



## **Class Participation**

Regular class participation is expected regardless of course modality. Students who fail to participate in class regularly are inviting scholastic difficulty. A portion of the grade for this course is directly tied to your participation in this class. It also includes engaging in group or other activities during class that solicit your feedback on homework assignments, readings, or materials covered in the lectures (and/or labs). Class participation is documented by faculty. Successful participation is defined as consistently adhering to University requirements, as presented in this syllabus. Failure to comply with these University requirements is a violation of the [Student Code of Conduct](#).

---

## **Class Materials**

The Instructor may provide class materials that will be made available to all students registered for this class as they are intended to supplement the classroom experience. These materials may be downloaded during the course, however, these materials are for registered students' use only. Classroom materials may not be reproduced or shared with those not in class, or uploaded to other online environments except to implement an approved Office of Student AccessAbility accommodation. Failure to comply with these University requirements is a violation of the [Student Code of Conduct](#).

## **Course Pre-requisites, Co-requisites, and/or Other Restrictions**

Natural Language Processing (NLP)/ Big Data Management and Analytics (CS 6350), JAVA (intermediate/advanced)/Python, Linux OS

## **Course Description**

Standard supervised learning requires labeled data, which are expensive to obtain due to the expertise required for quality annotation. This led scholars to seek alternative solutions based on the latest developments, Large Language Models (LLM) in natural language processing (NLP). Recent progress in NLP has been driven by pre-trained transformer language models. Self-supervision using large-scale unlabeled text can significantly alleviate the annotation bottleneck using transfer learning. The training parallelization of transformers also improves their efficiency on large datasets. As a result, the use of powerful computational devices and the advantage of transformer structures make large-scale language models' pre-training possible. We expect in this seminar class (1) we will study a family of LLMs, (2) pros and cons of each LLM model, (3) security issues and (4) applications.

The course will focus on reading papers related to LLM, and analyze very large amounts of text data beyond English or Big data using LLM. The students will give presentations in the class and do a project.

The course material will be drawn from papers from premier conferences to capture recent research.

### Required Textbooks and Materials

N/A

### Suggested Course Materials

#### Papers Related to Large Language Models [May Expand Further]

- Vaswani, Ashish & Shazeer, Noam & Parmar, Niki & Uszkoreit, Jakob & Jones, Llion & Gomez, Aidan & Kaiser, Lukasz & Polosukhin, Illia, [“Attention is all you need”](#), [Advances in Neural Information Processing Systems 30 \(NIPS 2017\)](#) 2017.
- Jacob Devlin and Ming-Wei Chang and Kenton Lee and Kristina Toutanova, BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, North American Chapter of the Association for Computational Linguistics (NACACL2019)
- [Basel Abdeen](#), [Ehab Al-Shaer](#), [Anoop Singhal](#), [Latifur Khan](#), [Kevin W. Hamlen](#): **SMET: Semantic Mapping of CVE to ATT&CK and Its Application to Cybersecurity**. [DBSec 2023](#): 243-260
- [Sadaf Md. Halim](#), [Saquib Irtiza](#), [Yibo Hu](#), [Latifur Khan](#), [Bhavani Thuraisingham](#): **WokeGPT: Improving Counterspeech Generation Against Online Hate Speech by Intelligently Augmenting Datasets Using a Novel Metric**. [IJCNN 2023](#): 1-10
- [Chen Zhao](#)<sup>id</sup>, [Feng Mi](#)<sup>id</sup>, [Xintao Wu](#)<sup>id</sup>, [Kai Jiang](#)<sup>id</sup>, [Latifur Khan](#)<sup>id</sup>, [Christan Grant](#)<sup>id</sup>, [Feng Chen](#)<sup>id</sup>: **Towards Fair Disentangled Online Learning for Changing Environments**. [KDD 2023](#): 3480-3491
- [Sultan Alsarra](#), [Luay Abdeljaber](#), [Wooseong Yang](#), [Niamat Zawad](#), [Latifur Khan](#), [Patrick T. Brandt](#), [Javier Osorio](#), [Vito D'Orazio](#): **ConfliBERT-Arabic: A Pre-trained Arabic Language Model for Politics, Conflicts and Violence**. [RANLP 2023](#): 98-108
- [Yibo Hu](#), [Erick Skorupa Parolin](#), [Latifur Khan](#), [Patrick T. Brandt](#), [Javier Osorio](#), [Vito J. D'Orazio](#): **Synthesizing Political Zero-Shot Relation Classification via Codebook Knowledge, NLI, and ChatGPT**. [CoRR abs/2308.07876](#) (2023)
- [Erick Skorupa Parolin](#), [Yibo Hu](#), [Latifur Khan](#), [Patrick T. Brandt](#), [Javier Osorio](#), [Vito D'Orazio](#): **Confli-T5: An AutoPrompt Pipeline for Conflict Related Text Augmentation**. [IEEE Big Data 2022](#): 1906-1913
- [Yibo Hu](#), [Yu Lin](#), [Erick Skorupa Parolin](#), [Latifur Khan](#), [Kevin W. Hamlen](#): **Controllable Fake Document Infilling for Cyber Deception**. [EMNLP\(Findings\) 2022](#): 6505-6519
- [Yibo Hu](#), [Seyyed MohammadSaleh Hosseini](#), [Erick Skorupa Parolin](#), [Javier Osorio](#), [Latifur Khan](#), [Patrick T. Brandt](#), [Vito D'Orazio](#): **ConfliBERT: A Pre-trained Language Model for Political Conflict and Violence**. [NAACL-HLT 2022](#): 5469-5482
- [Yibo Hu](#), [Latifur Khan](#): **Uncertainty-Aware Reliable Text Classification**. [KDD 2021](#): 628-636

- [Zhuoyi Wang, Yuqiao Chen, Chen Zhao](#), [Yu Lin, Xujiang Zhao, Hemeng Tao, Yigong Wang](#), Latifur Khan: **CLEAR: Contrastive-Prototype Learning with Drift Estimation for Resource Constrained Stream Mining**. *WWW 2021*: 1351-1362
- [Yang Gao, Yi-Fan Li, Yu Lin, Charu C. Aggarwal](#), Latifur Khan: **SetConv: A New Approach for Learning from Imbalanced Data**. *EMNLP (1) 2020*: 1284-1294
- Zheng-Xin Yong, Cristina Menghini, Stephen H. Bac, Low-Resource Languages Jailbreak GPT-4 [2310.02446.pdf \(arxiv.org\)](https://arxiv.org/abs/2310.02446)
- Deep contextualized word representations, <https://arxiv.org/pdf/1802.05365.pdf>
- Improving Language Understanding by Generative Pre-Training [https://s3-us-west-2.amazonaws.com/openai-assets/research-covers/language-unsupervised/language\\_understanding\\_paper.pdf](https://s3-us-west-2.amazonaws.com/openai-assets/research-covers/language-unsupervised/language_understanding_paper.pdf)
- RoBERTa: A Robustly Optimized Bert Pretraining Approach, <https://arxiv.org/pdf/1907.11692.pdf>
- Extracting Training Data from Large Language Models <https://arxiv.org/pdf/2012.07805.pdf>
- Recovering Private Text in Federated Learning of Language Models <https://arxiv.org/pdf/2205.08514.pdf>
- Multimodal Few-Shot Learning with Frozen Language Models <https://arxiv.org/pdf/2106.13884.pdf>
- Evaluating Large Language Models Trained on Code <https://arxiv.org/pdf/2107.03374.pdf>

## Software

Huggingface: <https://huggingface.co/>

<https://www.cs.princeton.edu/courses/archive/fall22/cos597G/>

<https://stanford-cs324.github.io/winter2022/lectures/modeling/>

## Technical Requirements

In addition to a confident level of computer and Internet literacy, certain minimum technical requirements must be met to enable a successful learning experience. Please review the important technical requirements on the [Getting Started with eLearning](#) webpage.

## Course Access and Navigation

This course can be accessed using your UT Dallas NetID account on the [eLearning](#) website.

Please see the course access and navigation section of the [Getting Started with eLearning](#) webpage for more information.

To become familiar with the eLearning tool, please see the [Student eLearning Tutorials](#) webpage.

UT Dallas provides eLearning technical support 24 hours a day, 7 days a week. The [eLearning Support Center](#) includes a toll-free telephone number for immediate assistance (1-866-588-3192), email request service, and an online chat service.

### **Communication**

This course utilizes online tools for interaction and communication. Some external communication tools such as regular email and a web conferencing tool may also be used during the semester. For more details, please visit the [Student eLearning Tutorials](#) webpage for video demonstrations on eLearning tools.

Student emails and discussion board messages will be answered within 3 working days under normal circumstances.

### **Distance Learning Student Resources**

Online students have access to resources including the McDermott Library, Academic Advising, The Office of Student AccessAbility, and many others. Please see the [eLearning Current Students](#) webpage for more information.

### **Server Unavailability or Other Technical Difficulties**

The University is committed to providing a reliable learning management system to all users. However, in the event of any unexpected server outage or any unusual technical difficulty which prevents students from completing a time sensitive assessment activity, the instructor will provide an appropriate accommodation based on the situation. Students should immediately report any problems to the instructor and also contact the online [eLearning Help Desk](#). The instructor and the eLearning Help Desk will work with the student to resolve any issues at the earliest possible time.

### **Academic Calendar**

#### **Lectures (to be added)**

Topic	Chapters/Paper	Homework/Lecture Notes
Introduction to LLM		
Basics of RNN, LSTM, Attention, Transformer		
Encoder-Decoderf		
Presentation	<b>TBD</b>	

- **Classes begin Monday, January 16**
- **The last day of classes: May 2.**
- **No classes: Spring break March 11 to 17**

### **Proctored Final Exam Procedures**

If your course has a proctored exam requirement, please see the [UTD Testing Center](#) webpage and [Distance Learning Proctored Exams](#) webpage to make arrangements.

### **Grading Policy**

Your course grade will be based on  
 Paper presentation (3 papers/student): 45%  
 Class Participation: 15%  
 Project” 40%

### **Course Policies**

*Make-up exams: N/A*  
*Extra Credit: N/A*

### **Comet Creed**

This creed was voted on by the UT Dallas student body in 2014. It is a standard that Comets choose to live by and encourage others to do the same:

*“As a Comet, I pledge honesty, integrity, and service in all that I do.”*

### **Academic Support Resources**

The information contained in the following link lists the University’s academic support resources for all students.

Please go to [Academic Support Resources](#) webpage for these policies.

### **UT Dallas Syllabus Policies and Procedures**

The information contained in the following link constitutes the University’s policies and procedures segment of the course syllabus.

Please go to [UT Dallas Syllabus Policies](#) webpage for these policies.

*The descriptions and timelines contained in this syllabus are subject to change at the discretion of the Professor.*

### **Field Trip Policies**

### **Off-campus Instruction and Course Activities**

*Off-campus, out-of-state, and foreign instruction and activities are subject to state law and University policies and procedures regarding travel and risk-related activities. Information regarding these rules and regulations may be found at the website address [http://www.utdallas.edu/BusinessAffairs/Travel\\_Risk\\_Activities.htm](http://www.utdallas.edu/BusinessAffairs/Travel_Risk_Activities.htm). Additional information is available from the office of the school dean. Below is a description of any travel and/or risk-related activity associated with this course.*

---

## **Student Conduct & Discipline**

The University of Texas System and The University of Texas at Dallas have rules and regulations for the orderly and efficient conduct of their business. It is the responsibility of each student and each student organization to be knowledgeable about the rules and regulations which govern student conduct and activities. General information on student conduct and discipline is contained in the UTD publication, *A to Z Guide*, which is provided to all registered students each academic year.

The University of Texas at Dallas administers student discipline within the procedures of recognized and established due process. Procedures are defined and described in the *Rules and Regulations, Board of Regents, The University of Texas System, Part 1, Chapter VI, Section 3*, and in Title V, Rules on Student Services and Activities of the university's *Handbook of Operating Procedures*. Copies of these rules and regulations are available to students in the Office of the Dean of Students, where staff members are available to assist students in interpreting the rules and regulations (SU 1.602, 972/883-6391).

A student at the university neither loses the rights nor escapes the responsibilities of citizenship. He or she is expected to obey federal, state, and local laws as well as the Regents' Rules, university regulations, and administrative rules. Students are subject to discipline for violating the standards of conduct whether such conduct takes place on or off campus, or whether civil or criminal penalties are also imposed for such conduct.

## **Academic Integrity**

The faculty expects from its students a high level of responsibility and academic honesty. Because the value of an academic degree depends upon the absolute integrity of the work done by the student for that degree, it is imperative that a student demonstrate a high standard of individual honor in his or her scholastic work.

Scholastic dishonesty includes, but is not limited to, statements, acts or omissions related to applications for enrollment or the award of a degree, and/or the submission as one's own work or material that is not one's own. As a general rule, scholastic dishonesty involves one of the following acts: cheating, plagiarism, collusion and/or falsifying academic records. Students suspected of academic dishonesty are subject to disciplinary proceedings.

Plagiarism, especially from the web, from portions of papers for other classes, and from any other source is unacceptable and will be dealt with under the university's policy on plagiarism (see general catalog for details). This course will use the resources of turnitin.com, which searches the web for possible plagiarism and is over 90% effective.

## **Email Use**

The University of Texas at Dallas recognizes the value and efficiency of communication between faculty/staff and students through electronic mail. At the same time, email raises some issues concerning security and the identity of each individual in an email exchange. The university encourages all official student email correspondence be sent only to a student's U.T. Dallas email address and that faculty and staff consider email from students official only if it originates from a UTD student account. This allows the university to maintain a high degree of confidence in the identity of all individual corresponding and the security of the transmitted information. UTD furnishes each student with a free email account that is to be used in all communication with university personnel. The Department of Information Resources at U.T. Dallas provides a method for students to have their U.T. Dallas mail forwarded to other accounts.

### **Withdrawal from Class**

The administration of this institution has set deadlines for withdrawal of any college-level courses. These dates and times are published in that semester's course catalog. Administration procedures must be followed. It is the student's responsibility to handle withdrawal requirements from any class. In other words, I cannot drop or withdraw any student. You must do the proper paperwork to ensure that you will not receive a final grade of "F" in a course if you choose not to attend the class once you are enrolled.

### **Student Grievance Procedures**

Procedures for student grievances are found in Title V, Rules on Student Services and Activities, of the university's *Handbook of Operating Procedures*.

In attempting to resolve any student grievance regarding grades, evaluations, or other fulfillments of academic responsibility, it is the obligation of the student first to make a serious effort to resolve the matter with the instructor, supervisor, administrator, or committee with whom the grievance originates (hereafter called "the respondent"). Individual faculty members retain primary responsibility for assigning grades and evaluations. If the matter cannot be resolved at that level, the grievance must be submitted in writing to the respondent with a copy of the respondent's School Dean. If the matter is not resolved by the written response provided by the respondent, the student may submit a written appeal to the School Dean. If the grievance is not resolved by the School Dean's decision, the student may make a written appeal to the Dean of Graduate or Undergraduate Education, and the dean will appoint and convene an Academic Appeals Panel. The decision of the Academic Appeals Panel is final. The results of the academic appeals process will be distributed to all involved parties.

Copies of these rules and regulations are available to students in the Office of the Dean of Students, where staff members are available to assist students in interpreting the rules and regulations.

### **Incomplete Grade Policy**

As per university policy, incomplete grades will be granted only for work unavoidably missed at the semester's end and only if 70% of the course work has been completed. An incomplete grade must be resolved within eight (8) weeks from the first day of the subsequent long semester. If the required work to complete the course and to remove the incomplete grade is not submitted by the specified deadline, the incomplete grade is changed automatically to a grade of **F**.

### **Disability Services**

The goal of Disability Services is to provide students with disabilities educational opportunities equal to those of their non-disabled peers. Disability Services is located in room 1.610 in the

Student Union. Office hours are Monday and Thursday, 8:30 a.m. to 6:30 p.m.; Tuesday and Wednesday, 8:30 a.m. to 7:30 p.m.; and Friday, 8:30 a.m. to 5:30 p.m.

The contact information for the Office of Disability Services is:  
The University of Texas at Dallas, SU 22  
PO Box 830688  
Richardson, Texas 75083-0688  
(972) 883-2098 (voice or TTY)

Essentially, the law requires that colleges and universities make those reasonable adjustments necessary to eliminate discrimination on the basis of disability. For example, it may be necessary to remove classroom prohibitions against tape recorders or animals (in the case of dog guides) for students who are blind. Occasionally an assignment requirement may be substituted (for example, a research paper versus an oral presentation for a student who is hearing impaired). Classes enrolled students with mobility impairments may have to be rescheduled in accessible facilities. The college or university may need to provide special services such as registration, note-taking, or mobility assistance.

It is the student's responsibility to notify his or her professors of the need for such an accommodation. Disability Services provides students with letters to present to faculty members to verify that the student has a disability and needs accommodations. Individuals requiring special accommodation should contact the professor after class or during office hours.

### **Religious Holy Days**

The University of Texas at Dallas will excuse a student from class or other required activities for the travel to and observance of a religious holy day for a religion whose places of worship are exempt from property tax under Section 11.20, Tax Code, Texas Code Annotated.

The student is encouraged to notify the instructor or activity sponsor as soon as possible regarding the absence, preferably in advance of the assignment. The student, so excused, will be allowed to take the exam or complete the assignment within a reasonable time after the absence: a period equal to the length of the absence, up to a maximum of one week. A student who notifies the instructor and completes any missed exam or assignment may not be penalized for the absence. A student who fails to complete the exam or assignment within the prescribed period may receive a failing grade for that exam or assignment.

If a student or an instructor disagrees about the nature of the absence [i.e., for the purpose of observing a religious holy day] or if there is similar disagreement about whether the student has been given a reasonable time to complete any missed assignments or examinations, either the student or the instructor may request a ruling from the chief executive officer of the institution, or his or her designee. The chief executive officer or designee must take into account the legislative intent of TEC 51.911(b), and the student and instructor will abide by the decision of the chief executive officer or designee.

***These descriptions and timelines are subject to change at the discretion of the Professor.***