

# Natural Language Understanding, Generation, and Machine Translation (2024–25)

*School of Informatics, University of Edinburgh*  
*Alexandra Birch*

## Tutorial 3: Ethics for NLP (Week 8)

### 1 Credit Risk Assessment

#### Question 1:

A startup wants you to help build an app that can help banks predict the creditworthiness of a person based on their social media activity. In simplified form: the input is the text of a user's social media posts, and system should classify the user into one of three categories that the bank will use to determine whether or not they receive a loan: LOW RISK, MEDIUM RISK, or HIGH RISK. A bank provides you with a large amount of training data based on their loan records. Each example consists of a user's twitter handle (from which the text of the user's tweets can be recovered) and the bank's classification of the user.

- Who are the stakeholders in the system? Your answer should include *direct stakeholders* who participated in system's creation (funders, developers, speakers or writers, annotators and curators) or who are the system's users. It should also include *indirect stakeholders* who do not use the system but are nonetheless impacted by it.
- What could go wrong?
- Who would be harmed? How?
- Who would benefit? How?
- What (if anything) might be a way mitigate such harm?

### 2 Automated Medical Coding

A company wants to hire you to develop an NLP system that fully automates **medical coding** from doctor's reports. One type of medical code identifies a diagnosis indicated in the report, while another identifies a prescribed treatment. Each code comes from a finite set. Here is a simplified example:

**Input** (doctor's report): *Patient is a 27-year-old white male. Height is 74 inches, weight 220 lbs. Patient states he is allergic to penicillin, but has no other outstanding medical history. Does not smoke, exercises moderately. Patient presents with chills, headache, cough, fever (101 degrees), difficulty breathing. Examination via stethoscope yields heavy rales. Percussion test on thorax suggests buildup in lungs. Streptococcal pneumoniae suspected. Prescribed patient two weeks of 500mg azithromycin (Zithromax), and scheduled follow-up for next week.*

**Output 1** (medical diagnosis code): pneumonia

**Output 2** (medical treatment code): azithromycin-500mg

**Question 2:**

Medical codes determine medical bills and authorize patients to receive certain treatments like prescription medicines, so they are a critical part of the modern healthcare system. They are also costly because they are produced by trained professionals. Many companies want to reduce this cost by using AI to automate medical coding.

- a. Who are the stakeholders in the system?
- b. What could go wrong?
- c. Who would be harmed? How?
- d. Who would benefit? How?
- e. What (if anything) might be a way mitigate such harm?

### 3 Bonus Question

**Question 3:**

Find a recent news article about harms caused by a machine learning system, ideally in natural language processing.

- a. Who are the stakeholders?
- b. What went wrong?
- c. Who was harmed? How?
- d. Who benefited? How?
- e. What (if anything) might be a way mitigate such harm in the future?

This is an open-ended question. Keeping in mind that ethics is an ongoing conversation, we encourage you to discuss your articles and findings with other participants in the course, including on piazza.