SAINT LOUIS UNIVERSITY.

Program-Level Assessment: Annual Report

Program Name (no acronyms): Artificial Intelligence	Department: Computer Science
Degree or Certificate Level: MS	College/School: School of Science and Engineering
Date (Month/Year):	Assessment Contact: Erin Chambers

In what year was the data upon which this report is based collected? AY 2022-2023

In what year was the program's assessment plan most recently reviewed/updated? 2018

Is this program accredited by an external program/disciplinary/specialized accrediting organization or subject to state/licensure requirements? No.

If yes, please share how this affects the program's assessment process (e.g., number of learning outcomes assessed, mandated exams or other assessment methods, schedule or timing of assessment, etc.):

1. Student Learning Outcomes

Which of the program's student learning outcomes were assessed in this annual assessment cycle? (Please provide the complete list of the program's learning outcome statements and **bold** the SLOs assessed in this cycle.)

This year, assessment was targeted at the following outcome:

PLO 5 - Make informed and ethical decisions regarding the impact of artificial intelligence technologies.

2. Assessment Methods: Artifacts of Student Learning

Which artifacts of student learning were used to determine if students achieved the outcome(s)? Please describe the artifacts in detail, identify the course(s) in which they were collected, and if they are from program majors/graduates and/or other students. Clarify if any such courses were offered a) online, b) at the Madrid campus, or c) at any other off-campus location.

CSCI 5050: Students were asked to submit a final paper on a topic in ethically responsible computing. The paper was required to be 2000 to 2500 words long. The audience was graduate students, and the assessment was scaffolded by a 10-minute presentation and an individual meeting with the instructor. Students were assessed on the focus and the depth of their discussion as well as their ability to integrate ethical perspectives from their own research or course work. The rationale for the rubric criteria Topic Focus and Depth of Discussion is a direct result of LLMs like ChatGPT becoming widely available. LLMs are terrific at superficial and broad discussions but some less proficient students do not prompt for depth or detail.

At the end of Fall 2022, many students were leaning too heavily on ChatGPT and other LLMs to the detriment of their final paper grades. Another assessment was introduced to increase both accountability and the chance to demonstrate bare knowledge of issues in socially responsible AI and software development. A simple multiple-choice section on an in-class exam seems to both test for an initial awareness of the impact of AI and other computing technologies and scaffolds for selection of paper and presentation topics. The required textbook has plenty of examples and is a good resource for initial discussion of the relevant issues.

3. Assessment Methods: Evaluation Process

What process was used to evaluate the artifacts of student learning, and by whom? Please identify the tools(s) (e.g., a rubric) used in the process and **include them in/with this report document** (please do not just refer to the assessment plan).

The rubric used to assess the final paper is in Appendix A of this document. The first 3 criteria are particularly relevant for determining their PLO 5-related skills, but overall student performance is shared since picking topics and sources is also relevant to their ability to make informed decisions regarding ethically controversial issues.

The multiple-choice exam (introduced in Spring 23 is designed to incentivize bare knowledge of the central issues in socially responsible technology design and implementation. Both historical and recent examples of problematic issues in technology are discussed in the book and integrated into the multiple-choice question bank. The book also introduces legal and professional issues that are germane to artificial intelligence in the professional world.

4. Data/Results

What were the results of the assessment of the learning outcome(s)? Please be specific. Does achievement differ by teaching modality (e.g., online vs. face-to-face) or on-ground location (e.g., STL campus, Madrid campus, other off-campus site)?

	Final Paper (F22&S23)			**Data covers all students in MS-AI, MS-CS, and MS-SE programs.**
Fail	Failure/Novice	9	5.88%	
D/C Range	Apprentice	63	41.18%	
B	Proficient	46	30.07%	
A	Full-Marks	35	22.88%	
	TOTAL	153	100.00%	
	Midterm Multiple Choice (S23)			**Data covers all students in MS-AI, MS-CS, and MS-SE programs**
D or Below	Failure/Novice	22	32%	
C	Apprentice	20	29%	
B	Proficient	25	36%	
A	Excellent	2	3%	

5. Findings: Interpretations & Conclusions

What have you learned from these results? What does the data tell you? Address both a) learning gaps and possible curricular or pedagogical remedies, and b) strengths of curriculum and pedagogy.

A significant number of students fail to achieve proficiency for the lowest levels of Bloom's Taxonomy in the newly introduced multiple-choice test. While the data indicates acceptable scores for the final paper. Many

students are merely regurgitating AI-generated answers, and this likely accounts for the papers that fall within the D/C range.

For the multiple-choice parts, students are not reading the book or are unable to read the book because of language proficiency issues. Part of the reason that the textbook was introduced was to lower the reading level required to reflect on the issues in AI and technology. The course is designed to lower some of the stakes of these assessments by frequent formative in-class activities, and the paper and presentation are scaffolded via multiple assignments like topic choice and bibliography development workshops and check-ins.

6. Closing the Loop: Dissemination and Use of Current Assessment Findings

A. When and how did your program faculty share and discuss the results and findings from this cycle of assessment?

The faculty discussed these assessment issues and saw pilot results from the Summer 2023 semester where the final paper was replaced with a presentation as the primary summative assessment. Given the issues surrounding academic integrity and LLMs, no faculty member objected to this change, so this will be implemented more strategically in the Fall of 2023.

There are certainly equity issues surrounding the assessment of many of these students, and many of the students take every assignment as equally high stakes, so continued academic integrity issues are likely. The introduction of the textbook, while initially an attempt to increase equity, may have backfired. The textbook though reasonably priced may create a financial strain or further exacerbate language comprehension issues. Alternative textbooks will be researched, but it doesn't appear that any Open Access materials fit the needs of the course. It is likely that the minimum TOEFL scores will have to be adjusted to account for this disparity as well.

B. How specifically have you decided to use these findings to improve teaching and learning in your program? For example, perhaps you've initiated one or more of the following:

Changes to the Curriculum or Pedagogies	• Course content • Teaching techniques • Improvements in technology • Prerequisites	• Course sequence • New courses • Deletion of courses • Changes in frequency or scheduling of course offerings
Changes to the Assessment Plan	• Student learning outcomes • Artifacts of student learning • Evaluation process	• Evaluation tools (e.g., rubrics) • Data collection methods • Frequency of data collection

Please describe the actions you are taking as a result of these findings.

The summative assessment will be changed to a 10-minute presentation and further scaffolding will be introduced to de-incentivize the use of plagiarism software like ChatGPT. The assessment will still focus on the depth and cohesiveness of the presentation, so they can demonstrate a mastery of the ethical issues.

The multiple-choice assessment will be revised based on some research presented at the Culturally Responsive Teaching Institute. The questions will be simplified, but they will still be based on the textbook. The. Assessment Committee will research alternative more accessible material from both a financial and a language-proficiency textbook, but there is a tension between accessibility and the goals of the course as a graduate-level course.

If no changes are being made, please explain why.

7. Closing the Loop: Review of Previous Assessment Findings and Changes

A. What is at least one change your program has implemented in recent years **as a result of previous assessment data**?

The prior round of assessment for this program involved software engineering, with overall positive scores. However, the rapid growth led faculty to be concerns about the introductory software engineering course. As a result, we have officially expanded foundational offerings to include CSCI 5020, which is under review, that will allow incoming MS students to remain stronger in this area, building up software engineering skills more gradually. We have also listed several advanced topics course, to grow the advanced material offered.

While not formally assessed in the prior round, the AI-area faculty also met to discuss the growth of interest in this area, as well as revisiting content in different courses. As a result, fall 2023 saw a major restricting of the content in the AI track classes, with a new "applied ML" course being offered as well as new electives related to computer vision and deep learning. The assessment plan will need to be revised in the coming year to include some of this content.

B. How has the change/have these changes identified in 7A been assessed?

N/A

C. What were the findings of the assessment?

N/A

D. How do you plan to (continue to) use this information moving forward?

The other notable issue from the prior round, while not directly related to 7A, was the lack of completed assessment in adjunct-taught courses. Faculty in the software engineering track have taken the lead in compiling a plan for common assessed content in any course in the required sequence.

IMPORTANT: Please submit any assessment tools (e.g., artifact prompts, rubrics) with this report as separate attachments or copied and pasted/appended into this Word document. Please do not just refer to the assessment plan; the report should serve as a stand-alone document. Thank you.

Appendix A

Paper Rubric

Paper Rubric

Criteria	Ratings					Pts
This criterion is linked to a Learning Outcome INTEGRATION OF KNOWLEDGE See the PDF of the Rubric from Cornell for more information, but this rubric contains specific problems for the course.	1.5 pts Full Marks The insights are more than superficial, and they demonstrate an understanding of the principles and issues in socially responsible computing.	1.2 pts Proficient This paper probably could have been written by ChatGPT, but some editing was necessary. Words like "Transparency" and "Privacy" appear, but they aren't always clearly defined.	0.8 pts Apprentice This paper doesn't demonstrate any ability to apply concepts or define them. It is repetitive because it also lacks focus as well.	0.4 pts Novice		1.5 pts
This criterion is linked to a Learning Outcome TOPIC FOCUS	2.5 pts Full Marks The paper has a clear direction and attempts to discuss 1 issue in a thorough manner.	2 pts Proficient There is a focus, but the author doesn't demonstrate any clear perspective on the issue.	1.33 pts Apprentice	0.67 pts Novice		2.5 pts
This criterion is linked to a Learning Outcome Depth of Discussion	2.5 pts Full Marks The paper brings together sources, examples, and reflection.	2 pts Proficient The paper does move from general information to specific information, but real interaction with the specific sources and examples is SOMETIMES lacking.	1.33 pts Apprentice	0.67 pts Novice		2.5 pts
This criterion is linked to a Learning Outcome Cohesiveness	1 pts Full Marks Ties together information from all sources. Paper flows from one issue to the next without the need for headings. Author's writing demonstrates an understanding of the relationship among material obtained from all sources.	1 pts Proficient The paper quickly turns into lists of relevant but fully general information.	0.53 pts Apprentice This paper appears to be a concatenation of lists.	0.27 pts Novice		1 pts
This criterion is linked to a Learning Outcome Spelling and Grammar	1 pts Full Marks	0.8 pts Proficient	0.55 pts Novice			1 pts
This criterion is linked to a Learning Outcome Sources and Citations Uses IEEE citation style, and the paper has at least 5 relevant academic sources	1.5 pts Full Marks	1.2 pts Proficient You probably neglected to cite things in the body of the text.	0.8 pts Apprentice Where are the citations? I can't tell that your bibliography is relevant to the research.	0.41 pts Novice		1.5 pts

Paper Rubric

Criteria	Ratings	Pts

Total Points: 10