Module Description

How should democratic societies respond to the amplification of propaganda, disinformation, and hate speech on digital forums designed to promote free expression? Content moderation - the regulation of the material that users create and disseminate online - has become a routine practice as a response to these new challenges presented by the use of social media platforms. However, such practices raise significant questions linked to democratic accountability and civil liberties.

This course will seek to outline the current regulatory practices being employed to control and restrict our online behaviour, and explain the underlying rationales for how, when, and why these policies are enforced. Topics will include the role of algorithms and curation in ranking content; the promise of labeling, fact-checking, and other interventions designed to counter misinformation; and case studies, such as Facebook’s Oversight Board.
Course Programme

<table>
<thead>
<tr>
<th>Date</th>
<th>Topic</th>
</tr>
</thead>
<tbody>
<tr>
<td>20 Feb</td>
<td>Introduction and Overview</td>
</tr>
<tr>
<td>27 Feb</td>
<td>The Promise of Social Media</td>
</tr>
<tr>
<td>6 Mar</td>
<td>Disinformation and Hate Speech</td>
</tr>
<tr>
<td>13 Mar</td>
<td>The Myth of the Neutral Platform</td>
</tr>
<tr>
<td>20 Mar</td>
<td>To Remove, Label, or Filter? Imperfect Solutions at Scale</td>
</tr>
<tr>
<td>27 Mar</td>
<td>Speech Police: Humans and Machines</td>
</tr>
<tr>
<td>3 Apr</td>
<td>Platform Governance and Democracy</td>
</tr>
<tr>
<td>24 Apr</td>
<td>Across-Platform Comparisons</td>
</tr>
<tr>
<td>8 May</td>
<td>What Platforms are and What They Should Be</td>
</tr>
<tr>
<td>15 May</td>
<td>Group Work and Peer-Review: Policy Solutions</td>
</tr>
<tr>
<td>22 May</td>
<td>No class</td>
</tr>
</tbody>
</table>

Learning Outcomes

Knowledge and Understanding

Following this course students will develop a range of important transferable skills.

Substantive Knowledge

By the end of the course, students should be able to:

- Gain knowledge of the different content moderation practices
- Understand the consequences of platform governance for public discourse
- Understand the current practices employed by different social media platforms
- Explain the underlying rationales for how, when, and why these policies are enforced
- Identify the possible consequences for democracy
- Identify real-life examples of platform governance and its known

Skills (Intellectual and Transferable)

The course will encourage you to:

- Listen carefully and critically to orally-presented arguments.
- Ability to understand the scientific literature, and in particular to identify research puzzles and knowledge gaps.
- Make links between material presented at different times, on different issues.
• Construct persuasive written, and oral arguments supported by evidence, orally
  and in writing.

• Read critically and with a clearly defined purpose.
• Apply your theoretical knowledge to the real world.
• Prepare, articulate and defend answers to set questions.
• Formulate and ask your own questions about course material.

The written work in the course will require you to:

• Select relevant material from lectures, literature, news sources, and the web.
• Understand, analyse and assess that material.
• Produce a sustained, structured and informed answer.
• Write in a concise and cogent style.

Assessment

Grade Component Breakdown

• Continuous assessment - 30%
  – Individual Presentation - 10%
  – Encounter 5%
  – Group Presentation and Peer-Review 15%

• Policy Memo - 70%

Individual Presentation - 10%

Over the course of the term, students will need to individually prepare and present one
“reaction piece”. The short (10 minutes) presentation will engage with at least two of
the readings assigned for that week. While you can spend a few sentences summarizing
the main points, these presentations should primarily analyze or critique the arguments,
identify tensions between them, and suggest constructive ways to synthesize or build on
these works.

Students will be asked to sign up to a specific topic/week they would like to cover in their
presentation at the beginning of the term. Slides should be submitted on OLAT 3 days
BEFORE the Monday seminar.

Each student will be required to present at least once in the term. No more than 1
student is allowed to present in a week, and topics will be assigned on a first come first
served basis.
Encounter - 5%

Over the course of the term, students will need to individually prepare and submit one "encounter". In 300/400 words, students will need to describe what they have "encountered" – a song, film, tweet, news story, book, etc. – provide a link (if applicable) and connect it to the material discussed in class. Your work should demonstrate your ability to:

- Identify important, relevant and recent developments.
- Understand and apply the main theoretical approaches covered in the course to analyse real world issues.
- Clearly describe what was "encountered" – a song, film, tweet, video game, book, conversation with a parent, etc. and connect it to the theoretical discussions developed in class (ESSENTIAL TO PASS!).
- Be able to explain in few words the relevance of the "encounter" to the topics being discussed in the course.

Students will be asked to sign up to a specific topic they would like to cover in their encounter at the beginning of the term. Encounters should be submitted on OLAT up until the day BEFORE the Monday seminar. Encounters CANNOT be submitted for the same topic as the individual presentation.

Each student will be required to submit at least one encounter. No more than 1 student is allowed to submit each week, and topics will be assigned on a first come first served basis.

Group Presentation and Peer-Review - 15%

In preparation for the final assignment, students will have the opportunity to present their proposed policies to the class. Working in pairs, students will have to provide a thorough (15 minutes) presentation of their policy proposal, explaining why their suggested content moderation regime should be adopted by policy-makers.

Moreover, students will have the opportunity to receive feedback from their peers on how to improve their proposals ahead of the final submission. To this purpose, each group will need to submit on OLAT a short (1500 - 2000 words) draft of their policy proposal BY 8 MAY. The drafts will be circulated with the entire class to allow the audience to prepare comments ahead of the presentations. Moreover, a short (10 minutes) Q&A session will be held after each presentation, allowing the speakers to receive feedback from the audience.

All group presentations will take place on 15 May during the seminar. Besides the draft, groups will have to submit their slides before the seminar taking place on 15 May.

Groups will be formed during the first seminar session and will remain unchanged until the end of the term.

Policy Memo - 70%

The final written assignment will take the form of a policy memo outlining and justifying a specific policy proposal or content moderation regime to relevant decision-makers. The
memo will be written in pairs and should be 5000 (10% +/-) words long. **Deadline: 1 June**

**Plagiarism**

Although this should be obvious, plagiarism – copying someone else’s text without acknowledgement or beyond ‘fair use’ quantities – is not allowed. Plagiarism is an issue we take very serious here in UZH.

Please familiarize yourself with the definition of plagiarism on UZH’s website and make sure not to engage in it.

**Late Submission Policy**

All written work must be submitted on or before the due dates.

When an extension is necessary, the student will need to contact our Prüfungsdelegierte Naome Czisch (pruefungen@ipz.uzh.ch) BEFORE THE DEADLINE to apply for extenuating circumstances.

**Grades**

I am very happy to schedule 1:1 meetings to provide students with further feedback when required. However, students should be advised that grades will not be modified after they are released.

**Participation in class**

This course is designed as a seminar. While a short lecture by the instructor will precede the discussion, students are expected to actively participate in class. For that purpose, students will need to follow the assigned readings and come to class ready to engage in dynamic discussions. Moreover, I will sometimes encourage debates by proposing different views and challenging students’ arguments. This is not a means of discouraging opposing views or imposing my own perspective on the students. On the contrary, it is a resource I employ in class to invite students to develop critical thinking and learn to construct arguments to support their own perspectives.

**Essay Grading Rubric**

The following guidelines should be adhered to when writing your final essay:

- **Statement of Purpose/ Focus and Organisation - 40%**
  - The response is fully sustained and consistently and purposefully focused:
    * Claims are clearly stated, focused, and strongly maintained
    * Claims are introduced and communicated appropriately for the purpose, audience, and task
    * Alternate or opposing claims are clearly addressed
  - The response has a clear and effective organisational structure creating unity and completeness:
* A variety of transitional strategies is consistently used to effectively clarify the relationships between and among ideas
* The progression of ideas from beginning to end is logical
* The introduction and conclusion are effective for audience and purpose
* Appropriate sentence structure variety produce strong connection between ideas

Evidence/Elaboration - 40%

- The response provides thorough and convincing support/evidence for the writer’s claim that includes the effective use of sources, facts, and details. The response achieves substantial depth that is specific and relevant:
  * Claims are supported with relevant evidence from credible sources and clear reasoning
  * Use of evidence from sources is smoothly integrated, cited, comprehensive, and concrete
  * A variety of effective argumentative techniques is used
- The response demonstrates strategic use of language to produce clear communication:
  * Precise language clearly and effectively expresses ideas
  * The use of academic and domain-specific vocabulary is clearly appropriate for the audience and purpose

Editing Conventions - 20%

- The response displays adequate command of all grade level and preceding level conventions of writing:
  * Some errors in usage and sentence formation may be present, but no systematic pattern of errors is displayed
  * The use of punctuation, capitalisation, and spelling is adequate

**OLAT**

Please make sure you have access to the module in OLAT as soon as possible. It is the student’s responsibility to make sure that they are signed up to the module correctly and they know how to submit coursework through the appropriate OLAT assignment tab. If you have any issues with OLAT contact the IT Helpdesk to resolve the issue.

Furthermore, module materials such as this syllabus and announcements made outside lectures shall be on OLAT. As such, OLAT is an important communication tool for the module.

**Emails**

I will seek to reply to emails within the following 48 hours. However, this might not always be the case. Additionally, I will not reply to emails during the weekend or after working hours.
Additional Covid-19 Guidelines

Covid-19 continues to pose a threat to our well-being and health. We all need to follow UZH’s guidelines. If you are not feeling well, stay home! I will try to make all relevant materials available to everyone using OLAT: I will share the slides after each session and upload all seminar materials.

Course Reading

Required Readings:

The following texts shall be used extensively throughout the course, so it is recommended that they are purchased:


In addition to these readings, students should keep up to date on current international affairs by reading daily newspapers, or one of the many websites and podcasts devoted to the Global South. This reading is essential as it will allow you to keep up to date with current affairs and identify potential encounter topics. These websites include the following:

- [http://www.foreignaffairs.com](http://www.foreignaffairs.com)
- [http://blogs.lse.ac.uk](http://blogs.lse.ac.uk)
- CCS Podcast - [https://open.spotify.com/show/0PLCDpeA5KyhPE5JObL5S3?si=Q6QiSwuTlqczIHc5YqiXw](https://open.spotify.com/show/0PLCDpeA5KyhPE5JObL5S3?si=Q6QiSwuTlqczIHc5YqiXw)
- Social Media and Politics podcast - [https://socialmediaandpolitics.org/](https://socialmediaandpolitics.org/)
Detailed Course Programme

20 February

Introduction and Overview

Key readings


Further reading


27 February

The Promise of Social Media

Key readings


Further reading


6 March

Disinformation and Hate Speech

Key readings


Further readings


• Shu, K., Bhattacharjee, A., Alatawi, F., Nazer, T. H., Ding, K., Karami, M., & Liu, H. (2020). Combating disinformation in a social media age. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 10(6), e1385.


13 March

The Myth of the Neutral Platform

Key readings


Further reading


20 March

To Remove, Label, or Filter? Imperfect Solutions at Scale

Key readings


Further reading


27 March

Speech Police: Humans and Machines

Key readings


Further reading


3 April

Platform Governance and Democracy

Key readings


Further reading

• Sablosky, J. (2021). Dangerous organizations: Facebook’s content moderation decisions and ethnic visibility in Myanmar. Media, Culture & Society, 43(6), 1017-1042.


24 April

Across-Platform Comparisons

Key readings


Further reading


8 May

What Platforms are and What They Should Be

Key readings


Further reading


15 May

Group Work and Peer-Review: Policy Solutions

22 May

No class