

# Resarching Responsible and Trustworthy Natural Language Processing

Session 1: Course Overview

---

Frank Keller, Catherine Lai, Caterina Moruzzi

18 September 2024

Informatics, PPLS, and ECA

University of Edinburgh

[keller@inf.ed.ac.uk](mailto:keller@inf.ed.ac.uk)

Introduction

Course Mechanics

Assessment

Your Immediate Tasks

We will start with a quick icebreaker task:

- Please briefly introduce yourself: name, pronouns, where you're from, which languages you speak.
- Say something about your academic background and research interests.
- Very briefly talk about a research project you've done recently (course project, MSc project, internship, etc.).

# Introduction

---

# Introduction

- This is a course on research methods.
- It covers generic things such as scientific writing, presentation skills, peer review.
- but also methods specific to responsible NLP:
  - replicability
  - design thinking
  - cluster computing
  - online experiments
  - ethical approval
- critically reading the scientific literature is another important skill; we will develop this in reading group sessions.
- We will also talk about public engagement and about working with industry.

## Relation to other Courses

- This course is designed so that you can apply the skills you learn in the Applied Interdisciplinary Project and in the Individual Exploratory Project.
- In addition, this course includes its own small project, the *methods project* in semester 1.
- Aspects of the Individual Exploratory Project (proposal writing, progress presentation) are integrated into this course in semester 2.

# Course Mechanics

---

## Course Format

- This is not a lecture course.
- But there will be tutorial-style material, some presented by the lecturers, some by guest speakers.
- We will rely on student participation for many aspects of the course.
- This can take the form of classroom discussion, contribution to Piazza, or student presentations.
- On some aspects of the course, students will work together in small groups (2–3 students).



- Two sessions per week. These will happen in person, but will also be recorded and live streamed.
- In some weeks, there will be small, non-assessed tasks. These will be announced in class or on Piazza.
- Lecturers: Frank Keller (Inf), Catherine Lai (PPLS), Caterina Moruzzi (ECA).
- Teaching assistants: Gautier Dagan (CDT in NLP) Anna Rezk (Design Informatics).
- The TAs will hold weekly office hours; details tba.

Like most Informatics courses, we will use a combination of Informatics OpenCourse, Learn, and Piazza:

- [The Informatics OpenCourse site](#) contains course materials (slides, readings, coursework exercises).
- [The University Learn site](#) contains “official” stuff like coursework deadlines, submission portal, lecture recordings.
- Piazza is linked from Learn will be used for discussion of course content, courseworks, and logistics.

# Assessment

---

This course is coursework-only, there is no exam. There will be three pieces of assessed coursework:

1. A research report on the methods project (semester 1), worth 40%.
2. A proposal for your individual exploratory project (semester 2), worth 30%.
3. A presentation reporting on the progress of your IEP (semester 2), worth 30%.

Please see the Learn course page for submission deadlines.

## Methods Project

- As part of this course, we will discuss various research methods across the participating disciplines.
- As a practical exercise, you will compare the methods used in two research papers.
- One of the papers will use NLP-oriented methods, the other one will use design-oriented methods.
- You will also be asked to design a followup study for one of the papers. This could be a replication or an extension of the paper.
- You will write up your work as an (assessed) report, and present your results to the class.
- This exercise will be done in pairs.

Timeline of the methods project:

- Week 4: We release the papers you will work with and form project groups.
- Week 8: Students submit their project reports (deadline: 8 Nov).
- Week 9: Each group gives a short presentation on their methods project.

## Your Immediate Tasks

---

## Task 1: Reading

For the writing skills part, we will use Alley's (2018) *The Craft of Scientific Writing*.

Alley covers scientific writing generally, but is also very relevant to responsible NLP. The e-book is linked from the OpenCourse site.

We will set reading for each session; a reading list will appear on OpenCourse.



## Task 2: Find Research Papers

- We will have a number of *reading group* sessions.
- In these, 2–3 students lead a discussion of a research paper.
- We would like to discuss papers of broad relevance to responsible NLP.
- You should start thinking about which papers you'd like to present.
- The TAs will soon start collecting your paper suggestions.
- The lecturers will vet them :)

We will look at the content of these paper, but also at their *writing* (structure, style, abstract, intro, figures/tables) and at *research methods*.

# On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?

Emily M. Bender\*

ebender@uw.edu

University of Washington

Seattle, WA, USA

Angelina McMillan-Major

aymm@uw.edu

University of Washington

Seattle, WA, USA

Timnit Gebru\*

timnit@blackinai.org

Black in AI

Palo Alto, CA, USA

Shmargaret Shmitchell

shmargaret.shmitchell@gmail.com

The Aether

### ABSTRACT

The past 3 years of work in NLP have been characterized by the development and deployment of ever larger language models, especially for English. BERT, its variants, GPT-2/3, and others, most recently Switch-C, have pushed the boundaries of the possible both through architectural innovations and through sheer size. Using these pretrained models and the methodology of fine-tuning them

alone, we have seen the emergence of BERT and its variants [39, 70, 74, 113, 146], GPT-2 [106], T-NLG [112], GPT-3 [25], and most recently Switch-C [43], with institutions seemingly competing to produce ever larger LMs. While investigating properties of LMs and how they change with size holds scientific interest, and large LMs have shown improvements on various tasks (§2), we ask whether enough thought has been put into the potential risks associated with developing them and strategies to mitigate these risks

## Task 3: Writing Exercise

For the next session (on Friday):

Please read *On the Dangers of Stochastic Parrots* by Bender et al. (2021). Then write summaries of the main points of the paper for the following audiences:

- A twelve year old child.
- The CEO of a tech startup.

Each summary should be no more than 200 words.

On Friday, we will discuss *the role of the audience in scientific writing*.

# References

Alley, Michael. 2018. *The Craft of Scientific Writing*. Springer, New York, NY, 4 edition.

Bender, Emily M., Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell. 2021. On the dangers of stochastic parrots: Can language models be too big? In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*. Association for Computing Machinery, New York, NY, USA, FAccT '21, page 610–623.