# Course Change Request

## New Course Proposal

Date Submitted: 04/17/24 2:52 pm

## Viewing: CSS 671 : Natural Language Processing for Complex Systems

Last edit: 04/17/24 5:29 pm

Changes proposed by: blaisten

### In Workflow

1. **CDS Chair**
2. **SC Curriculum Committee**
3. SC Assistant Dean
4. Assoc Provost-Graduate
5. Registrar-Courses
6. Banner

### Approval Path

1. 04/17/24 3:16 pm Jason Kinser (jkinser): Approved for CDS Chair

**Are you completing this form on someone else's behalf?**

Yes

**Requestor:**

| Name | Extension | Email |
|---|---|---|
| Estela Blaisten | 31988 | blaisten@gmu.edu |

**Effective Term:** Fall 2024

**Subject Code:** CSS - Computational Social Science

**Course Number:** 671

**Bundled Courses:**

**Is this course replacing another course?** No

**Equivalent Courses:**

**Catalog Title:** Natural Language Processing for Complex Systems

**Banner Title:** NLP for Complex Systems

**Will section titles vary by semester?** No

**Credits:** 3

**Schedule Type:** Lecture

**Hours of Lecture or Seminar per week:** 2.5

**Repeatable:** May only be taken once for credit, limited to 2 attempts (N2)

**Max Allowable Credits:** 6

**Default Grade Mode:** Graduate Regular

**Recommended Prerequisite(s):**
CSS 610

**Recommended Corequisite(s):**

**Required Prerequisite(s) / Corequisite(s) (Updates only):**
CSS 605 or permission of the instructor

**Registrar's Office Use Only - Required Prerequisite(s)/Corequisite(s):**

| And/Or | ( | Course/Test Code | Min Grade/Score | Academic Level | ) | Concurrency? |
|---|---|---|---|---|---|---|
| | | | | | | |

**Registration Restrictions (Updates only):**
NA

**Registrar's Office Use Only - Registration Restrictions:**

**Field(s) of Study:**

**Class(es):**

**Level(s):**

**Degree(s):**

**School(s):**

**Catalog Description:**
This course focuses on the fundamentals of Natural Language Processing (NLP), Natural Language Understanding (NLU) and Large Language Models (LLMs), comparatively or coupled with other computational methods used in modeling communication, such as agent-based modeling, social network

analysis, audio and image processing. It teaches topic modeling, topic2vec, speech recognition, clustering methods, and other relevant methods applied to social science and complex systems.

**Justification:**

What: This course fills a niche in the CSS, PhD instruction by teaching doctoral students a set of specific applications of NLP/NLU and large language models (LLMs) in complex systems applications. Specifically, students in the CSS, PhD are researching various aspects of communication, or use text databases, or apply advanced computational methods without having formal training and access to NLP/NLU/LLMs methods and their cross-methodological use with other data-based approaches, such as machine learning.
Why: The CSS, PhD doctoral students need to become proficient in a range of computational methodologies in order to successfully graduate and competently represent Mason in the workplace. Currently, the graduate curriculum of this doctoral program lacks NLP/NLU/LLM courses that focus on complex systems and on the use of these methods aligned with other computational methods, e.g. machine learning and deep learning. The Computational and Data Science department has assessed that this line of expertise enhances the quality of the graduate students portfolios.

**Does this course cover material which crosses into another department?** No

**Learning Outcomes:**

1. Students will acquire competency with state-of-the art NLP/NLU/LLM methods and applications in current complex systems phenomena, such as information and opinion dynamics, speech recognition problems, social media analysis, news in different foreign languages.
2. Students will reach proficiency in the use of NLP/NLU/LLM methods comparatively or coupled with other methods, such as agent-based models, social network analysis, image processing, audio processing, from both human and cetacean (dolphins and whales) recordings data.
3. Students will improve competency in the analysis of NLP/NLU/LLM applications of complex systems and social science theories with respect to language convergence, origins of language theories, cultural complexity, the future of human languages, communication patterns and scalability in living systems, and communication for space missions.

**Will this course be scheduled as a cross-level cross listed section?** No

**Attach Syllabus**

CSS_671_Syllabus_April10-2024.pdf

**Additional Attachments**

**Staffing:**

Dr. Anamaria Berea, Associate professor, Department of Computational and Data Sciences

**Relationship to Existing Programs:**

The CSS, PhD and the CSI, PhD are administered by the Department of Computational Sciences. It is in the interest of the department to cross-list courses when the course subject is applicable to both doctoral

programs. This is the case between the new course proposal CSS 671 Natural Language Processing for Complex Systems and CSI 671 Natural Language Processing for Complex Systems. The CIM platform does not allow the courses to be equivalent because of the nuances in the prerequisites, banner title, and several mentioning of either one of the two doctorate programs.

**Relationship to Existing Courses:**

Provides intermediate and advanced coding and logical elements that can be used with Agent-Based Models in CSS 605 and CSS 610.

This course parallels, but does not overlap, with the content, methods and applications taught in the following courses: CS 678; AIT 526 and AIT 726.

**Additional Comments:**

This course will be added to the Elective Courses category of the CSS, PhD requisites.

**Reviewer Comments**

Key: 18562

**CSS 671 (cross-listed with CSI 671)**

**NATURAL LANGUAGE PROCESSING FOR COMPLEX SYSTEMS**

## 1. General Information

| | |
|---|---|
| **Instructor**: | Dr. Anamaria Berea, aberea@gmu.edu |
| **Department**: | Computational and Data Sciences |
| **When, where, website:** | TBD |
| **Credits:** | 3 |
| **Recommended prereq:** | **CSS 610 or CSI 690** |
| **Required prerequisites:** | **CSS 605 or CSI 500 or permission of the instructor** |

## 2. Course Description

This course focuses on the fundamentals of Natural Language Processing (NLP), Natural Language Understanding (NLU) and Large Language Models (LLMs), comparatively or coupled with other computational methods used in modeling communication, such as agent-based modeling, social network analysis, audio and image processing. The course teaches topic modeling, topic2vec, speech recognition, clustering methods, and more, applied to social sciences and complex systems.

## 3. Learning Outcomes

- Students will acquire competency with state-of-the art NLP/NLU/LLM methods and applications in current complex systems phenomena, such as information and opinion dynamics, speech recognition problems, social media analysis, news in different foreign languages.
- Students will reach proficiency in the use of NLP/NLU/LLM methods comparatively or coupled with other methods, such as agent-based models, social network analysis, image processing, audio processing, from both human and cetacean (dolphins and whales) recordings data.
- Students will be competent in the analysis of NLP/NLU/LLM applications of complex systems and social science theories with respect to language convergence, origins of language theories, cultural complexity, the future of human languages, communication patterns and scalability in living systems, and communication for space missions.

## 4. Textbooks

Required textbook:

1. Computational Analysis of Communication: Van Atteveldt, Wouter, Damian Trilling, and Carlos Arcíla Calderón. Computational analysis of communication. John Wiley & Sons, 2022. https://cssbook.net

Recommended books:

1. Bird, Steven, Ewan Klein, and Edward Loper. Natural language processing with Python: analyzing text with the natural language toolkit. O'Reilly Media, Inc., 2009. https://tjzhifei.github.io/resources/NLTK.pdf
2. Adger, David. Language unlimited: The science behind our most creative power. Oxford University Press, USA, 2019.
3. Jockers, Matthew L., and Rosamond Thalken. Text analysis with R. Springer International Publishing, 2020.

Additional useful online resources:
https://www.complexityexplorer.org/courses/135-foundations-applications-of-humanities-analytics.

## 5. Technology Requirements

*Hardware:* Students need access to a Windows, Mac, or Linux computer with at least 8 GB of RAM and a fast, reliable broadband Internet connection. An average laptop will be sufficient to address most of the programming needs of each student. For computationally intensive programming projects, students will be instructed how to access open source cloud systems (i.e., Google Colab) and/or the use the servers available in our department.

*Software:* This course will be using primarily Python, and/or secondarily R as programming languages. Prior intermediate level knowledge of these languages is required.

**6. Course outline (Tentative)**

| Week | Topic | What is Due |
|---|---|---|
| 1 | Natural Language Processing and its relationship with information theory | |
| 2 | Fundamentals of text mining and topic modeling | SP |
| 3 | Categorizing and tagging words; text classification and n-gram structures | SP |
| 4 | Extracting information from text, audio and image files | SP |
| 5 | Syntax and Context-Free Grammars with applications to non-human language data | SP |
| 6 | Feature-based grammars and neural network models of language inference | SP |
| 7 | Semantic analysis, Semantic ML and Natural Language Understanding (NLU) in foreign news and social media | SP |
| 8 | Applications of NLP/NLU to underlying complex informational structures in societies | SP |
| 9 | Speech recognition, communication patterns and scalability in living systems | SP |
| 10 | Fundamentals of Large Language Models (Bert, PaLM and GPT systems) | SP |
| 11 | Pattern recognition in complex datasets; language design for space missions | SP |
| 12 | Universals patterns of communication in complex systems | SP |
| 13 | Summary of the semester | SP |
| 14 | Final project presentation | FP |

**7. Grades**

| | |
|---|---|
| Final Project (FP) | **60%** |
| Student presentations (SP) | **40%** |

| Final Mark | Corresponding Grade |
|---|---|
| >99 | A+ |
| 95.1 – 99 | A |
| 90.1 – 95 | A- |
| 87.1 – 90 | B+ |
| 83.1 – 87 | B |
| 80.1 – 83 | B- |
| 70.1 – 80 | C |
| <=70 | F |

**8. Student Presentations and Final Project**

Each student will present once during the semester, at the end of the class, a paper or a model or an algorithm of their choice, but related to the topic of that week.

Final project consists of an original computational project in NLP/NLU/LLM of student choice, that students can work on throughout the semester.

**9. General University Policies**

The course abides by all the Mason policies. A list of important policies is given below. For distribution to students and extramural posting, the syllabus will contain the established requirement of including policies as described in: https://stearnscenter.gmu.edu/knowledge-center/designing-your-syllabus/.

Mason Honor Code (quote from catalog.gmu.edu): "To promote a stronger sense of mutual responsibility, respect, trust, and fairness among all members of the George Mason University community and with the desire for greater academic and personal achievement, we, the student members

of the university community, have set forth this honor code: Student members of the George Mason University community pledge not to cheat, plagiarize, steal, or lie in matters related to academic work."

Email policy: Mason's electronic mail provides any official information to students. Any class materials, assignments, questions, and instructor feedback should use the email Mason email. Students are responsible for maintaining their email account active, working correctly, and should check their content regularly (review details in catalog.gmu.edu).

Plagiarism policy for Internet materials: Copyright rules apply to users of the Internet who employ elements downloaded from Internet sources. Any information in the form of graphics, text, tables, or data accessed electronically and used in homework, presentations, exams, email, reports, must be cited giving credit to the pertaining sources. Even if credit is given, students must obtain permission from any copyrighted source to use any material not created by them. Inserting someone's else material in your work is stealing intellectual property. Including a link to the site URL is currently an appropriate citation.

Student privacy policy: Mason complies with FERPA by protecting the privacy of student records and judiciously evaluating requests for release of information from those records. It is not permitted for faculty to share class progress or grade information with parents/guardians under any circumstances. Student privacy policy: https://registrar.gmu.edu/students/privacy/

Academic integrity: This course embodies the value that we all have differing perspectives and ideas, and we each deserve the opportunity to share our thoughts. Therefore, we will conduct our discussions with respect for those differences. That means, we each have the freedom to express our ideas, but we should also do so keeping in mind that our colleagues deserve to hear differing thoughts in a respectful manner, i.e. we may disagree without being disagreeable. https://oai.gmu.edu/

Students with disabilities: Students with disabilities should contact the Office of Disability Services (ODS). Students requiring special accommodations should inform the instructor the first week of classes. Accommodations may be appropriate for situations that directly affect the student academic performance. ODS requires pertinent medical documentation of a physical, mental health, attention, or other health challenge. https://ds.gmu.edu/.