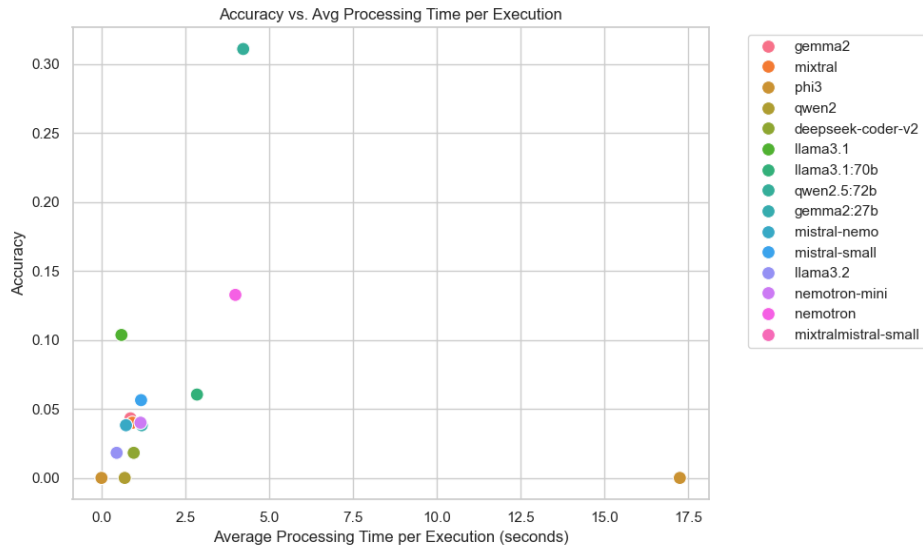


Figure 2—Accuracy vs. Processing time for each model

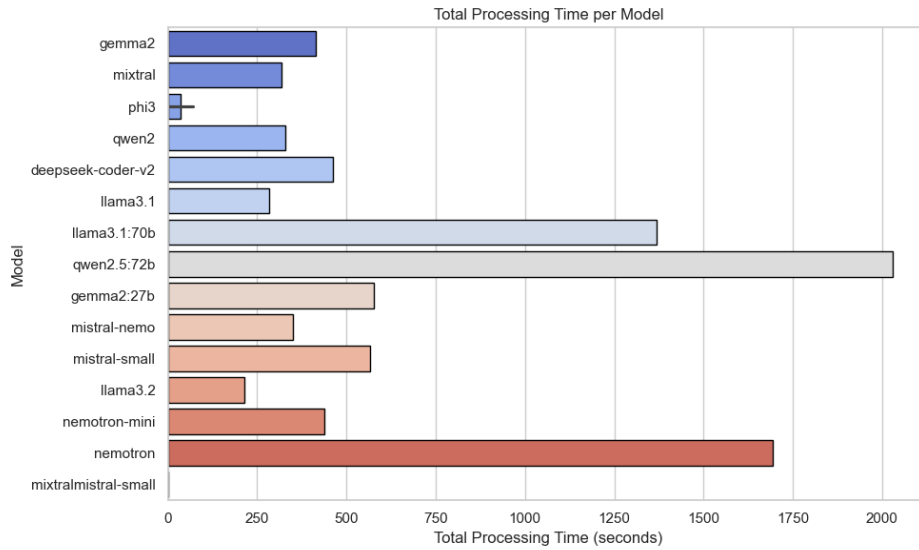


Figure 3—Total processing time for each model

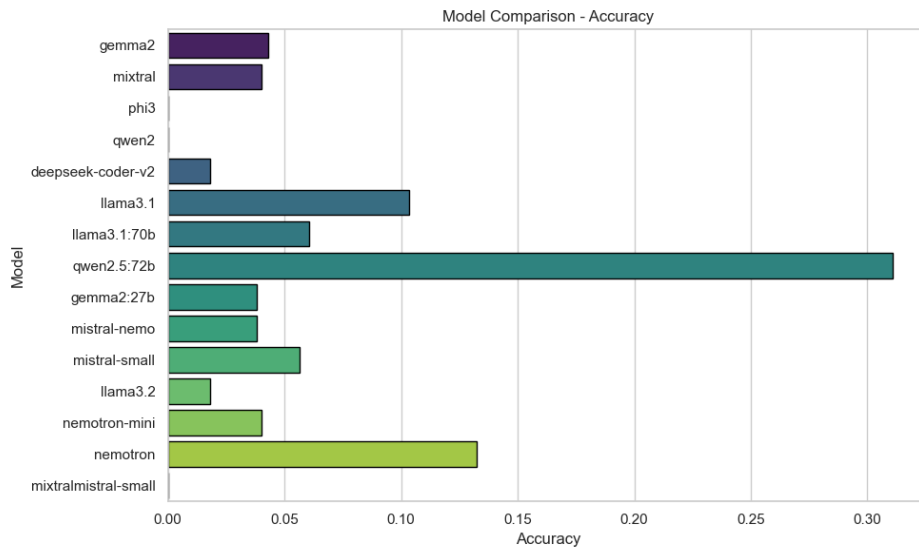


Figure 4—Accuracy comparison between models

7 PROOF OF WORK

The implemented function now uses a smaller piece of context to retrieve the class related to the date. This should both improve the model’s accuracy and performance.

Below is an example of a date and context retrieved by the updated code to be used for verifying the model's new accuracy. Similarly to what was performed on the previous week, to ensure stability in the results, each input will be used 10 times to generate an output, with the expectations that given the low temperature and the exact same input the model would generate the exact same output.

```
'seis (06) días del mes de enero del año dos mil  
veintitrés (2023)'
```

Code 3—Date to be classified

```
\nPRESIDENCIA DE LA CÁMARA CIVIL Y COMERCIAL DEL JUZGADO
→ DE PRIMERA INSTANCIA DEL DISTRITO NACIONAL \n
→ Ordenanza civil núm. 123-4567-ABCD-8901 Número único
→ de caso (NUC) 1234-0158080 EN NOMBRE DE LA REPÚBLICA
→ \n Ordenanza civil núm. 504-2023-SORD-0013 Número
→ único de caso (NUC) 1234-0158080 En la ciudad de Santo
→ Domingo de Guzmán, Distrito Nacional, capital de la
→ República Dominicana, a los seis (06) días del mes de
→ enero del año dos mil veintitrés (2023); años ciento
→ setenta y nueve (179) de la Independencia y ciento
→ sesenta (160) de la Restauración. \n \nPresidencia de
→ la Cámara Civil y Comercial del Juzgado de Primera
→ Instancia del Distrito Nacional, localizada en el
→ primer piso del Palacio de Justicia del Centro de los
→ Héroes de Constanza, Maimón y Estero Hondo, en el
→ Distrito Nacional, República Dominicana, presidida por
→ XXXXXXXXXXXX, quien dicta esta ordenanza en sus
→ atribuciones de juez presidente de los referimientos y
→ en audiencia pública constituida por la secretaria
→ XXXXXXXXX. \nXXXXXXXXX, y el alguacil de estrados de
→ turno. \n \nCon motivo de la demanda en referimiento
→ sobre producción forzosa y entrega inmediata de
→ certificado de matrícula interpuesta por la señora
→ XXXXXXXXX, dominicana, mayor de edad, titular de la
→ cédula de identidad y electoral núm. 123-45678-1, con
→ su domicilio en la calle XXXXXX, núm. 1, torre XXXXXX,
→ apartamento núm. 1, urbanización XXXX, XXXX
```

Code 4—Window for retrieving date class

8 NEXT WEEK'S PROPOSAL

1. Keep working on improving context retrieval for the dates.
2. Finish preparing slides for call with Judge Miguel.
3. Finalize implementation of a simple pipeline to run all processes end to end.

4. Keep working on a simple UI to interact with the pipeline and upload and process a single file.
5. Keep working on getting more content in for our paper.

Week 11 | HAAG - NLP | Fall 2024

Alejandro Gomez

November 1st, 2024

1 Time-log

1.1 What progress did you make in the last week?

- This week I started a large refactor to the NER model. Architecting the spaghetti-code into something that could be more modular took many iterations. I wanted to create a pipeline that was easy to tweak levers with, easy to read, and easy to debug. My current pipeline model is an assortment of dozens of files with even more functions and all living in the same directory. As it ballooned, it became impossible to track and debug which is a critical flaw since I'm currently chasing some odd bugs. This is in part due to preparing to wrap up the coding and transition to writing more, as well as finding bugs that have been causing odd behavior. Refactoring the code to be more decoupled, readable, maintainable can help also open a path to receive assistance as other devs and mentors can follow the workflow. The new pipeline will have a main driver file that will call different modules to perform the core business logic. Any properties that can be modified will live in a configuration file for rapid modifications on model iterations

1.2 What are you planning on working on next?

- I'll be working on the presentation to the DR Judge where we hope to receive feedback on our current efforts and be able to hone in on the outstanding problems in the DR legal system with document triaging. This can help us understand better what to focus on with the date recognition and context summarization.
- Need to continue gathering more data for training the model to improve it
- Continue the refactoring and scaffolding for our finetuning pipeline submission. This includes documentation to be used on HuggingFace
- Continuing: The labelling for the model on HuggingFace also shows "1,2,3" in the JSON file instead of "DATE" in the BIO format so I need to look into this next.
- Following the meeting with the judge, need to try to wrap up coding and focus on writing.

1.3 Is anything blocking you from getting work done?

N/A

2 Article Review

2.1 Abstract

The automated translation of natural language text into structured logical representations is a critical task in various applications, including legal reasoning and decision-making. This paper presents a Name Entity Recognition (NER) based approach for translating the legal case descriptions written in natural language into PROLEG fact formulas. The approach comprises (1) extracting legal entities from the case description using a specialized NER model, namely LegalCaseNER and (2) transforming

the extracted entities into PROLEG fact formulas using PROLEG rules. The experimental results demonstrate the efficacy of our proposed approach in accurately extracting relevant entities from legal case descriptions and translating them into the appropriate PROLEG fact formulas. Our approach provides a promising solution for handling complex and diverse case descriptions, enabling their representation in a structured format. This work provides a foundation for future research in the application of logical fact formulas in legal reasoning and decision-making. doi[ZNS+23]

2.2 Summary

This paper discusses an approach that fine-tunes a bert model using legal data to achieve 97% accuracy with NER. This is extremely relevant to our paper and it is a few page journal. This will be a good baseline for the team to review because of its similarity. This confirms what we aim to do is novel and gives me confidence in our future submission.

3 Scripts and Code Blocks

3.1 Code

```
1
2 # sample
3
4 data = {
5     "sentence": "Circunscripci n de Sato Domingo Oeste, la cual hace constar que
6     la joven Emely, es hija de los se ores ngel Gonz lez y Lorenza Mej a Ramos; \n
7     D. Declaraci n jurada de uni n libre, de fecha dieciocho (18) de mayo del a o
8     dos mil veintid s (2022), instrumentado por el doctor Miguel Cabral Hern ndez ,
9     notario p blico de los del n mero del Distrito Nacional, mediante el cual los
10    se ores ngel Gonz lez y Lorenza Mej a , declararon que se encontraban unidos
11    sentimentalmente en pareja, mediante la figura legal de esta civil de uni n libre,
12    desde el 28 de diciembre del a o 2006, hasta la fecha de dicha declaraci n , y
13    que fruto de esa relaci n han procreado dos \n(2) hijas que responden a los
14    nombres de Crismery Gonz lez Mej a y Emely Gonz lez Mej a ; \n E.",
15    "entities": [
16        {
17            "entity_group": "DATE",
18            "score": 0.9849426746368408,
19            "word": "dieciocho (18) de mayo del a o dos mil veintid s (2022)",
20            "start": 190,
21            "end": 244
22        },
23        {
24            "entity_group": "DATE",
25            "score": 0.9999849796295166,
26            "word": "28 de diciembre del a o 2006,",
27            "start": 547,
28            "end": 576
29        }
30    ]
31 }
```

Listing 1: result

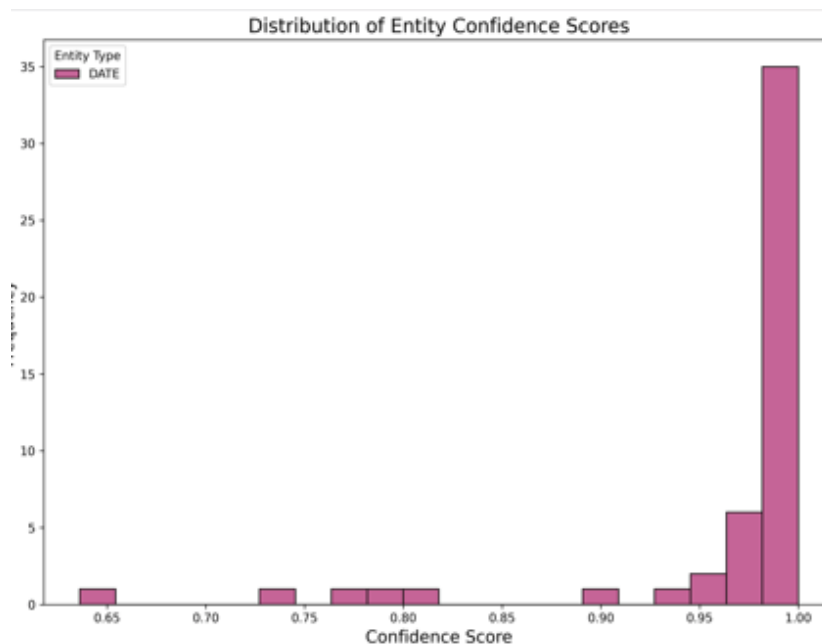


Figure 1: confidence score

As seen in the images above, the modifications made to the model for this week have not shown significant improvements. It seems like adding all the epochs made the model far more confident, however it is not always correct in its NER. I tried to allow the model to execute 11 epochs without exiting early (with the same amount of data as week 10) but there was no improvement. Note the missing parenthesis in the first NER. This is nearly identical results to week 10, so at least it's not deteriorating in performance.

```

1
2 import logging
3 from pathlib import Path
4
5 from huggingface_hub import HfApi
6
7 from modules.configurations.config import Config
8 from modules.utils import Utility
9 from modules import (
10     data_collection,
11     data_cleaning,
12     data_converting,
13     model_finetuning,
14 )
15
16 logging.basicConfig(
17     format="%(asctime)s - %(levelname)s - %(message)s", level=logging.INFO
18 )
19
20
21 class NERModelPipeline:
22     """
23     Pipeline for a finetuned Named Entity Recognition (NER) model for dates in Spanish
24     legal texts:
25     - preprocesses data
26     - finetunes a model
27     - deploys the model to HuggingFace (HF)
28
29     Using this script:
30     - finetune/create
31       - python main.py
32     - finetune/create AND deploy
33       - python main.py -d OR python main.py --deploy
34     """

```

```

34
35 def __init__(self):
36     self.deployment_config: dict[str, any] = Config.DEPLOY
37     self.general_config: dict[str, any] = Config.GENERAL
38
39 def run_full_ner_pipeline(self, current_dir: Path) -> None:
40     """runs the full finetuning pipeline"""
41     try:
42         logging.info("Starting the finetuning pipeline...")
43
44         data_collection.collect_raw_data()
45         data_cleaning.clean_raw_data()
46         data_converting.convert_cleaned_data()
47         model_finetuning.finetune_model(current_dir)
48
49         logging.info("Success! the finetuning pipeline is complete.")
50     except Exception as e:
51         logging.error(f"Error finetuning the model: {e}")
52         raise
53
54 def deploy_model_to_hf(self) -> None:
55     """
56     Deploy to the model to HuggingFace (HF).
57     Make sure you have set up your .env file with you HF environment variables.
58     """
59     try:
60         logging.info("Deploying model to HF...")
61         api = HfApi()
62         api.upload_folder(
63             folder_path=self.general_config["model_output_dir"],
64             repo_id=self.deployment_config["repo_id"],
65             token=self.deployment_config["token"],
66             repo_type="model",
67         )
68         url: str = f"https://huggingface.co/{self.deployment_config['repo_id']}"
69         logging.info(f"Success! See your deployed model at: {url}")
70     except Exception as e:
71         logging.error(f"Error deploying to HuggingFace: {e}")
72         raise
73
74
75 def main() -> None:
76     args = Utility.create_parser().parse_args()
77
78     ner_pipeline: NERModelPipeline = NERModelPipeline()
79     model_output_path: Path = Path(__file__).parent.resolve()
80     ner_pipeline.run_full_ner_pipeline(model_output_path)
81
82     if args.deploy:
83         ner_pipeline.deploy_model_to_hf()
84
85
86 if __name__ == "__main__":
87     main()

```

Listing 2: full pipeline

This refactor makes it clear how much more readable and maintainable the code will be rather than the cluster of dozens of functions. This main driver will pull from the modules for business logic and for configurations for any parameters that can and should be modified for easier fine-tuning modifications.

```

1
2 ag2004@t14s:~/gt/haag/sentencias/nlp/finetuning$ tree
3 .
4 |-- README.md
5 |-- _manual_model_tests
6 |   |-- deployed_model_test.py
7 |   |-- local_model_test.py
8 |-- data
9 |   |-- keep.txt
10 |-- main.py
11 |-- modules

```



```

12 | |-- configurations
13 | | |-- config.py
14 | |-- data_cleaning.py
15 | |-- data_collection.py
16 | |-- data_converting.py
17 | |-- model_finetuning.py
18 | |-- utils.py
19 |-- requirements.txt

```

Listing 3: code refactor tree

3.2 List of Scripts

- Full NER pipeline Scaffolding (wip)
 - Driver code to finetune and create model with an option to deploy to HuggingFace
 - The configurations needed for this pipeline
 - The modules that will have the core logic for the fine-tuning pipeline
 - scaffolding for testing the local model created and the deployed model

3.3 Documentation

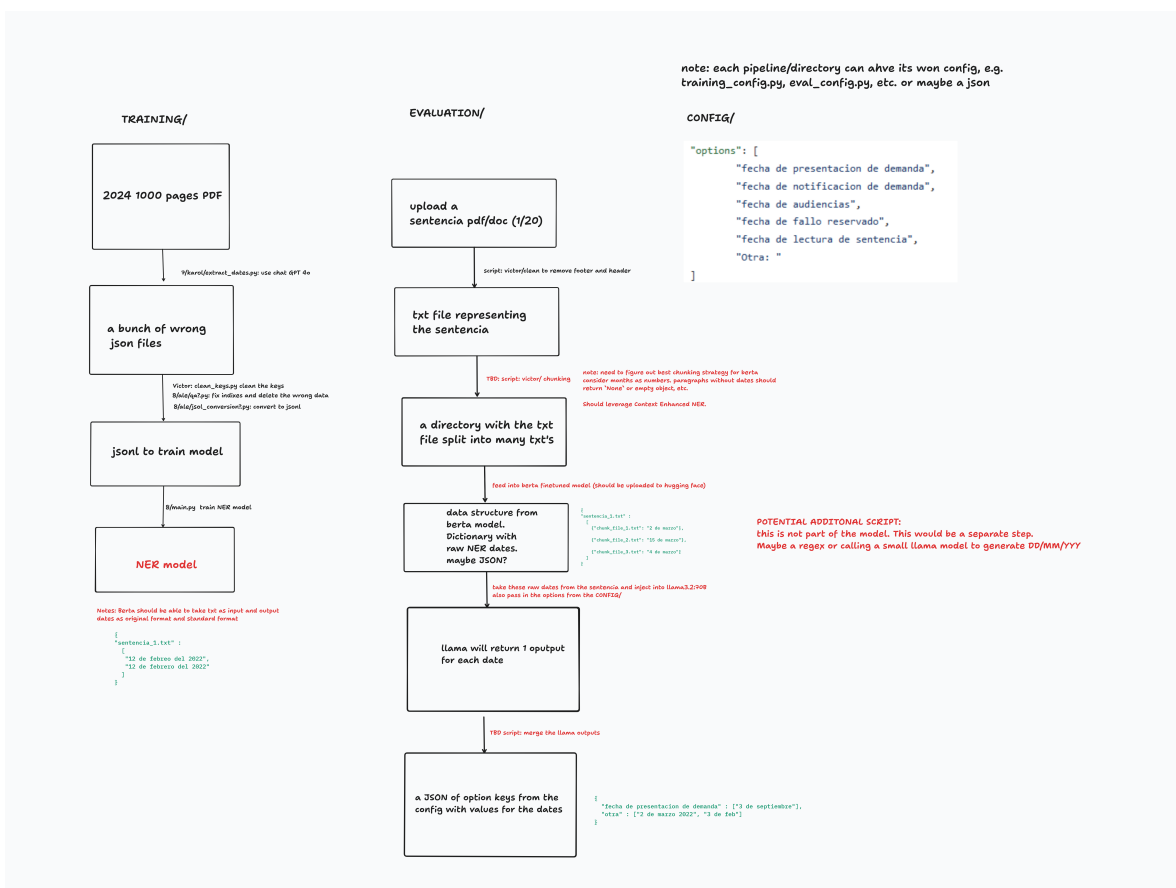


Figure 2: pipeline visualization for code structure

Given the current refactor this flow chart can be refined. Following the update from last week, the config was changed to python successfully and the high level steps are still the same. I can also check with the team to see if there have been iterations on the evaluation flowchart to update that formally as well.

3.4 Script Validation (optional)

N/A

3.5 Results Visualization

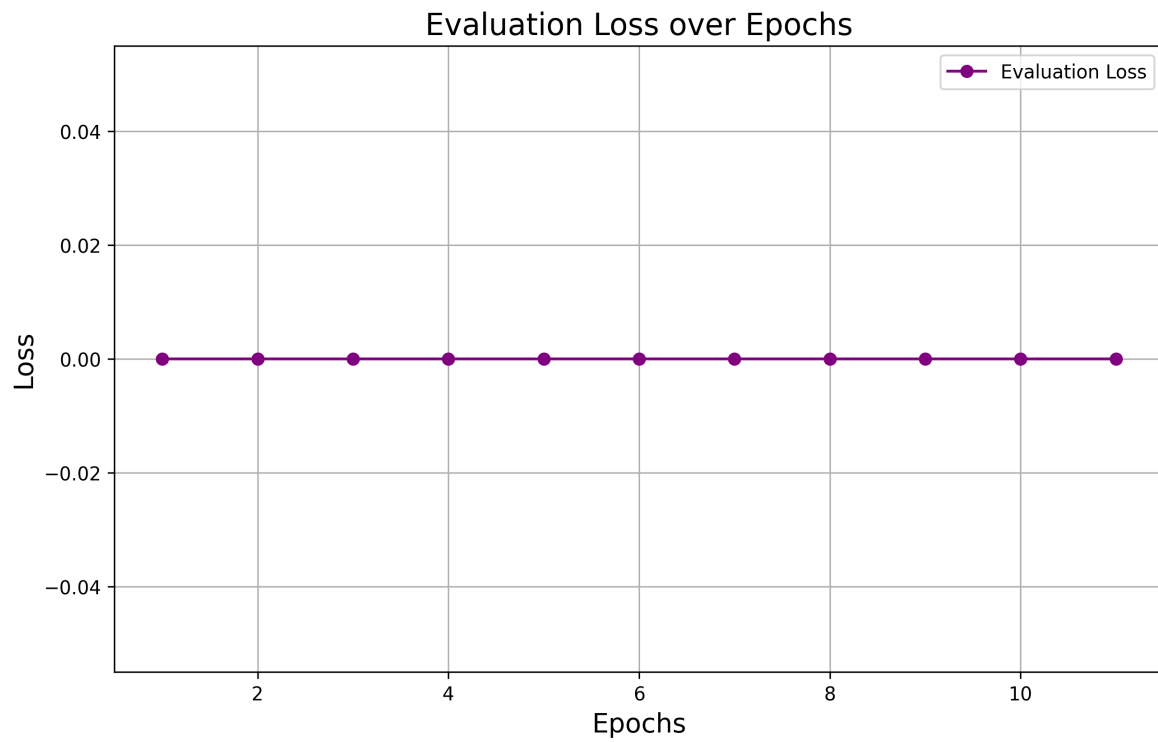


Figure 3: eval loss

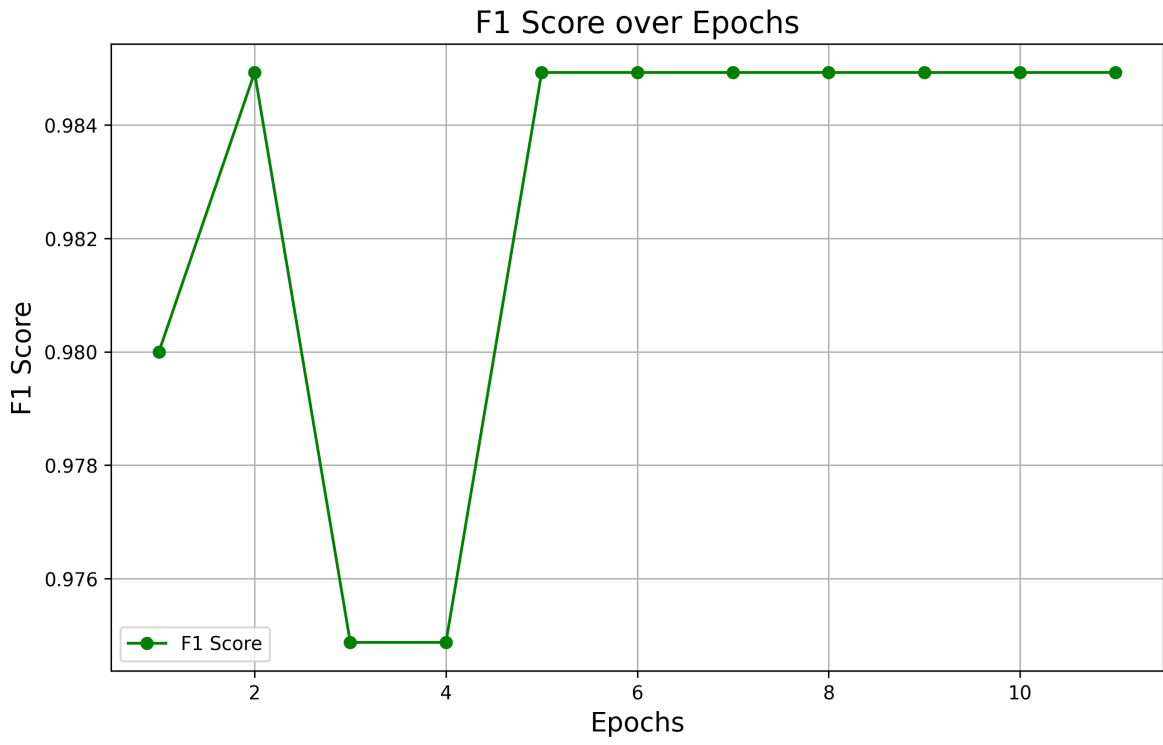


Figure 4: f1 score

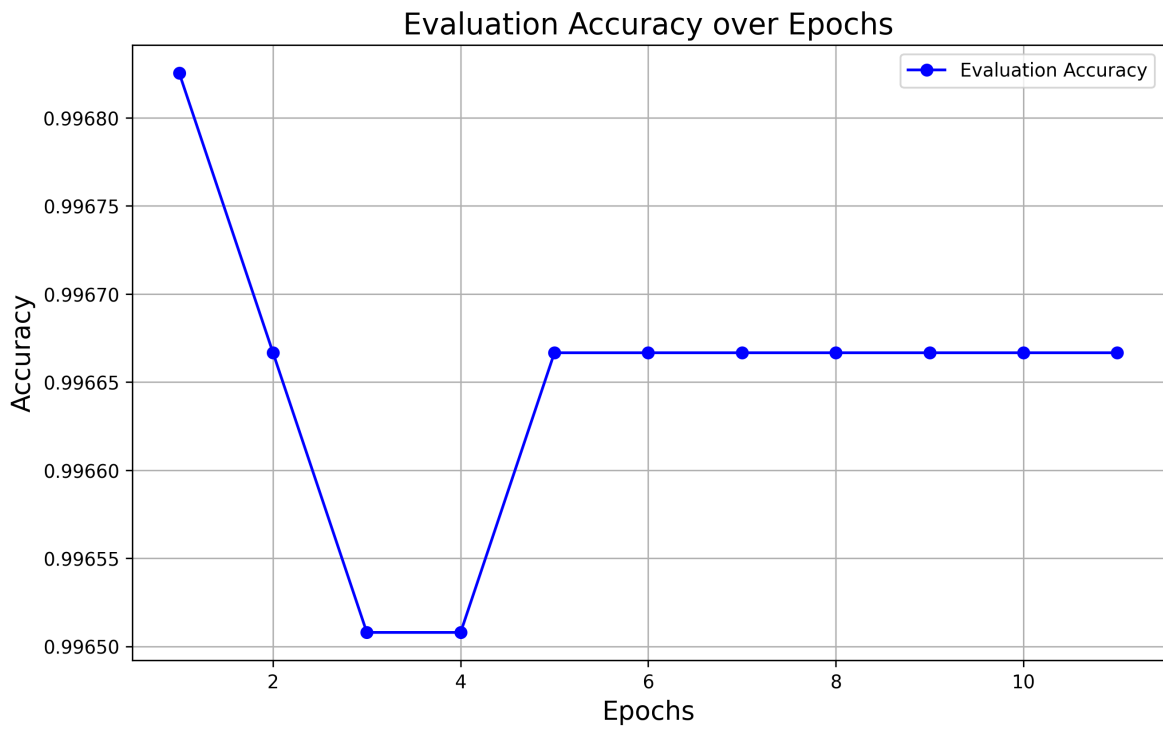


Figure 5: eval accuracy

3.6 Results Visualization Summary

All of the above graphs look highly suspicious. Their shapes are no consistent with expected behavior so this makes me believe that this could be due to running too many training epochs with the same amount of data. To fix these so they look as they did in the past couple weeks, I want to keep the epochs high or higher than initial and add more data.

3.7 Proof of Work

[Scripts in GitHub Repo](#)

4 Next Week's Proposal

- (See first section for full list. Brief summary below)
- Work on presentation for DR judge
- Improve model for POC during presentation and for publication
- Contribute to publication writing content
- Update current documentation, e.g. NLP website.

References

- [ZNS⁺23] May Myo Zin, Ha Thanh Nguyen, Ken Satoh, Saku Sugawara, and Fumihito Nishino. Improving translation of case descriptions into logical fact formulas using legalcasener. In *Proceedings of the Nineteenth International Conference on Artificial Intelligence and Law*, ICAIL '23, page 462–466, New York, NY, USA, 2023. Association for Computing Machinery.

HAAG NLP Sentencias — Week 11 Report

NLP-Gen Team

Karol Gutierrez

November 2, 2024

1 Weekly Project Update

1.1 What progress did you make in the last week?

- Run model with new training data.
- Research on clustering sample for the context of sentencia.
- Add K means clustering for context.
- Fulfill my role as Meet Manager/Documentor by working on the tasks expected for my position.
- Continuous meetings with Dr. Alexander, Nathan and team to discuss progress on project and publication options, as well as internal meetings with team to sync on next steps.

1.2 What are you planning on working on next?

- Improve clustering solution to provide more meaningful categories.
- Sync with team to get feedback on the categories to use in our analysis as well as current performance.
- Continue fulfilling my role as Meet Manager/Documentor by working on the tasks expected for my position (gather notes from meetings and prepare recordings).

1.3 Is anything blocking you from getting work done?

No.

2 Literature Review

Paper: LegalNLP – Natural Language Processing methods for the Brazilian Legal Language [[PMP+21](#)].

2.1 Abstract

We present and make available pre-trained language models (Phraser, Word2Vec, Doc2Vec, FastText, and BERT) for the Brazilian legal language, a Python package with functions to facilitate their use, and a set of demonstrations/tutorials containing some applications involving them. Given that our material is built upon legal texts coming from several Brazilian courts, this initiative is extremely helpful for the Brazilian legal field, which lacks other open and specific tools and language models. Our main objective is to catalyze the use of natural language processing tools for legal texts analysis by the Brazilian industry, government, and academia, providing the necessary tools and accessible material.

2.2 Summary

The paper introduces LegalNLP, a toolkit specifically designed for processing Brazilian legal texts. It offers various pre-trained language models, including Phraser, Word2Vec, Doc2Vec, FastText, and BERT, optimized for Portuguese legal language. The authors provide a Python package with functionalities and tutorials to facilitate the use of these models across legal applications, supporting efforts to advance NLP in Brazilian legal contexts.

- **Data Handling:** LegalNLP is capable of processing diverse legal documents in Portuguese, including court rulings, contracts, and statutes, helping users to extract valuable insights from unstructured legal text.
- **Entity Recognition:** The toolkit includes algorithms tailored to Brazilian legal language, improving the identification of relevant entities and legal terms compared to traditional NLP models not trained on Portuguese legal text.
- **Text Classification:** LegalNLP offers built-in support for classifying legal documents, which allows categorization of texts based on legal themes, aiding in organization and retrieval within Brazilian legal domains.
- **Practical Applications:** Designed for legal professionals, researchers, and developers, LegalNLP supports the creation of applications that leverage Brazilian legal data, contributing to the modernization of legal technology and enhancing access to legal information in Brazil.

2.3 Relevance

This paper is directly relevant to our Sentencias project with the difference that it is designed to work in Portuguese and it is fine tuned for the Brazilian judicial system. Nevertheless, it demonstrates the potential of specialized NLP tools for legal language, which aligns with our goal of analyzing Spanish language court decisions. By following a similar model and adapting techniques to the Spanish language, we can improve our approach and provide promising results.

3 Scripts and code blocks

The code is in the private repository [repository](#). The progress for this week is in `./karol/week11/`.

3.1 Code developed

The following items were developed this week. The full workflow of the code is shown in [Figure 1](#).

- Script to implement clustering using K-Means after processing semantic categories using bert tokenizer in [Figure 2](#)
- Plot results in [Fig 3](#).

4 Documentation

The documentation is present in the README.md file in the [repository](#).

For this week, the only added library is: `pip install langdetect`.

5 Script Validation

[Figure 4](#) shows the score obtained after using different number of clusters. This validation of the performance is still low to be meaningful in the classification, yet it is a proof of concept of the clusterization approaches that can be tried in the final part of the project.

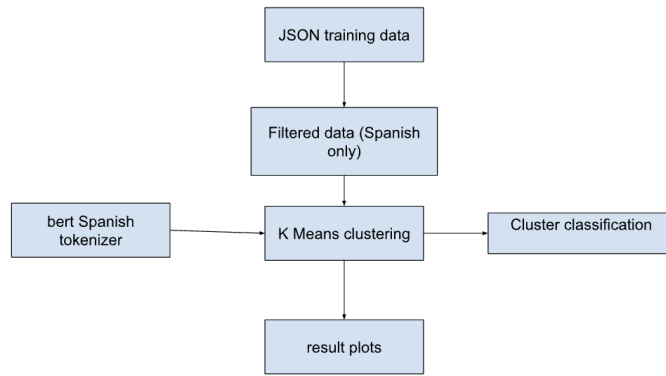


Figure 1: Code logic workflow to process data and train model.

6 Results Visualization

Final result of K-Means using 5 clusters for the different types of sentencias. This can be seen in Figure 5.

7 Proof of Work

Figures 6 and Figure 7 show the final distribution of the data into the 5 predefined clusters, as well as the classification during runtime of the program.

8 Next Week’s Proposal

Refer to section 1.2 for details (avoid repetition).

References

- [PMP⁺21] Felipe Maia Polo, Gabriel Caiaffa Floriano Mendonça, Kauê Capellato J. Parreira, Lucka Gianvechio, Peterson Cordeiro, Jonathan Batista Ferreira, Leticia Maria Paz de Lima, Antônio Carlos do Amaral Maia, and Renato Vicente. Legalnlp – natural language processing methods for the brazilian legal language, 2021.

```

clusterization.py x finetune.py
clusterization.py > get_bert_embedding
12
13 # Folder path with JSON files
14 folder_path = "../week10/dates_marzo"
15 output_path = "./plots1"
16
17 # Create the output directory if it doesn't exist
18 os.makedirs(output_path, exist_ok=True)
19
20 # Load the Spanish BERT model and tokenizer
21 tokenizer = AutoTokenizer.from_pretrained("dccuchile/bert-base-spanish-wwm-cased")
22 model = AutoModel.from_pretrained("dccuchile/bert-base-spanish-wwm-cased")
23
24 # Step 1: Load and filter Spanish contexts
25 contexts = []
26 print("Loading and filtering contexts...")
27 for filename in tqdm(os.listdir(folder_path), desc="Files processed"):
28     if filename.endswith(".json"):
29         with open(os.path.join(folder_path, filename), "r", encoding="utf-8") as file:
30             try:
31                 data = json.load(file)
32                 if isinstance(data, list):
33                     for entry in data:
34                         context = entry.get("context", "")
35                         if detect(context) == "es": # Ensure text is in Spanish
36                             contexts.append(context)
37             except json.JSONDecodeError:
38                 continue
39
40 # Step 2: Generate BERT embeddings for each context
41 def get_bert_embedding(text):
42     inputs = tokenizer(text, return_tensors="pt", truncation=True, padding=True)
43     with torch.no_grad():
44         outputs = model(**inputs)
45         # Use the mean of the last hidden state as the embedding
46         embedding = outputs.last_hidden_state.mean(dim=1).squeeze().numpy()
47     return embedding
48
49 print("Generating BERT embeddings...")
50 embeddings = np.array([get_bert_embedding(context) for context in tqdm(contexts, desc="Embedding contexts")])
51
52 # Step 3: Clustering using KMeans and calculating Silhouette Score
53 n_clusters = 5 # Number of clusters based on categories identified
54 print("Performing KMeans clustering...")
55 kmeans = KMeans(n_clusters=n_clusters, random_state=0)
56 kmeans.fit(embeddings)
57 labels = kmeans.labels_
58
59 # Calculate silhouette score for quality of clustering
60 silhouette_avg = silhouette_score(embeddings, labels)
61 print(f"Silhouette Score for {n_clusters} clusters: {silhouette_avg}")
62
63 # Optional: Calculate and plot silhouette scores for different cluster counts
64 print("Calculating silhouette scores for different numbers of clusters...")
65 silhouette_scores = []
66 cluster_range = range(2, 10)
67 for n in cluster_range:
68     km = KMeans(n_clusters=n, random_state=0)
69     km.fit(embeddings)
70     score = silhouette_score(embeddings, km.labels_)
71     silhouette_scores.append(score)
72
73 # Define representative phrases for each category
74 category_phrases = {
75     "Supreme Court Judgments": "sentencia de la Suprema Corte de Justicia",
76     "Appeal Court Decisions": "decisión de la Corte de Apelación",
77     "Initial Trial Judgments": "juzgado de primera instancia dicta sentencia",
78     "Filing and Appeal Actions": "recurso de apelación interpuesto",
79     "Scheduled Dates and Readings": "fecha de lectura de sentencia"
80 }
81
82 # Function to calculate embedding
83 def get_category_embedding(text):
84     inputs = tokenizer(text, return_tensors="pt", truncation=True, padding=True)
85     with torch.no_grad():
86         embeddings = model(**inputs).last_hidden_state.mean(dim=1)
87     return embeddings
88

```

Figure 2: Clustering of categories for context


```

89 # Precompute category embeddings
90 print("Calculating category embeddings...")
91 category_embeddings = {category: get_category_embedding(phrase) for category, phrase in category_phrases.items()}
92
93 # Step 4: Categorize each context based on similarity to category phrases
94 def categorize_cluster(context):
95     context_embedding = torch.tensor(get_bert_embedding(context)).unsqueeze(0)
96     similarities = {category: torch.cosine_similarity(context_embedding, category_embedding).item()
97                   for category, category_embedding in category_embeddings.items()}
98     best_category = max(similarities, key=similarities.get)
99     return best_category
100
101 # Apply categories to clusters and count occurrences
102 print("Categorizing contexts...")
103 clustered_contexts = {i: [] for i in range(n_clusters)}
104 category_counts = {category: 0 for category in category_phrases.keys()}
105
106 for i, context in tqdm(enumerate(contexts), desc="Contexts categorized", total=len(contexts)):
107     label = labels[i]
108     category = categorize_cluster(context)
109     clustered_contexts[label].append({"context": context, "category": category})
110     category_counts[category] += 1
111
112 # Step 6: Plotting
113
114 # 6a. Scatter Plot with PCA Reduction
115 print("Generating PCA plot...")
116 pca = PCA(n_components=2)
117 reduced_X = pca.fit_transform(embeddings)
118
119 plt.figure(figsize=(10, 6))
120 scatter = plt.scatter(reduced_X[:, 0], reduced_X[:, 1], c=labels, cmap="viridis", alpha=0.7)
121 plt.colorbar(scatter, label="Cluster ID")
122 plt.title("PCA of Contexts with KMeans Clusters")
123 plt.xlabel("PCA Component 1")
124 plt.ylabel("PCA Component 2")
125 plt.savefig(os.path.join(output_path, "pca_clusters.png"))
126 plt.close()
127
128 # 6b. Bar Plot of Categories
129 print("Generating category bar plot...")
130 plt.figure(figsize=(10, 6))
131 categories, counts = zip(*category_counts.items())
132 plt.bar(categories, counts, alpha=0.7)
133 plt.title("Number of Contexts in Each Category")
134 plt.xlabel("Category")
135 plt.ylabel("Count")
136 plt.xticks(rotation=45)
137 plt.tight_layout()
138 plt.savefig(os.path.join(output_path, "category_counts.png"))
139 plt.close()
140
141 # 6c. Silhouette Score Plot for Different Cluster Counts
142 print("Generating silhouette score plot...")
143 plt.figure(figsize=(10, 6))
144 plt.plot(cluster_range, silhouette_scores, marker='o')
145 plt.title("Silhouette Score for Different Cluster Counts")
146 plt.xlabel("Number of Clusters")
147 plt.ylabel("Silhouette Score")
148 plt.xticks(cluster_range)
149 plt.grid()
150 plt.savefig(os.path.join(output_path, "silhouette_scores.png"))
151 plt.close()
152
153 # Display results by cluster and category
154 for cluster_id, cluster_contexts in clustered_contexts.items():
155     print(f"\nCluster {cluster_id}:")
156     for entry in cluster_contexts:
157         print(f"Category: {entry['category']} - Context: {entry['context']}")
158

```

Figure 3: Plotting of results

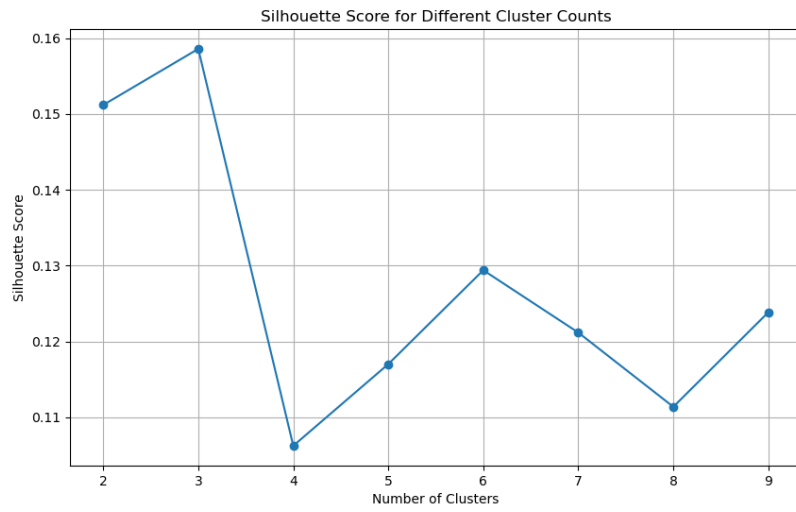


Figure 4: Proof of validation of results giving Silhouette scores for number of clusters

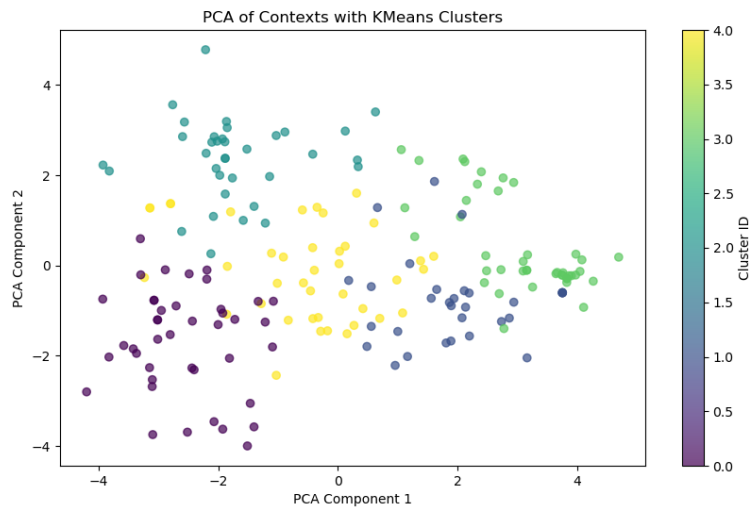


Figure 5: Generated clusters and visualization

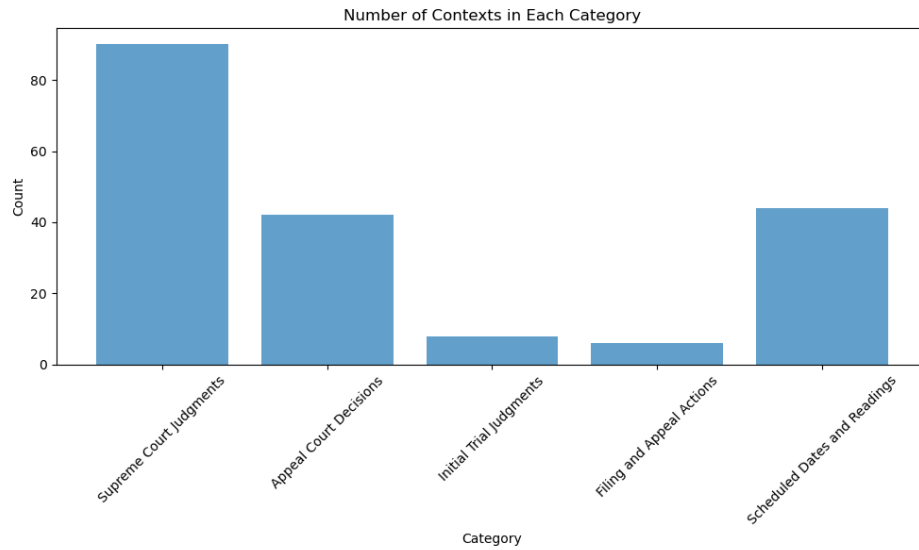


Figure 6: Distribution of categories

```

karo@Karols-MacBook-Pro:~/haag/sentencias/karo/week1
(haag-rlp) * week11 git:(main) * python clusterization.py
Some weights of BertModel were not initialized from the model checkpoint at dccuchile/bert-base-spanish-wn-cased and are newly initialized: ['bert.pooler.dense.bias', 'bert.pooler.dense.weight']
You should probably TRAIN this model on a down-stream task to be able to use it for predictions and inference.
Loading and filtering contexts...
Files processed: 100%
Generating BERT embeddings... | 17/17 [00:00:00:00, 21.96it/s]
Embedding contexts: 100% | 190/190 [00:55:00:00, 3.45it/s]
Performing kMeans clustering...
huggingface/tokenizers: The current process just got forked, after parallelism has already been used. Disabling parallelism to avoid deadlocks...
To disable this warning, you can either:
- Avoid using 'tokenizers' before the fork if possible
- Explicitly set the environment variable TOKENIZERS_PARALLELISM=(true | false)
Silhouette Score for 5 clusters: 0.11702653765678406
Calculating silhouette scores for different numbers of clusters...
Calculating category embeddings...
Categorizing contexts...
Contexts categorized: 100% | 190/190 [00:57:00:00, 3.30it/s]
Generating PCA plot...
Generating category bar plot...
Generating silhouette score plot...

Cluster 0:
Category: Initial Trial Judgments - Context: Primera Sala ordena la fusión de expedientes relativos a la casación.
Category: Appeal Court Decisions - Context: Dictamen de la procuradora general adjunta, Ana María Burgos, sobre recurso de casación.
Category: Scheduled Dates and Readings - Context: Expediente remitido de la secretaria general a la secretaria de la Primera Sala.
Category: Scheduled Dates and Readings - Context: Primera Sala establece la necesidad de motivar las decisiones en relación a la fijación de indemnizaciones.
Category: Filing and Appeal Actions - Context: Parte recurrente invoca sus medios contra la sentencia impugnada mediante memorial de casación.
Category: Filing and Appeal Actions - Context: El recurrido invoca sus medios de defensa mediante el memorial de defensa.
Category: Appeal Court Decisions - Context: Este expediente fue remitido de la Secretaría General a la Secretaría de esta Sala.
Category: Scheduled Dates and Readings - Context: Primera Sala prescinde de la necesidad de celebración de audiencia y del dictamen del Ministerio Público.
Category: Scheduled Dates and Readings - Context: Juicio celebrado donde se declara la inconstitucionalidad y la sentencia dictada al efecto.
Category: Scheduled Dates and Readings - Context: Este expediente fue remitido de la secretaria general a la secretaria de esta sala.
Category: Scheduled Dates and Readings - Context: Notificación de la parte recurrida en la oficina de su abogado apoderado.
Category: Scheduled Dates and Readings - Context: Notificación del memorial de casación fuera del plazo perentorio establecido en la ley.
Category: Scheduled Dates and Readings - Context: Notificación del memorial de casación fuera del plazo establecido.
Category: Scheduled Dates and Readings - Context: Memorial de casación contra la ordenanza impugnada.
Category: Scheduled Dates and Readings - Context: Memorial de defensa de la parte correcurrida.
Category: Scheduled Dates and Readings - Context: Expediente remitido de la Secretaría General a la Secretaría de Esta Sala.
Category: Scheduled Dates and Readings - Context: La fecha en la que se elabora la referencia de la designación del seccionario judicial.
Category: Initial Trial Judgments - Context: se deposita memorial de casación invocando medios contra la sentencia recurrida.
Category: Scheduled Dates and Readings - Context: Depositación memorial de defensa exponiendo medios de defensa.
Category: Scheduled Dates and Readings - Context: Expediente remitido de la secretaria general a la secretaria de esta sala.
Category: Scheduled Dates and Readings - Context: Plazo para realizar la notificación debido a la distancia.
Category: Filing and Appeal Actions - Context: la parte recurrente deposita el memorial de casación.
Category: Initial Trial Judgments - Context: La parte recurrida deposita el memorial de defensa.
Category: Scheduled Dates and Readings - Context: El expediente se remite de la secretaria general a la secretaria de la sala.

```

Figure 7: Code working and classifying JSON elements

Week 11 Research Report

Thomas Orth (NLP Summarization / NLP Gen Team)

October 2024

0.1 What did you work on this week?

1. Wrote script for basic settlement task
2. Experimented on Settlement Documents with Anthropic
3. Coordinated with Interview team for further summary reviews
4. Experimented with more of Anthropic workbench to test outputs

0.2 What are you planning on working on next?

1. Continue testing settlements to see how to extract the correct information for summarization
2. Meet with Clearinghouse team to discuss how we would demo / integrate more of the summarization work in an application

0.3 Is anything blocking you from getting work done?

1. For further testing, I'll need the settlement dataset that our OCR team is preparing.

1 Abstracts

- Title: Word Matters: What Influences Domain Adaptation in Summarization?. Conference / Venue: Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Link: <https://aclanthology.org/2024.acl-long.51/>
- Domain adaptation aims to enable Large Language Models (LLMs) to generalize domain datasets unseen effectively during the training phase. However, factors such as the size of the model parameters and the scale of training data are general influencers and do not reflect the nuances of domain adaptation performance. This paper investigates the fine-grained factors affecting domain adaptation performance, analyzing the specific

impact of ‘words’ in training data on summarization tasks. We propose quantifying dataset learning difficulty as the learning difficulty of generative summarization, which is determined by two indicators: word-based compression rate and abstraction level. Our experiments conclude that, when considering dataset learning difficulty, the cross-domain overlap and the performance gain in summarization tasks exhibit an approximate linear relationship, which is not directly related to the number of words. Based on this finding, predicting a model’s performance on unknown domain datasets is possible without undergoing training. Source code and scripts are available at <https://github.com/li-aolong/Word-Matters>.

- **Summary:** This paper explores domain adaption and its impact on the summarization process. They measure the performance gain from domain adaption through metrics such as the increase in ROUGE score. Based on the experiments provided, cross-domain overlap plays a key part in the performance for summarization tasks. They assert that by figuring out the cross-domain overlap between a source and target domain, that you can predict how well the model will do at summarizing that dataset.
- **Relevance:** This could prove useful to follow as we grow our summarization dataset.

2 Relevant Info

- Summary Chain of Thought (CoT) is a technique to prompt LLMs for information to provide context for summarization. I took a domain-centric approach in this experiment to extract entities the Clearinghouse is looking for specifically.
- Llama 3.2 is a popular LLM given its performance
- Ollama is a way to serve LLMs locally
- Langchain is a popular library for interacting with LLMs
- Anthropic is a company that produces the Claude family of models that compete with GPT-4.
- The two best models in terms of accuracy and cost tradeoff is Claude 3.5 Sonnet and Claude 3 Haiku

3 Scripts

1. All scripts uploaded to <https://github.com/Human-Augment-Analytics/NLP-Gen>
2. Scripts were run with the following file for testing: <https://gatech.box.com/s/63c0mulnwyzrauydaherfaihzrgtg32>

3. Thomas-Orth/anthropic/settlements/domain_specific_scot_chunked.py

- Brief Description: Run a domain specific version of Summary Chain-of-thought (CoT) on settlements with Anthropic models.
- Status: Tested by running the pipeline to completion without issue
- Important Code Blocks:
 - (a) First block: Read in CSV file, choose document
 - (b) Second block: Run through prompts, chunking documents, save summaries
 - (c) Third Block: Evaluate via manual inspection
- Screenshot of code: No screenshots provided due to the code being largely the same as last week, just with different prompts. Prompts will be pasted at the bottom of the report

4. Thomas-Orth/anthropic/settlements/make_temp_dataset.py

- Brief Description: Take a set of PDFs and OCR them
- Status: Tested by running the pipeline to completion without issue
- Important Code Blocks:
 - (a) One for loop to read through the PDFs and OCR them.
- Screenshot of code:

```
1 from unstructured.partition.pdf import partition_pdf
2 ids = [152022, 151557, 138154]
3 cases = ["Adams v Kentucky", "Progeny v. City of Wichita", "Terrill v. Oregon", "Remick v. City of Phi
4
5 def extract_text_from_pdf(pdf_path):
6     elements = partition_pdf(pdf_path)
7     parsed_text = "\n".join([element.text for element in elements])
8     return parsed_text
9
10 data = []
11 for id, case in zip(ids, cases):
12     path = "pdfs/{id}.pdf"
13     text = extract_text_from_pdf(path)
14     data.append((case, id, text))
15
16 import pandas as pd
17
18 df = pd.DataFrame(data, columns=["Case Name", "Doc ID", "Text"])
19 df.to_csv("settlements_tmp.csv", index=False)
```

Figure 1: OCR code

5. Flow Diagram:

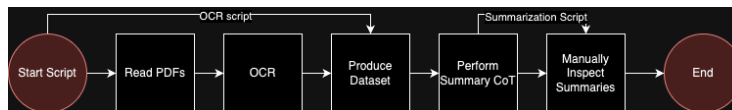


Figure 2: Flow diagram

6. Running scripts:

- (a) Download the scripts, the csv from the box link and llm.requirements.txt
- (b) Install unstructured from the README: <https://github.com/Unstructured-IO/unstructured/blob/main/README.md>
- (c) Run: `python -m pip install -r llm.requirements.txt`
- (d) Sign up for an Anthropic account, generate an API Key, and set "ANTHROPIC_API_KEY" in your environment.
- (e) Run: `python (chosen python script)`

4 Documentation

1. Download CSV file and PDFS
2. Update scripts to point to CSV file PDFS
3. Run scripts to output generated summaries
4. Manually evaluate summary

5 Results

5.1 Example Settlement Summary

Below is the example of a settlement summary from Claude Haiku:

"The Settlement Agreement is between the Plaintiffs, Oscar Adams and Michael Knights, and the Defendants, the Commonwealth of Kentucky et al. The settlement was filed on June 24, 2015 and aims to resolve claims related to the accessibility of services and accommodations for deaf inmates in the Kentucky Department of Corrections (KDOC).

Key Terms of the Settlement:

- The KDOC will ensure that deaf inmates have full and equal access to all services, privileges, facilities, and accommodations available to non-deaf inmates.
- The KDOC will provide deaf inmates with appropriate auxiliary aids and services, including qualified interpreters, to ensure effective communication.
- The KDOC will provide qualified interpreters in specific situations, such as medical care, disciplinary hearings, classification and transfer hearings, and educational/rehabilitative programs.
- The Effective Date of the Agreement is 30 days after it is filed with the Court.
- The KDOC must assign ADA Coordinators at each KDOC Adult Institution within 30 days of the Effective Date.
- The KDOC must provide deaf inmates with materials on auxiliary aids and services within 90 days of the Effective Date.
- The KDOC will pay \$250,000.00 to the Plaintiffs' counsel within 90 days of the Effective Date for attorneys' fees and costs.

- The KDOC will pay \$1,500.00 in damages to Plaintiff Oscar Adams and \$1,500.00 in damages plus \$77.62 to compensate for lien charges to Plaintiff Michael Knights, all within 90 days of the Effective Date.

The settlement includes detailed provisions to ensure that the KDOC provides appropriate accommodations and services for deaf inmates, as required by the U.S. Constitution, the Americans with Disabilities Act, and other applicable laws."

5.2 Prompts

Below are the prompts used by the anthropic model. First prompt will extract key details. The second will take that information to make a summary.

First prompt:

You are a law student tasked with extracting key information from a chunk of text belonging to a settlement document. Your goal is to identify and extract important details, provide citations for each extracted piece of information, and then create a partial summary based on the extracted information.

Here is the chunk of text from the settlement document:

<settlement_text> document </settlement_text>

Your task is to:

1. Extract key information from the text. This may include, but is not limited to:
 - Parties involved
 - Filing date
 - Monetary amounts
 - Key terms or conditions
 - Deadlines or important dates
 - Any specific legal language or clauses

2. For each piece of extracted information, provide a citation. The citation should be the exact quote from the original text that supports the extracted information. Format your extracted information and citations as follows:

<extracted_info>

(Key Information): (Your extracted information)

(Citation): "Exact quote from the text"

</extracted_info>

3. After extracting the key information, provide a partial summary of this chunk of text using the extracted information. The summary should be concise but comprehensive, capturing the main points of the settlement chunk.

Present your findings in the following format:

<key_information>

(List your extracted information with citations here)

</key_information>

<summary>
(Your partial summary of the chunk based on the extracted information)
</summary>

Remember to be thorough in your extraction of key information, accurate in your citations, and concise yet comprehensive in your summary.

Second Prompt:

You are a law student tasked with creating a concise summary of a settlement document. You will be provided with chunks of extracted information from the settlement document. Your goal is to analyze this information and create a clear, concise summary of the settlement, including a bulleted list of its key terms.

Here are the chunks of information from the settlement document:

<settlement_chunks> chunks </settlement_chunks>

Please follow these steps to complete your task:

1. Carefully read and analyze all the provided chunks of information from the settlement document.

2. Identify the key components of the settlement, including but not limited to:

- Parties involved
- Filing date
- Key terms and conditions
- Financial arrangements (if any)
- Timelines or deadlines
- Any special provisions or clauses

3. Create a brief introductory paragraph summarizing the overall nature and purpose of the settlement. This should include the parties involved, the date of the settlement, and a high-level overview of what the settlement aims to resolve.

4. List the key terms of the settlement in a bulleted format. Each bullet point should be concise but informative, capturing the essential details of each term.

5. If there are any notable or unusual aspects of the settlement, briefly mention these in a concluding paragraph.

Present your summary in the following format:

<summary> (Introductory paragraph here)

Key Terms of the Settlement: • (Bullet point 1) • (Bullet point 2) • (Bullet point 3) (...continue with all relevant bullet points)

(Concluding paragraph here, if applicable) </summary>

Remember to use clear, professional language throughout your summary. Avoid legal jargon where possible, but retain any necessary legal terms that are crucial to understanding the settlement. Your goal is to create a summary that would be easily understood by someone with basic legal knowledge.

6 Proof of work

The prompts were generated using Anthropic Workbench and ran using their LLMs, so the results are relatively reliable.

6.1 Known Limitations

The criteria for information extraction was based off what the workbench inferred. I received the criteria from our interview / expert review team this week for what the Clearinghouse looks for in settlement summaries so I'll need to refine the prompts around that.

6.2 Replies to Higher Ed Feedback

- The criteria for manual evaluation of the summaries is based on a checklist and directions receive from law students who work for the clearinghouse. Based on that feedback, we work to craft prompts that guide the LLM to summarize. Criteria can be provided in the next report once I have results with the new prompts
- Performance is measured based of the interview team reviewing the summaries. They compare the generated summaries to ground truth summaries that law students wrote, compare the generated summaries to the criteria we are given from the Clearinghouse, and then compare the summaries to the original document to test if they are hallucinating.
- Haiku and Sonnet are the names of the LLMs produced by Anthropic. They don't have any bearing on performance.