I. Course Description and Objectives

In this experimental course, Pardee students will work with computer-science students to map different kinds of dangerous speech. We will assess scholars’ understanding of the causes of prejudice and hate speech, and consider a variety of approaches to prevent harm. We will explore scholars’ understanding of what makes content go viral, evaluate attempts to automate the identification of hate speech and to measure its prevalence; and consider how hate speech is countered around the world. We will also consider the relative utility of censoring versus countering hate speech.

Students will be able to choose their project: 1) Collecting data from online platforms, (2) Mapping hate-speech content, and (3) To be determined by students by working with faculty.

Course Objectives:

This course fulfills the following BU Hub learning outcomes:

1. Teamwork/Collaboration (TWC) (2 Units)

   Outcome 1: As a result of explicit training in teamwork and sustained experiences of collaborating with others, students will be able to identify the characteristics of a well-functioning team.

   - As a result of explicit training in teamwork and sustained experiences of collaborating with others, students will be able to identify the characteristics of a well-functioning team. Students will achieve this objective through working in teams to (1) research hate speech on social media websites; (2) present research to the class; and (3) write a final paper summarizing the group’s research.

   Outcome 2: Students will demonstrate an ability to use the tools and strategies of working successfully with a diverse group, such as assigning roles and responsibilities, giving and receiving feedback, and engaging in meaningful group reflection that inspires collective ownership of results.

   - Students will demonstrate an ability to use the tools and strategies of working successfully with a diverse group, such as assigning roles and responsibilities, giving and receiving feedback, and engaging in meaningful group reflection that inspires collective ownership of results, through (1)
a group research project on online hate speech in which student groups are comprised of around 5 students and a team manager organizes roles and responsibilities; (2) repeated oral group presentations of research findings throughout the semester to gain feedback from other groups; and (3) student groups will collaborate in writing a final paper to relay their research findings.

2. Writing Intensive Course (WIN) (2 Units)

**Outcome 1:** Students will be able to craft responsible, considered, and well-structured written arguments, using media and modes of expression appropriate to the situation.

- Students will read important empirical and theoretical articles and gray literature on the study of hate speech. Through writing weekly analytic reading responses, students will practice crafting measured arguments about how to combat online hate speech and extremism, and the role of government in doing so.

**Outcome 2:** Students will be able to read with understanding, engagement, appreciation, and critical judgment.

- The ability to read with understanding and critical judgment, as demonstrated by applying reading material to form questions for guest speaker sessions, is critical to success in IR 558. In both written assignments and discussions, students will be expected to demonstrate an ability to critically evaluate the readings, and develop an argument in the debate on the government’s role in regulating hate speech.

**Outcome 3:** Students will be able to write clearly and coherently in a range of genres and styles, integrating graphic and multimedia elements as appropriate.

- Students will practice writing about the weekly readings such that the responses will incorporate class discussions, previously assigned readings, documentaries shown in class, and guest speaker sessions. Students will also complete a final paper on their group’s research. Feedback will be given regularly and students are encouraged to work with the writing center.

3. Individual in Community (IIC) (1 Unit)

**Outcome 1:** Students will analyze at least one of the dimensions of experience—historical, racial, socio-economic, political, gender, linguistic, religious, or cultural—that inform their own worldviews and beliefs as well as those of other individuals and societies.

- Students will analyze at least one of the dimensions of experience—historical, racial, socio-economic, political, gender, linguistic, religious, or cultural—that inform their own worldviews and beliefs as well as those of other individuals and societies through, (1) analyzing similarities between users that post hate speech against the same sect of society; and understanding (2) how individuals conceive of stereotypes based on a two-dimensional classification system of “warmth” and “competence;” (3) how societal breakdowns can activate resentment toward minorities who are seen as competitors for limited resources; and (4) how exposure to hate speech increases prejudice through desensitization.
Outcome 2: Students will participate respectfully in different communities such as campus, citywide, national, and international groups, and recognize and reflect on the issues relevant to those communities.

- Students will participate respectfully in different communities such as campus, citywide, national, and international groups, and recognize and reflect on the issues relevant to those communities, through (1) communicating online with individuals who post hate speech by using fake accounts to understand their grievances and their desires; and (2) developing an understanding of the debate on the role different government and private companies play in regulating, or not regulating, hate speech online.

Prerequisites: All students enrolled in IR 558 must have completed the first-year writing sequence (WR 120 and WR 150 or the equivalent).

II. Course Requirements

Class Participation (30%): Students will be expected to attend regularly, read thoughtfully, and participate actively in class discussions and group projects. Students are required to come see Professor Stern during her office hours in the first month of class.

Response Papers/Presentations (30%): Students are required to write six short (2-3 pages, double-spaced, 1 inch margins) analytical response papers throughout the semester. The lowest grade will be dropped. Any quotations must be cited appropriately. Help with citations is available at the Writing Center as well as online. The response paper will summarize and analyze the core arguments for each of the day’s readings, and how the readings build on what we have already discussed in class. Response papers will be evaluated on the basis of comprehension, accuracy, clarity, critical engagement, and organization. Students must submit the response papers to Blackboard by midnight Thursday night before their assigned topic is discussed in class. (Ex. If Class is on Monday September 28, the response paper would be due on Thursday September 24.)

Students are also required to give one 10-minute presentation on their findings from one of the response papers, and raise questions for the class to discuss. The purpose of these presentations is to give you an opportunity to practice public speaking, and experience creating succinct and informative presentations. The grade for this assignment will count as part of your class participation.

Team Work: Students will work together to determine the best approach to identify and evaluate hate speech online, partially in collaboration with the Anti-Defamation League.

All teams are required to devise a group contract (see below) during the 4th week of class. In order to succeed, team members must be committed to accomplishing the team's goals, understand members' expectations, and establish positive team norms, processes, and interactions. The contract will explicitly identify each team’s goals, roles, and processes for communication. The team contract will help guide each group toward a successful final product. It is expected that the goals and roles of the project may change throughout the semester. Any changes should be documented, and groups should submit both an original and final version of the team contract at the end of the semester.

Team Contract
- Goals
  o What outcomes are we trying to achieve?
  o Why does our work matter?
What matters most to the team?
Who does our work matter to?
What do we imagine the final product will look like?

- **Processes**
  - What work are we doing as a team?
  - What is the schedule for our team’s interactions?
  - How will we communicate? On what platform?

- **Roles**
  - Who is our team manager?
  - What are all the roles of the team members?

- **Interactions**
  - What is our contingency plan in the case that there is tension among team members, for example, if one member of the team free rides?

**Team Purpose** (may take some time to determine):

- Team purpose is an aspiration, not a metric. Team purpose is your team’s guiding light – it should be both inspiring and grounding.
- Team purpose statement: Our team exists to _____ for _____.

---

**Specific projects requested by the Anti-defamation League:**

1. **Discord** – With limited visibility into the various servers housed by Discord, it could be useful to identify extremist narratives on available (public) servers, track individual users on those servers and try to gain access into other servers. This effort to map potential extremist activity/narratives on Discord could not only broaden our visibility in that space, but could also lead to an analysis of how Discord is being used to spread or amplify extremist narratives.

2. **Extremist Scorecard** – Individual extremist use various social platforms to reach, recruit and radicalize. COE can provide a list of 100 key extremists for research that would result in a scorecard listing how many platforms each individual extremist uses (Facebook, Discord, PayPal, etc.). This scorecard could also lend itself to a network analysis explaining how individual extremists use specific platforms and multiple platforms together, shedding light on online ecosystems.

3. **Stormfront** – As one of the oldest extremist forums with millions of posts, it is unclear how its user base has changed over the years. An analysis of the membership and posts may provide insight into its utility for a new generation of extremists, or if it remains mostly popular with an older generation.

4. **Sports Council** – With so many athletes active and visible on social media, an analysis of the type of hate or abuse athletes receive on social media (Twitter, Instagram, etc.) would be a first of its kind. The finding could also be leveraged with ADL’s Sports Council.

5. **Enemy Journalists** – With journalists increasingly targeted online and in the public discourse – including by violent extremists (Hasson, Sayoc, etc.) – an analysis of Reddit, 4Chan, etc. to gauge the volume and type of narratives targeting journalists could provide insight into the impact of public attacks on journalists (“enemies of the people”) – seeing if there are links between public attacks against journalists with online activity and behavior.
The project will lay the groundwork for the development of repository that includes an open source library containing data that can be made available for researchers from around the world. The repository will be kept by Professor Gianluca Stringhini. The BU team will seek to leverage tools and data collections already developed by the ADL such as the machine learning models utilized in the Online Hate Index or the Hate Symbols Database.

**Final Project (40%)**: Prepare a presentation and memo describing how your team identified and evaluated hate speech online, and what you learned from the conference. A model will be shown during Week 11.

**III. Course Policies**

**Etiquette**: **Computers and phones may not be used in class, unless required for class-work.** Texting or personal use of social media in professional settings makes students/employees appear unserious and unprofessional.

Drinks and dry food are allowed.

**Communications with the faculty**: We try our best to respond to students, but we cannot guarantee a reply in the first 48 hours. We get hundreds of emails a week - **please help minimize email traffic by not requesting information that can be easily found elsewhere, especially if available on the syllabus.** Substantive questions are best left for office hours.

**Academic Integrity**: Cheating and plagiarism will be punished in accordance with BU’s Academic Conduct Code: [http://www.bu.edu/academics/resources/academic-conduct-code/](http://www.bu.edu/academics/resources/academic-conduct-code/).

**Special Needs and Other Concerns**: If you have any special needs or circumstances, such as a learning disability or health issue, please feel free to speak with me, and we can discuss suitable accommodations. The Office of Disability Services is the only office that may grant accommodations for students. When a student is in need of a short-term release from an academic requirement and there is no clear “disability” involved, students may request clemency from their professors.

Clinicians at Student Health Services generally do not write excuse letters or emails advocating for a student who is requesting release from an academic requirement. When meeting with a student our practice is to encourage students to speak directly with their professors. Clinicians will provide our business card to the students as proof of the visit. If a student is in an acute crisis and requires immediate clinical attention and will miss a class or assignment, an email may be sent to the necessary academic staff from the University Service Center.

**Attendance**: Unexplained absences will result in a significant reduction in your grade. If you plan to miss class for a justifiable reason – religious observance, serious illness, etc. - notify me before you will miss the class, not after.

**Late Work**: Late response paper and final project submissions will have a full-letter grade deducted each day beginning with the assignment deadline.


Another way to interpret grades:
A. The student displays mastery of the material and more. The student performed far beyond my expectations in the course, displaying a grasp of the analytical and empirical material as well as creativity or insight beyond the material itself. The student will be among the top students in the degree program based on her/his performance in this course.

A- The student displays complete mastery of the course material. I was very impressed by the student’s performance, and the student has acquired the analytical, theoretical, and empirical skills to achieve at a very high level in the degree program.

B+ The student fully comprehends the course material. The student met all of my expectations in the course; the student will perform very well in the degree program.

B The student met most of the requirements of the course, but demonstrated weakness in either analytical or empirical skills. The student will perform well in the degree program when those skills are further developed.

B- The student demonstrated weakness in both analytical and empirical skills, but clearly attempted to prepare for evaluated assignments. It is difficult to evaluate whether the student will succeed in the degree program.

C The student demonstrated disregard of the course requirements. Continuing in the degree program is not recommended.

D The student demonstrated blatant negligence and disrespect during the course. Continuing in the degree program is not recommended.

F The student did not attend class or did not perform to a level that I knew they were attending. Continuing in the degree program is not recommended.

V. Topic Schedule:

<table>
<thead>
<tr>
<th>Week</th>
<th>Class Date</th>
<th>Topic</th>
<th>Assignment Due Date</th>
<th>Assignment Due</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>9/14</td>
<td>Comparative approaches to free speech</td>
<td>N/A</td>
<td>None</td>
</tr>
<tr>
<td>2</td>
<td>9/21</td>
<td>Case Study: Incels + Internationally-Linked Right-Wing Groups</td>
<td>Thursday 9/17</td>
<td>Week 2 Reading Response</td>
</tr>
<tr>
<td>3</td>
<td>9/28</td>
<td>Social identity and hate</td>
<td>Thursday 9/24</td>
<td>Week 3 Reading Response</td>
</tr>
<tr>
<td>4</td>
<td>10/5</td>
<td>Why is hate speech so hard to define? -In Class: Group contracts</td>
<td>Thursday 10/1</td>
<td>Week 4 Reading Response</td>
</tr>
<tr>
<td>5</td>
<td>10/13 [Note: This is a Tuesday]</td>
<td>Is hate speech correlated with violence?</td>
<td>Friday 10/9</td>
<td>Week 5 Reading Response</td>
</tr>
<tr>
<td>6</td>
<td>10/19</td>
<td>Case Study: Rwandan Genocide</td>
<td>Thursday 10/15</td>
<td>Week 6 Reading Response</td>
</tr>
</tbody>
</table>
VI. Weekly Readings

XXX = Computer Science Readings

**Week 1 (9/14): Comparative approaches to free speech**

Readings:


**Week 2 (9/21): Case Study: Incels + Internationally-Linked Right-Wing Groups**

**Readings:**


**Week 3 (9/28): Social Identity and Hate**

**Readings:**


Week 4 (10/5): Why is hate speech so hard to define?


Week 5 (10/13) [This is a Tuesday]: Is hate speech correlated with violence?

Readings:


Week 6 (10/19): Case Study: Rwandan Genocide

Readings:


**Week 7 (10/26): What makes content go viral?**

Readings:


**Week 8 (11/2): How hate speech spreads**

Readings:


**Week 9 (11/9): Instigators and their targets**

Readings:


**Week 10 (11/16): Measuring the prevalence of hate speech on specific online platforms**

Readings:


**Week 11 (11/23): Detecting hate speech – does automated identification of hate speech work?**

Readings:


**Week 12 (11/30): Censoring hate speech**

Readings:


Week 13 (12/7): Countering hate speech

Readings:


