Introduction

Natural Language Processing and Large Language Models
Jan 13, 2026

RIT Rochester Institute of Technology

Outline

Outline

Course logistics

- Instructor: Dr. Hakyung Sung
  - PhD in Linguistics, MS in CS @ University of Oregon
  - Research interests: How to understand how people use their (second) languages by using NLP-based approach
- Grader: Bea Pulido
  - Grad student, Psychology
- Time: Tu/Th 8:00-9:15 AM
- Office hours: Tu/Th 9:30-10:30 AM (in-person, Office: EAS 3173) or Zoom by appointment
- Course website: `https://hksung.github.io/Spring26_PSYC681/` (Link on the MyCourses)
- Email: `hksgla@rit.edu`

RIT Rochester Institute of Technology

Learning goals

Let's recall the course title:
Natural Language Processing (NLP) &
Large Language Models (LLMs)

Learning goals

Let's recall the course title:
Natural Language Processing (NLP) &
Large Language Models (LLMs)

- What is NLP?
- What is LLM?

Learning goals

What is **NLP**?

Learning goals

What is **NLP**?

- "When scientists consider artificial intelligence, they mostly think of modeling or recreating the capabilities of an individual human brain... The power of language is fundamental to human societal intelligence, and language will retain an important role in a future world in which human abilities are augmented by artificial intelligence (AI) tools ... For these reasons, the field of natural language processing (NLP) emerged in tandem with the earliest development in AI." (Manning, 2022, pp. 127-128)

Sourced from: Manning, C. D. (2022). Human language understanding & reasoning. Daedalus, 151(2), 127-138.

RIT Rochester Institute of Technology

Learning goals

What is NLP?

Natural language processing (NLP) is a subfield of computer science and artificial intelligence (AI) that uses machine learning to enable computers to understand and communicate with human language.

Sourced from: `https://www.ibm.com/think/topics/natural-language-processing`

Learning goals

What is **NLP**?

- Started from Computer Science
- Now, people in different fields are so excited about NLP because they:
  - enables real-world applications (e.g., search, translation, chatbots, assessment)
  - leverage NLP to automatically analyze large-scale text data
  - can extract patterns, meanings, and structures that are difficult to capture manually
  - want to understand human minds; leverage nlp-techniques to generate and explore cogsi hypotheses

Learning goals

What is **LLM**?

Learning goals

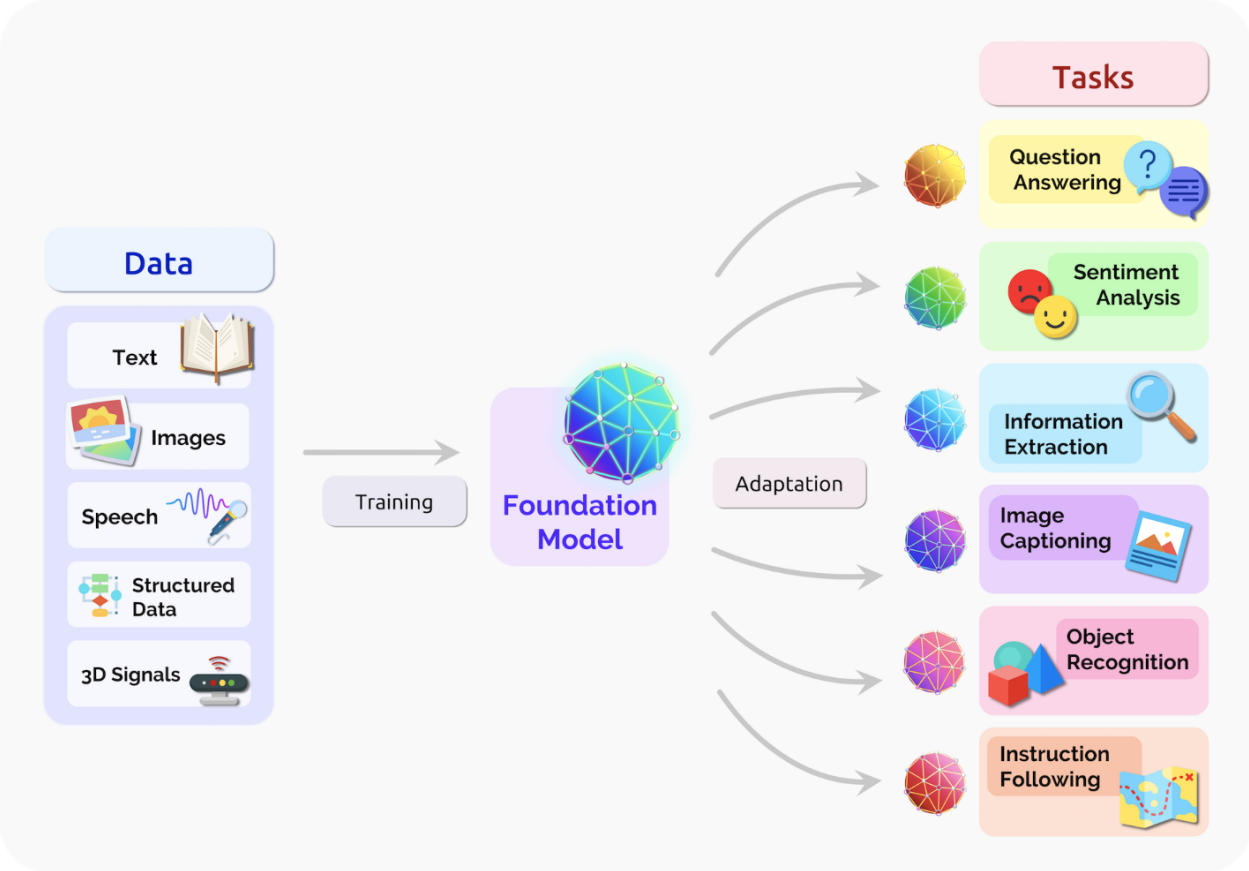
What is **LLM**?

- "A *foundation model* is any model that is trained on broad data (generally using self-supervision at scale) that can be adapted (e.g., fine-tuned) to a wide range of downstream tasks; current examples include BERT, GPT-3... We choose the term *foundation models* to capture the unfinished yet important status of these models... Existing terms (e.g., pretrained , self-supervised) partially capture the technical dimension of these models, but fail to capture the **significance of the paradigm shift in an accessible manner for those beyond machine learning**" (Bommasani et al., 2021, p. 3)

Sourced from: Bommasani, R. et al. (2021). On the opportunities and risks of foundation models. arXiv preprint arXiv:2108.07258.

RIT Rochester Institute of Technology

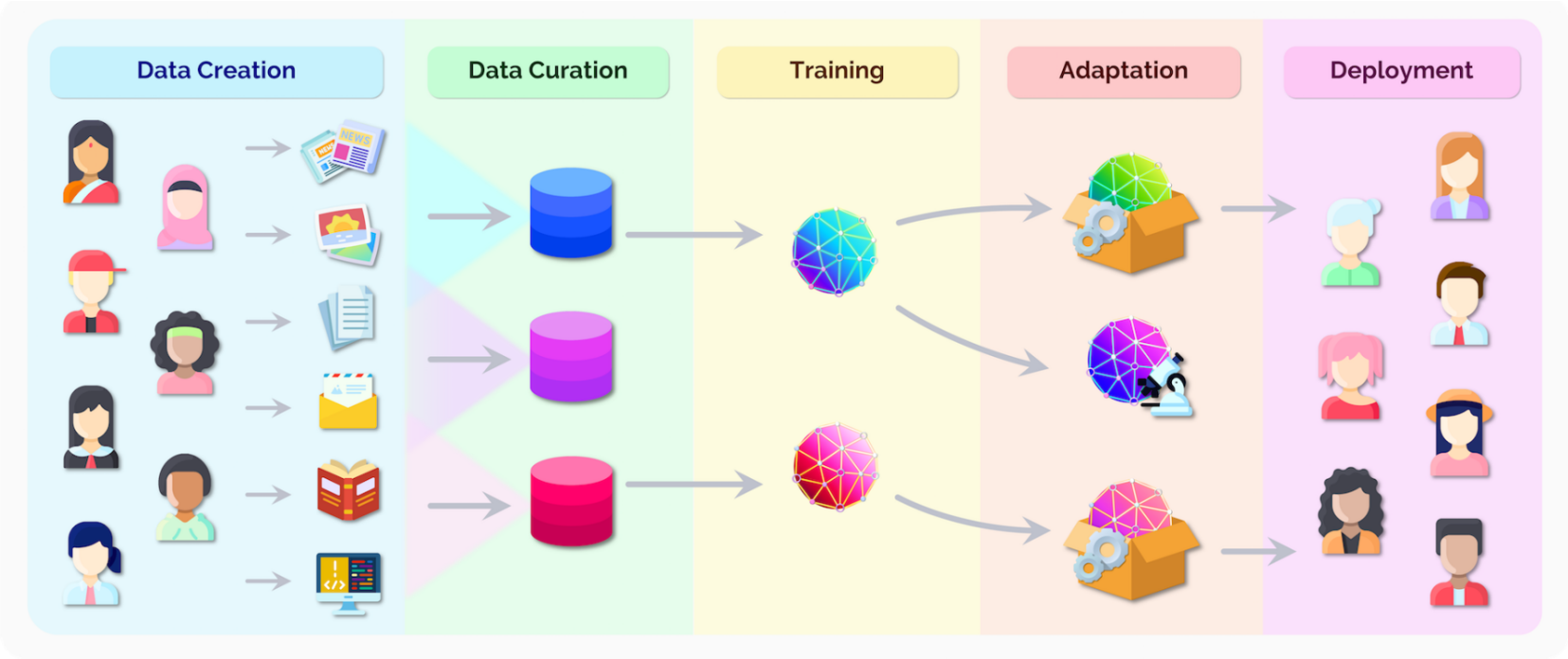
Learning goals



Sourced from: Bommasani, R. et al. (2021).

RIT Rochester Institute of Technology

Learning goals



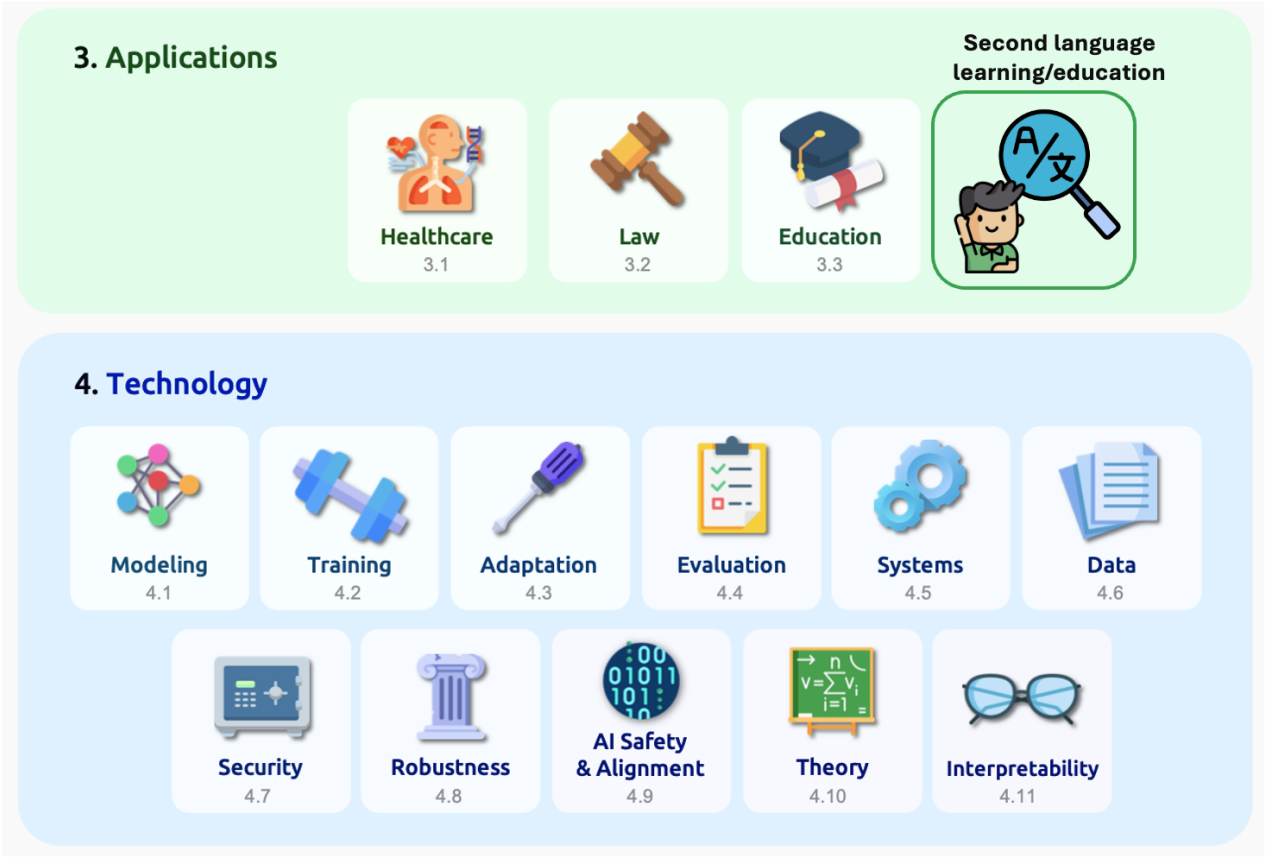
Sourced from: Bommasani, R. et al. (2021).

RIT Rochester Institute of Technology

Learning goals

Sourced from: Bommasani, R. et al. (2021).

Learning goals



RIT Rochester Institute of Technology

Learning goals

- Explain core NLP techniques and recent advancements (mostly the changes of `technology`)
- Gain hands-on experience applying basic NLP tools (including LLMs) and explore their use within domains of individual interest (`tools for applications`)
- Report on how they collaborated in *teams* to design and develop customized NLP solutions leveraged LLMs for real-world contexts (`applications`)

RIT Rochester Institute of Technology

Outline

Grading components: Lab

[a × b] a = number; b = points

- Lab [6 × 6]: 36%
- 8 labs in total (on Thursday)
- *Lab 1 and the lowest-scoring lab assignment will be dropped.

on the Syllabus

Week	Date	Topic	Paper assigned	Due (Friday, 11:59 pm)
1	1/13	Introduction, Word vectors		
	1/15	Lab1 Python basics		Lab 1
2	1/20	Word vectors	[1] Mikolov et al. (2013) [2] Pennington et al. (2014)	
	1/22	Lab2 Word vectors	[3] Levy et al. (2015)	Lab 2
3	1/27	Neural network	[4] Collobert et al. (2011) §3	
	1/29	Lab 3 PyTorch, Project guide		Lab 3
4	2/3	Dependency parsing	[5] Chen & Manning (2014)	
	2/5	Lab 4 Dependency parsing	[6] de Marneffe et al. (2021)	Identify team members
5	2/10	RNNs, LSTMs	[7] Sak et al. (2014)	
	2/12	Lab 5 Sentiment analysis	[8] Du et al. (2024)	Lab 4
6	2/17	Self-attention & Transformer	[9] Vaswani et al. (2017)	
	2/19	Lab 6 Hugging face	[10] Huang et al. (2018)	Lab 5
7	2/24	Pre-training	[11] Devlin (2019)	
	2/26	Lab 7 Ollama	[12] Smith (2020)	Lab 6
8	3/3	Project group meeting		
	3/5	Project proposal presentation		Lab 7
9	3/10	**Spring break (No class)**		
	3/12			Project proposal
10	3/17	Post-training	[13] Chung et al. (2022) [14] Wang et al. (2022)	
	3/19	Lab 8 Constructed language	[15] Taguchi & Sproat (2025)	
11	3/24	LLMs in 2025 (async)		
	3/26	Efficient Adaptation	[16] Brown et al. (2020) [17] Hu et al. (2021)	Lab 8

Grading components: Paper presentation

[a × b] a = number; b = points

- Paper presentations [1 × 6] 6%

on the Syllabus

Week	Date	Topic	Paper assigned
1	1/13	Introduction, Word vectors	
	1/15	Lab1 Python basics	
2	1/20	Word vectors	[1] Mikolov et al. (2013) [2] Pennington et al. (2014)
	1/22	Lab2 Word vectors	[3] Levy et al. (2015)
3	1/27	Neural network	[4] Collobert et al. (2011) §3
	1/29	Lab 3 PyTorch, Project guide	
4	2/3	Dependency parsing	[5] Chen & Manning (2014)
	2/5	Lab 4 Dependency parsing	[6] de Marneffe et al. (2021)
5	2/10	RNNs, LSTMs	[7] Sak et al. (2014)
	2/12	Lab 5 Sentiment analysis	[8] Du et al. (2024)
6	2/17	Self-attention & Transformer	[9] Vaswani et al. (2017)
	2/19	Lab 6 Hugging face	[10] Huang et al. (2018)
7	2/24	Pre-training	[11] Devlin (2019)
	2/26	Lab 7 Ollama	[12] Smith (2020)
8	3/3	Project group meeting	
	3/5	Project proposal presentation	
9	3/10	**Spring break (No class)**	
	3/12		
10	3/17	Post-training	[13] Chung et al. (2022) [14] Wang et al. (2022)
	3/19	Lab 8 Constructed language	[15] Taguchi & Sproat (2025)
11	3/24	LLMs in 2025 (async)	
	3/26	Efficient Adaptation	[16] Brown et al. (2020) [17] Hu et al. (2021)
12	3/31	Background research presentation 1	
	4/2	Background research presentation 2	
13	4/7	Benchmarking and evaluation	[18] Hendrycks et al. (2021) [19] Liang et al. (2023) §2-3
	4/9	Agent, Tool use, RAG	[20] Yao et al. (2023) [21] Shick et al. (2023)
14	4/14	Reasoning 1	[22] Wei et al. (2023) [23] Wang et al. (2023)
	4/16	Reasoning 2	[24] Lightman et al. (2023) [25] Snell et al. (2024)

Grading components: Paper Presentation

- Guidelines for preparing the presentation:
  `https://hksung.github.io/Spring26_PSYC681/assignments/1_paper%20presentation`
- At the end of today's class, you will have time to choose the paper you want to present.
- Presentations will begin next Tuesday.
  Tip: The first two papers are low-hanging fruits.

RIT Rochester Institute of Technology

Grading components: Project

Project 48%

- Identify team members [1 × 2] 2%
- Project proposal [1 × 4] 4%
- Midway report [1 × 7] 7%
- Background research presentation [1 × 8] 8%
- Final presentation [1 × 12] 12%
- Final paper [1 × 15] 15%

Grading component: Project

- This is the **core** of the course: `application` or `technology`
- Students will work in teams of 3–4.
- The most important date for now:
  - **Feb 5**: Deadline to finalize team members.
  - Earlier decisions on teams/topics will make the course easier, as you can directly apply your project ideas during lab sessions and presentations.
  - *All other due dates are on the syllabus.
- Welcome projects that align with my research interests (e.g., language education/learning); Let me know.
- Ideal goals: Submit to conference/workshop (e.g., EMNLP mid May)

RIT Rochester Institute of Technology

Grading components: Participation

- Completing ad-hoc events that the instructor announces in class 1%
- Active participation in paper presentation 3%
- Active participation in background research presentation 3%
- Active participation in final presentation 3%

Any questions?

Grading policy

- 2-hr grading window: Any assignment submitted online will automatically have a 2-hour grading window. This will be applied by the system, and no action is required from students.
- Late penalty: Late assignments will incur a 10% deduction per day, for up to 5 days (e.g., 1 day late = 10% off). After 5 days, the assignment will receive a grade of zero.
- Extenuating circumstances: Whenever possible, please request an official document that can prove the circumstances—this allows me to accommodate you fairly while respecting your privacy.
- No extensions will be granted for the final paper.
- **Will be strict about this policy**

RIT Rochester Institute of Technology

Collaboration policy

- For the group works, all members are expected to contribute their time and effort equally.
- Collaboration with AI tools is fine, but you are responsible for the quality and integrity of the work produced (e.g., double check the references)
- Please acknowledge and document how AI tools were used in your work (including individual exercises).

Outline

In Class

- You are welcome to take notes on your laptop.
- **BUT please** be respectful and attentive when others are speaking (e.g., the instructor or your peers).
- Early morning classes can be challenging (especially in this weather) but being present and engaged will help.
- If weather conditions are severe, I will notify you by 7:00 am on the day of class regarding cancellation or a switch to Zoom.

Before you leave

Please complete the following tasks:

- Go to MyCourses and check how to access the course website.
- Check the **Assignments** tab → *1. Paper Presentation*.
- Paper Selection: Find the link at the bottom, review the papers, and select your paper (first come, first served).
- Check the **Lectures** tab → find today's *slides* (This should be how I share all materials).
- Click the link below to complete the mini survey: `https://forms.gle/G6n2RVbPD12PewfN7`
- Check the syllabus while answering to the mini survey (which is attached to the homepage or MyCourses).

RIT Rochester Institute of Technology

Preview: On Thursday

- We will have our first lab session on Thursday to brush up Python skills. Please bring your laptop. (You will need your laptop for every lab session.)

Thank you and see you on Thursday!