CS889 W23 InfoVis for AI Explainability

INSTRUCTOR
Dr. Jian Zhao, Assistant Professor, Cheriton School of Computer Science

CONTACT
jianzhao-at-uwaterloo-dot-ca

OFFICE HOURS
By appointment

WHEN AND WHERE
Wednesdays 10:00am-12:50pm, DC2568

Course Communication

MS Teams: for announcements, Q & A, access to class materials, team communication, etc. Download

Zoom: for remote class meetings and office hours when necessary. Download

Announcements

- If you want to apply for a permission number to enroll the course, please fill in this form.
- Please let the instructor know immediately if you have difficulty to access the MS Teams course channels within the first week.

Overview

Recent advances in artificial intelligence (AI), while offering effective means of transforming huge amount data into useful knowledge, generate many challenges in understanding and utilizing the models developed. For example, a deep neural network often operates like a black box, and users do not know when it works, how it works, and why it works (or not). Information Visualization is a

powerful tool to amplify human cognition of abstract data and models with interactive visual representations.

In this course, students will get an overview of state-of-the-art research in visualization methods for AI explainability, and learn how to design, develop, and evaluate visualization techniques for solving real-world data problems. Also, students will gain hands-on experience in prototyping visualization designs. This is a seminar-style course, which covers topics in both theoretical foundations and practical techniques in visualization, such as human perception and cognition, design methods and principles, data representations and transformations, as well as various techniques for visually explaining AI models.

For more details, please see the course information (for marking scheme and graded components) and course schedule (for readings, seminars, exercises, and project deadlines). Also check the resources page for useful materials.

Course Policies

Course Enrollment

Enrollment to the course *after the first week* needs an instructor's approval, since this is a project-based course which requires teams to be formed by the second week.

Participation and Late Panelties

Students are required to attend all the classes, and participation grade will be partially based on how many classes are missed. However, special situations can be accommodated (e.g., academic travel, illness, and emergencies). Students must notify the instructor regarding their absence and provide the necessary justification. Moreover, students must inform the instructor if they have to miss a deadline for such special situations. For other cases, the general policy is that late work will be *deducted 10% of the total marks* per calendar day late. The instructor reserves the right to accept late work or not.

Note due to COVID-19

In the event of a short-term cancellation of in-person classes due to COVID-19 related causes (e.g., an outbreak on campus, the instructor having COVID-19-like symptoms), lecture videos will be uploaded for students to watch remotely. A longer-term cancellation will be handled similarly. Students who cannot attend in-person classes due to self-isolation will still be able to watch recorded lectures.

Academic Integrity

In order to maintain a culture of academic integrity, members of the University of Waterloo community are expected to promote honesty, trust, fairness, respect and responsibility. [Check the Office of Academic Integrity for more information.]

Grievance

A student who believes that a decision affecting some aspect of his/her university life has been unfair or unreasonable may have grounds for initiating a grievance. Read Policy 70, Student Petitions and Grievances, Section 4. When in doubt, please be certain to contact the department's administrative assistant who will provide further assistance.

Discipline

A student is expected to know what constitutes academic integrity to avoid committing an academic offence, and to take responsibility for his/her actions. [Check the Office of Academic Integrity for more information.] A student who is unsure whether an action constitutes an offence, or who needs help in learning how to avoid offences (e.g., plagiarism, cheating) or about "rules" for group work/collaboration should seek guidance from the course instructor, academic advisor, or the undergraduate associate dean. For information on categories of offences and types of penalties, students should refer to Policy 71, Student Discipline. For typical penalties, check Guidelines for the Assessment of Penalties.

Appeals

A decision made or penalty imposed under Policy 70, Student Petitions and Grievances (other than a petition) or Policy 71, Student Discipline may be appealed if there is a ground. A student who believes he/she has a ground for an appeal should refer to Policy 72, Student Appeals.

Note for Students with Disabilities

AccessAbility Services, located in Needles Hall, Room 1401, collaborates with all academic departments to arrange appropriate accommodations for students with disabilities without compromising the academic integrity of the curriculum. If you require academic accommodations to lessen the impact of your disability, please register with AccessAbility Services at the beginning of each academic term.

Intellectual Property

Students should be aware that this course contains the intellectual property of their instructor, TA, and/or the University of Waterloo. Intellectual property includes items such as:

- Lecture content, spoken and written (and any audio/video recording thereof);

- Lecture handouts, presentations, and other materials prepared for the course (e.g., PowerPoint slides);
- Questions or solution sets from various types of assessments (e.g., assignments, quizzes, tests, final exams); and
- Work protected by copyright (e.g., any work authored by the instructor or TA or used by the instructor or TA with permission of the copyright owner).

Course materials and the intellectual property contained therein, are used to enhance a student's educational experience. However, sharing this intellectual property without the intellectual property owner's permission is a violation of intellectual property rights. For this reason, it is necessary to ask the instructor, TA and/or the University of Waterloo for permission before uploading and sharing the intellectual property of others online (e.g., to an online repository).

Permission from an instructor, TA or the University is also necessary before sharing the intellectual property of others from completed courses with students taking the same/similar courses in subsequent terms/years. In many cases, instructors might be happy to allow distribution of certain materials. However, doing so without expressed permission is considered a violation of intellectual property rights.

Please alert the instructor if you become aware of intellectual property belonging to others (past or present) circulating, either through the student body or online. The intellectual property rights owner deserves to know (and may have already given their consent).

CS889 W23

Textbooks

There are no textbooks required for this course. Students are expected to read the lecture notes and attend class. However, the following textbooks are recommended for reference:

- Tamara Munzner, Visualization Analysis and Design, CRC Press, 2014 (Referred as VAD). Free access via the University online library
- Christoph Molnar, Interpretable Machine Learning: A Guide for Making Black Box Models Explainable, Online, 2021.
- Scott Murray, Interactive Data Visualization for the Web, 2nd Ed., O'Reilly, 2017. Free access via the University online library
- Anne-Marie Dufour, D3.js in Action, manning.com, 2022. [Free access online]

Grading

- Participation and Preparation
- Visualization Design Exercises
- Micro Paper Reviews
- Class Seminar
  - Paper Presentation
  - Technique Tutorial
- Class Project
  - Project Proposal
  - Mid-class and Final Presentations
  - Final Paper and Video Demo

The format of the course is mainly a seminar style. It will contain lectures by the instructor, paper discussion seminars by students, and presentations of group projects. Students will be evaluated by the quality of their in-class participation, assignments, paper presentations, and class projects.

This course consists of a mixture of individual and group work. Students will work in teams of *three or four* to complete a term-long class project. A maximum of *seven teams* can be formed in this term due to the capacity. In addition, students will work individually on a set of programming expertise related to visualization and analysis, weekly paper reading and reviews, as well as a

seminar of paper presentation, technique tutorial, and discussion. The philosophy of this seminar based course is *learning by doing and learning together*.

The grading components of the course include:

10%	Participation and preparation (individual)
10%	Visualization design exercises (individual)
15%	Micro Paper Reviews (individual)
20%	Class seminar (group)
45%	Class project (group)

Students' work will be graded based on a *10-point scale with a 0.5-point increment*, summed and rescaled based on the below weights, where in general 0 = not participate, 2 = poor, 5 = fair, 8 = good, 10 = excellent.

Participation and Preparation

Attendance, team building, and in-class activities such as group discussion, paper critiques, and feedback to other students' work. Students are required to attend every class and actively participate in discussions. As a rule of thumb, you are expected to contribute to class discussion at least *two times* each class: ask a question, comment on a topic, clarify a point, etc. Your attendance and participation level will be recorded during the class.

Students must form teams at the beginning of the term. After finalizing a team, designate one person from the team to submit the Project Team Information form (link can be found on MS Teams). In the end of the term, students are required to complete a Team Peer Evaluation form to provide their individual perspectives on the teamwork and their teammates. The responses may be used for adjusting individual marks based on the final team marks.

Visualization Design Exercises

Three (3) programming exercises for creating visualizations using Javascript and/or Python. Each of the visualization design exercises will have one or more objectives that you must write code to solve. You will use popular data visualization and analysis toolkits such as D3 and scikit-learn. In addition to submitting final code, you will need to include a very short solution description (less than *500 words*) explaining what you did, what was hard, and an annotated list of the main resource(s) you used (blog posts, online tutorials, StackOverflow posts, papers, textbooks, etc.). The

Each exercise is graded using the following marking scheme:

5 marks	quality and degree of completion of code implementation
5 marks	quality and clarity of solution description and documentation

Submit your code, data, and solution description in a single zip file, named *[your last name]_[your student ID]_VD[n].zip*, to the corresponding LEARN Dropbox by the posted deadlines.

Micro Paper Reviews

Seven (7) weekly reading assignments of papers provided by the instructor. You will be required to submit micro reviews (less than *500 words* each) for the assigned papers every week with a predefined template. You are also recommended to read other papers on the list to get yourself more familiar with the week's topic. A micro paper review must include four parts:

1. *Summary:* One or two sentences to summarize the paper's key contributions.
2. *Strengths:* Two or three specific strengths of the paper, with rationales.
3. *Weaknesses:* Two or three specific weaknesses of the paper, with rationales.
4. *Future work:* One or two sentences to state any work that could be done to extend or build on this.

You should consider the following aspects when writing your micro reviews:

- Be specific and provide details.
- Do not simply paraphrase the paper content: your goal is to demonstrate your read the paper and thought about it.
- Do not comment on the quality of writing or paper presentation: your goal is to focus on the research described in the paper.
- Use grammatically correct English writing.

Each micro paper review is graded using the following marking scheme:

3 marks	style and grammar
7 marks	quality of content (addressing each aspect well)

Submit your micro paper reviews as PDF files with a predefined template or similar format, named *[your last name]_[your student ID]_MR[n].pdf*, to under the appropriate directory in the shared

folder on MS Teams by the posted deadlines. All micro reviews will be made public to the class to foster discussions.

Class Seminar

One (1) seminar of discussing a designated research topic from the list provided by the instructor.
Each team will lead a seminar on a specific topic during one week of the class, including the presentation and discussion of two related research papers as well as one AI explanation technique:

10%	Paper presentations
5%	Technique tutorial

Students must fill in the Seminar Schedule form (link can be found on MS Teams) at the beginning of the class, to select which topics/weeks they will present for the seminars. All the team members need to contribute to the paper presentations and technique tutorial, although not every student needs to present. Include a task split slide at the beginning or the end of your seminar. The format of the seminar (in terms of the order of the activities) is flexible; for example, you could do the technique tutorial first, or combine the two paper presentations together.

PAPER PRESENTATION

The paper presentation needs to clearly demonstrate the problem, approach, and key results of the paper. More importantly, you should use the paper content and examples as a way to *introduce and address problems beyond the paper*, related to the week's topic and other papers. You could consider the following questions:

- What is the problem area? What is the relevant background? Why do we care?
- What is the idea and what are the main contributions?
- What are the key related work? What are the differences and similarities?
- What is the core approach and results?

In addition, you will be responsible for answering questions regarding the paper and leading a discussion after the presentation. You will have access to all others' micro reviews at least 24 hours before the class, so budget some time to read through them, summarize the responses. You are encouraged to call on individuals in the class during the discussion to expand or justify their responses. You will have to manage the class, by keeping people on topic, encouraging everyone to speak, and making sure the discussion is not dominated by a few people. You should carefully prepare the discussion points, such as:

- What are the strengths and weaknesses of the paper?
- What are potential extensions and future work?
- What critical reflections could we have?

The paper presentation is graded using the following marking scheme:

2 marks	length and delivery (clear speaking, clear slides)
4 marks	paper presentation (more critique than summary with relevant and insightful comments)
4 marks	leading discussion (prepared questions, reading others' summaries, class management)

Each paper presentation should last about *35 minutes*, roughly *20 minutes* for the presentation and *15 minutes* for the discussion. Remember that you do not need to cover every nitty gritty details about the paper, because the whole class also read the paper beforehand; aim for a profound thinking rather than an on-surface description of the paper.

TECHNIQUE TUTORIAL

The technique tutorial needs to clearly introduce the theory and mechanism behind the AI explanation technique, as well as the advantages and disadvantages of the technique. When appropriate, you could relate the technique to the papers in the week. You will have to find secondary sources to gain a deep understanding about the technique. Your objective is to provide sufficient interpretation to give intuition behind how the technique works and to get someone started using it.

The technique tutorial is graded using the following marking scheme:

2 marks	length and delivery (clear speaking, clear slides)
6 marks	quality of content (clear and thorough demonstration of the technique)
2 marks	quality of Q & A (good understanding of the technique)

The technique tutorial should last about *20 minutes* including Q & A. Be prepared for questions from the class.

Class Project

One (1) term-long research project on designing, developing, and evaluating new visualization techniques. Each team will need to complete a class project on a topic related to explainable AI and visualization. The project topic must be approved by the instructor. Through this project, each team will progressively build a final artifact deliverable that can be *a standalone visualization system prototype or an interactive article with several smaller visualization components*. It needs to be something actually running (i.e., not a mockup). Please refer to the project page for ideas and potential datasets/problems. You are encouraged propose a project that is relevant to your own research and background.

The class project will be split into different stages as the following:

5%	Project proposal
10%	Mid-class project presentation
10%	Final project presentation
20%	Final paper and video demo

Paper writings must be in the IEEE TVCG Journal format as well as include the *References* and an *Acknowledgement* that briefly describes what each team member contributes to the overall assignment (e.g., roles and task distributions; not necessarily everyone has to contribute to the writing).

Submit your writing, slides, or video, named *[team member last names or a cool project name]_[deliverable name].[pdf/pptx/mp4]*, to the corresponding LEARN Dropbox by the posted deadlines.

PROJECT PROPOSAL

You should start thinking about your project early. The idea should be non-trivial and have some level of novelty, e.g., designing new visualization techniques, creating a new data processing pipeline for visualization, applying visualizations to a new problem, domain, or dataset. You are encouraged to talk to the instructor before finalizing the topic. You will write a brief proposal (*1 to 2 pages*) that outlines your project idea. The proposal must contain the following parts:

- *Title:* The title should be catchy, short, and descriptive.
- *Abstract:* The abstract should be less than 150 words and summarize your paper, including contributions, approach, and results. You could use the following template: 1) context and background (1 sentence), 2) challenges and problems to solve (1 sentence), 3) your proposed approach and main contributions (1-3 sentences), and 4) expected results (1-2 sentences).

- *Introduction:* The goal of the introduction is to provide the context, deliver the motivation, summarize the paper, basically, why do you do it, and how? You could use this template: 1) provide the background or context of the research and present the unaddressed research problems or questions (1 paragraph), 2) summarize a few key relevant work and why they do not work for your problem (1 paragraph), 3) describe your approach and how it addresses the problem (1-2 paragraphs), and 4) list 1-3 key contributions of the paper (1 paragraph).
- *Proposed Solution:* This section describes the high level technical approach and/or design, languages/libraries/toolkits you plan to use, and how you will address the research questions or problems using the technical approach. Try to be as detailed as possible and include sketches or figures if it makes explaining your approach easier.

The marking scheme is:

3 marks	style (academic writing style, correct grammar, formatting)
7 marks	quality of content (completion of all sections, clarity of description)

MID-CLASS AND FINAL PRESENTATIONS

You will present your work in class, answer questions, and get feedback from other students. You will need to do two presentations: one in the middle and the other in the end of the term, in a style of academic conferences/workshops. Each presentation session is limited to a *10-minute talk* and a *5-minute Q & A*, in total, *15 minutes maximum*.

- *Mid-class Presentation:* You will need to present your progress of the class project, including the *motivation, problem, key related work, and preliminary design idea.* The main purpose of this presentation is to get early feedback from the class and the instructor. Mockups, diagrams, and figures to demonstrate your design idea is strongly encouraged.
- *Final Presentation:* You will need to give a more thorough talk about the your project, including the *motivation, problem, key related work, proposed design, and initial evaluation (optional)* of the project. You can reuse the slides in your mid-class presentation, but should put more weight on showing your solution and results. The final presentation must include a *video or live demo* of a working prototype (i.e., not a mockup), which should have the core functionalities. You can later continue developing your prototype for the final video demo.
- Please include a *contribution slide* in the end to indicate the task distribution among team members.

The marking scheme is:

3 marks	length and delivery (clear speaking, clear slides)

4 marks	project presentation (all aspects included, questions answered)
3 marks	quality of idea or demo/video (complete and clear)

FINAL PAPER AND VIDEO DEMO

You will write an academic paper (*8 to 10 pages*, excluding references) that describes the entire class project in detail. The ideal final paper should be in a quality that is ready to submit to a conference workshop. You will refine the project proposal and extend it by summarizing the work throughout the term. Besides a revised *Introduction*, you need to have the following parts in your final paper:

- *Related Work:* The literature review should include scholarly articles, industry and trade articles (as appropriate), books and magazines, and a review of existing products (where applicable). As a general rule of thumb, you should survey at least *15 related references*. You need to summarize each related work concisely and emphasize the differences or novelty of your idea. You want your literature survey to: 1) Be comprehensive in breadth and depth and related theories have been investigated; 2) Point out aspects of the problem space that have not been solved or examined in detail; and 3) Distinguish your work from the state of the art and emphasize the novelty.
- *Proposed Design:* You will extend the *Proposed Solution* of your proposal with more details including the following aspects. You are encouraged to provide *diagrams, sketches, and mockups* in the paper to demonstrate your solution.
  - *System architecture* describes how you process the raw data to extract useful information for your visualization to solve the problem as well as provides a high-level overview of your system data flow and key components.
  - *Visual representations* demonstrate your visual design based on the processed data, for example, the main views of your system, how the extracted information is encoded visually, etc. Basically, it describes the static aspect of your approach.
  - *User interactions* illustrates the interactions on your visualization such as how a user manipulates the views and visual representations to gain knowledge. Basically, it describes the dynamic aspect of your approach.
- *Example Use Cases/ Initial Evaluation:* Example use cases demonstrate a *detailed walkthrough* of your approach that is applied to real-world datasets for solving the proposed problem. This normally includes steps of interacting with the system and obtaining information for decision making. Figures illustrating the key steps are necessary. On the other hand, an initial evaluation can be *interviews or trial studies* with real target users. This often reports the procedure and results of this evaluation, including who the users are, how you perform the evaluation, what

the users' tasks are, as well as the users' actions, impression, and feedback on your system or technique.

- *Discussion:* This reflects on what was done, implications for designers and researchers, some issues or problems encountered in the evaluation, and limitations of the approach. This section is your chance to reflect on any lessons learned.
- *Future Work:* This presents advice and directions for others doing work similar to your project, suggesting how things could be improved or better.

In addition, you will need to submit a final video (less than *5 minutes*) to demonstrate your system. The ideal video should showcase a *usage scenario* of the system while demonstrating the key features. Voice narration is strongly recommended. Ensure your video is encoded with MPEG-4 using H.264 (check Handbrake for encoding/re-encoding your video). The video could be made public on YouTube or the class website.

The marking scheme is:

1 mark	style (academic writing style, correct grammar, formatting)
7 marks	quality of content (completion of all sections, clarity of description, sufficient details)
2 marks	quality of video (compete and clear)
