UNIVERSITÄT BERN
Master of Science
Artificial Intelligence in Medicine

Natural Language Processing Project Presentation:

Group C: Medical transcription data

Choekyel Nyungmartsang, Marc Nanzer, Matthias Peterhans, Vincent Gaspoz

Bern, 16.12.2025

Motivation

- Project's aim:
  - Construct NLP algorithms that can extract **meaningful data** from **medical notes**

  - ✓ Named entity recognition:
    - For each token, assign a label summarizing its medical entity (i.e. : 50 mg [dose] , lorazepam [durg]) → **useful for medical queries**

  - ✓ Predict category of medical notes:
    - Predict from which medical domain a note comes from (surgery, discharge summary, orthopedics, …)

Table of Contents

- Task 1: Dataset exploration
  - Understand data structure and composition
  - Creating a pre-processing pipeline (tokenizing, normalizing, filtering)
- Task 2: Named entity recognition:
  - Apply spaCy's NER algorithm and a transformer to perform NER on medical notes
- Task 3: Transformers
  - Use encoder and decoder transformers to correctly predict the category of medical notes

Argilla data set

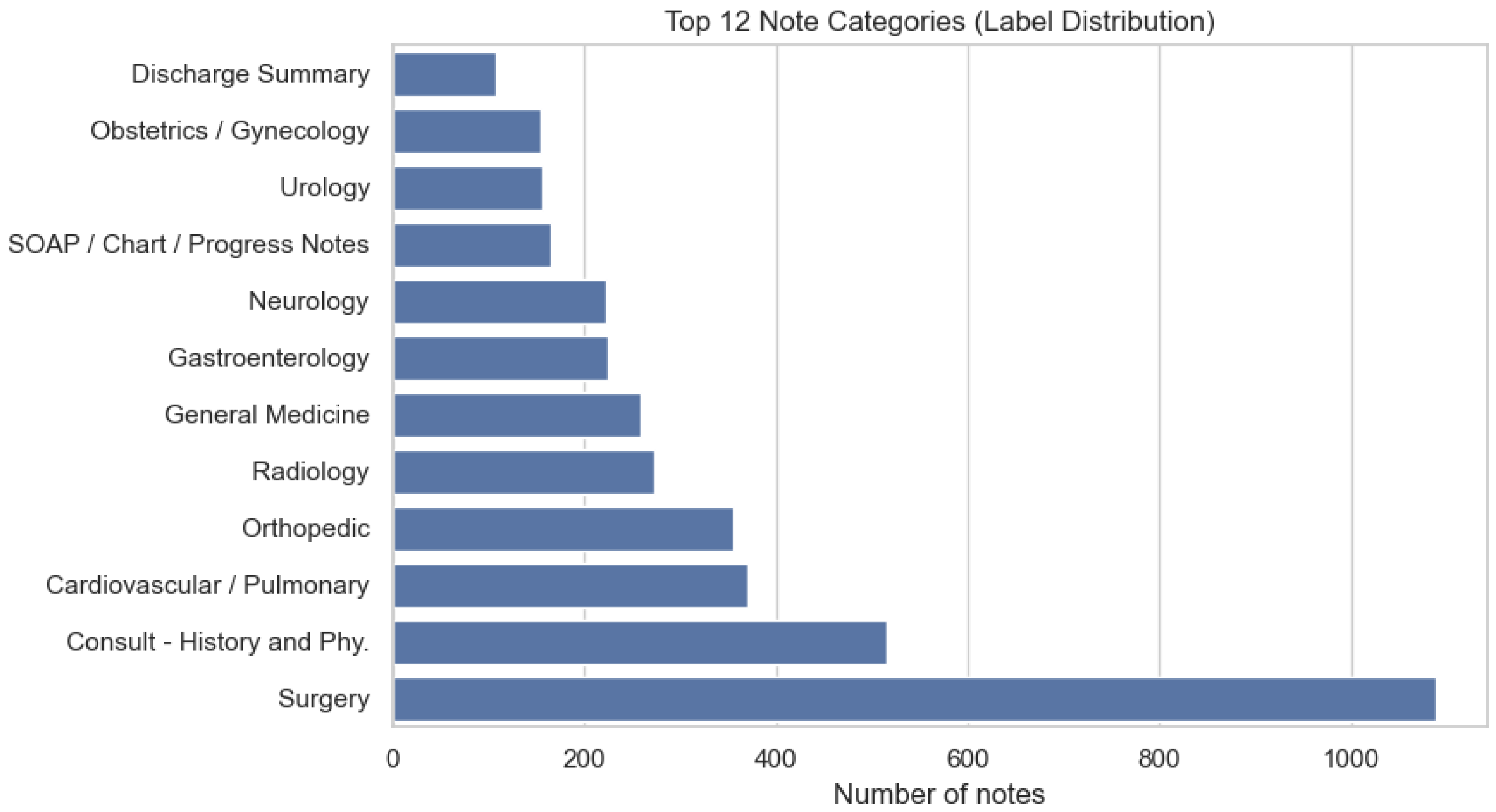
- **Unstructured** medical notes:
  - Source: https://mtsamples.com/
    - Transcribed medical reports
    - Reports provided directly by transcriptionist:
      - Over **40 domains**, but not **exhaustive** and **evenly distributed** across medical domains.

Task 1: Dataset Exploration

id	text	text_length	label
00001265-03e2-47b2-b6cf-bed	PREOPERATIVE DIAGNOSIS:, Iron deficiency anemia.,POSTOPERATIVE DIAGNOSIS:, Diverticulosis.,PROCEDURE:, Colonoscopy.,MEDICATIONS:	1085	Gastroenterology

- 4’966 notes
- 40 categories

Surgery dominates→



UNIVERSITÄT BERN
Master of Science
Artificial Intelligence in Medicine

Task 1: Duplicates

- 2’357 unique texts
- 2’148 duplicated clinical notes
- Small counts (max repeat = 5)
- Appear across different labels

Master of Science
Artificial Intelligence in Medicine

Task 1: Template-like Structure

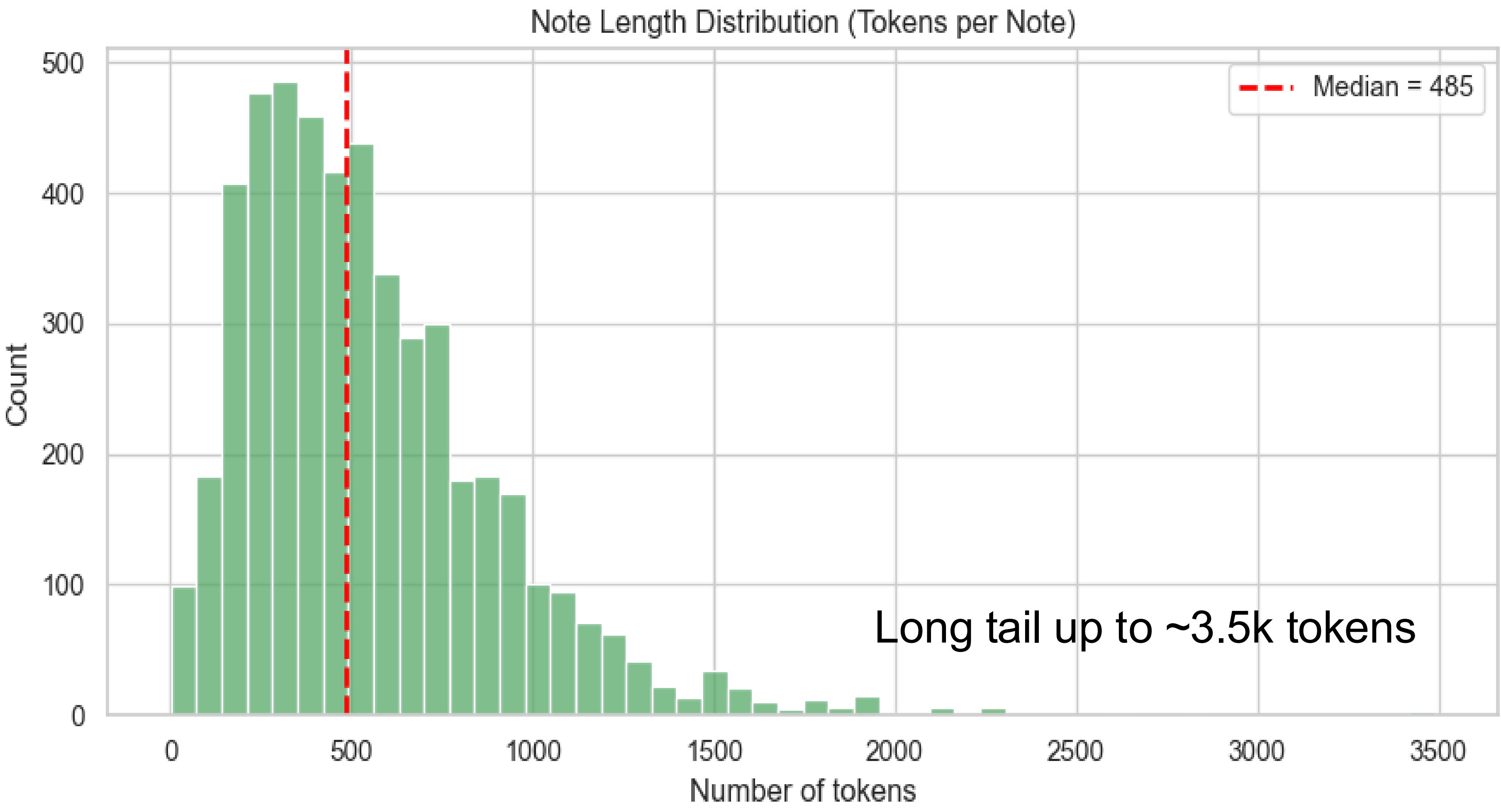
PREOPERATIVE DIAGNOSES:,1. Painful enlarged navicula, right foot.,2. Osteochondroma of right fifth metatarsal.,
POSTOPERATIVE DIAGNOSES:,1. Painful enlarged navicula, right foot.,2. Osteochondroma of right fifth metatarsal.,
PROCEDURE PERFORMED:,1. Partial tarsectomy navicula, right foot.,2. Partial metatarsectomy, right foot.,
HISTORY: ,This 41-year-old Caucasian female who presents to ABCD General Hospital with the above chief complaint. The patient states that she has extreme pain over the navicular bone with shoe gear as well as history of multiple osteochondromas of unknown origin. She states that she has been diagnosed with hereditary osteochondromas. She has had previous dissection of osteochondromas in the past and currently has not been diagnosed in her feet as well as spine and back. The patient desires surgical treatment at this time.,
PROCEDURE: ,An IV was instituted by the Department of Anesthesia in the preoperative holding area. The patient was transported to the operating room and placed on operating table in the supine position with a safety belt across her lap. Copious amounts of Webril were placed on the left ankle followed by a blood pressure cuff. After adequate se

Label: Orthopedic

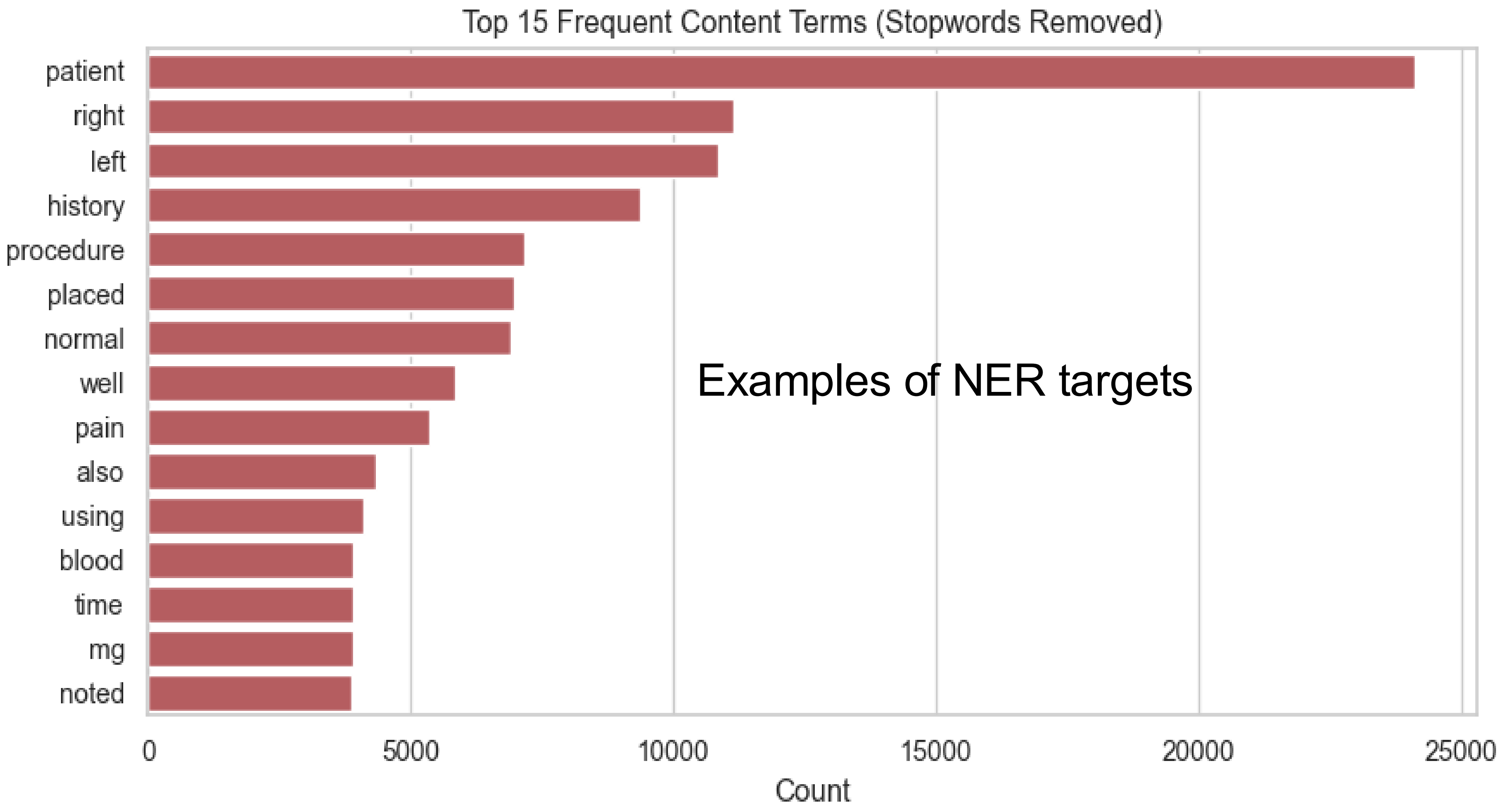
PREOPERATIVE DIAGNOSIS: , Bunion, left foot.,
POSTOPERATIVE DIAGNOSIS: ,Bunion, left foot.,
PROCEDURE PERFORMED:,1. Bunionectomy with first metatarsal osteotomy base wedge type with internal screw fixation.,2. Akin osteotomy with internal wire fixation of left foot.,
HISTORY: , This 19-year-old Caucasian female presents to ABCD General Hospital with the above chief complaint. The patient states she has had worsening bunion deformity for as long as she could not remember. She does have a history of Charcot-Marie tooth disease and desires surgical treatment at this time.,
PROCEDURE: , An IV was instituted by the Department of Anesthesia in the preoperative holding area. The patient was transported to the operating room and placed on operating table in the supine position with a safety belt across her lap. Copious amounts of Webril were placed on the left ankle followed by a blood pressure cuff. After adequate se

Label: Podiatry

Task 1: Text Properties (1/2)



Task 1: Text Properties (2/2)



Task 2: spaCy Baseline & Evaluation

- Investigate which standard and new NER types are most prominent in your dataset:
  - Standard NER: Person and Date
  - New NER: Disease, Anatomy and Medication

- Apply the standard NER classifier of sapCy to your data and evaluate it automatically and manually with 100 random sampled entities:
  - Automatical evaluation: no gold standard NER annotations

Task 2: spaCy Baseline & Evaluation

- Manual evaluation:
  - Correct: ~30%, Incorrect: ~70%
  - Does well: Dates: "05/26/1999", Person names: "Jackson"
  - Does poorly: Medications and Anatomy often labeled as ORG or Person

Entity	spaCy Label	Correct?	Expected Medical Category	Comment
Esophagogastroduodenoscopy	ORG	✗	PROCEDURE	Misread as organization
Melena	PERSON	✗	SYMPTOM / FINDING	Disease labeled as person
Demerol	ORG	✗	MEDICATION	Drug interpreted as an organization
Betadine	NORP	✗	MEDICATION / ANTISEPTIC	Not a nationality/group
CT Abdomen & Pelvis	ORG	✗	IMAGING PROCEDURE	Imaging test mislabeled
Foley	PERSON	✗	DEVICE / CATHETER	Mistaken as a person
L4–L5	ORG	✗	ANATOMY	Vertebral level mislabeled
3.4 cm	QUANTITY	✓	MEASUREMENT	Correct

Task 2: Custom Medical NER

Extend the standard NER and re-run the NER classification and evaluation:

- First attempt:
  - 5 Medical entities: ANATOMY, DISEASE, PROCEDURE, DRUG, DEVICE

```
 ('Both groins were prepped and draped in the usual sterile fashion.  After local anesthesi
a with 2% lidocaine, a 6-French sheath was inserted in the right femoral artery.\r',
  {'entities': [[5, 11, 'ANATOMY'],
    [73, 89, 'PROCEDURE'],
    [98, 107, 'DRUG'],
    [111, 126, 'DEVICE'],
    [147, 167, 'ANATOMY']]}),
```

  - Accuracy: ~30%
  - Problem: little training data, confusion with original NER types

Task 2: Custom Medical NER

Extend the standard NER and re-run the NER classification and evaluation:

- Second attempt:
  - 6 Medical entities: DISEASE, MEDICATION, SYMPTOM, PROCEDURE, ANATOMY, LAB_VALUE
  - Around 1000 labeled entities across all classes
  - Blank spaCy model and trained NER from scratch
  - Accuracy: ~43%
  - Remaining errors: confusion symptom vs diseases

Task 2: Transformer Exploration

- Goal: explore transformer models for medical NER
- Two approaches:
  - Encoder-based (BERT)
  - Decoder-based (zero-shot)
- Same clinical dataset, manual evaluation

Encoder-only Transformer (BERT)

- Model: biomedical BERT (token classification)
- Evaluated on 100 manually labeled entities

Observation

- **Overall accuracy ~54%**
- Stronger on DISEASE, LAB_VALUE, ANATOMY
- Weaker on PROCEDURE and SYMPTOM

Decoder-only Transformer (Zero-shot)

- **Model:** Mistral-7B-Instruct (decoder-only, spaCy-LLM)
- **Setup:** Zero-shot medical NER on clinical notes

- **Observation:**
  - Few entities extracted
  - Section headers often misclassified (e.g. *“CLINICAL INDICATION”* → *DISEASE*)

- **Interpretation:**
  - Zero-shot setting + long clinical text is challenging

Key Learnings

- **General-purpose NER models** do not transfer well to clinical text
- **Custom medical labels** improve results, but performance is limited by dataset size
- **Encoder-based BERT models** perform best out of the box for medical NER
- **Biomedical LLMs** struggle in zero-shot settings and likely require few-shot examples or adaptation

Outlook: Task 3

- Compare **encoder (BERT)** vs. **decoder (GPT-like)** models
- Test **fine-tuning vs. prompting (zero-/few-shot)**
- Goal: **best-performing and most efficient model for clinical NER**

Q&A

Group C: Medical transcription data

Choekyel Nyungmartsang, Marc Nanzer, Matthias Peterhans, Vincent Gaspoz

- {'prompt': ['You are an expert Named Entity Recognition ' '(NER) system.\n' 'Your task is to accept Text as input and ' 'extract named entities.\n' 'Entities must have one of the following ' 'labels: ANATOMY, DISEASE, LAB_VALUE, ' 'MEDICATION, PROCEDURE, SYMPTOM.\n' 'If a span is not an entity label it: ' '`==NONE==`.\n'