AI Human Rights Impact Assessment

for Students

Purpose: To provide a structured guide for students wishing to assess an AI application that they are considering for use in course work.

Quick Start Guide

Get warmed up by considering this Quick Start guide:

Review the full guide (page 4) for more in-depth questions, supporting information and examples.

AI Tool Assessment Select questions from the in-depth guide. Numbering reflects their location within the guide.	**Example: AI pizza ordering and delivery application** 
Q1: What is the primary function of the AI application?	**What is the primary function of the AI application?** • To streamline the pizza ordering and delivery process.

Q3: Why is this AI application needed or beneficial?
☐ Provides me with personalized learning
☐ Improves accessibility for me
☐ Provides me with 24/7 support
☐ Deepens my learning
☐ Supports my need for additional learning modalities
☐ Other
	Why is this AI application needed or beneficial?
	• Provides customers
	  ○ with personalized pizza recommendations
	  ○ improved customer accessibility
	  ○ 24/7 support
	• Frees up restaurant resources by automating the ordering process
	• Reduces wait times and errors in orders.

Q5-16: What potential harm could arise from using this AI application? ☐ Provide me with inaccurate/biased information ☐ Share my private Information. ☐ Share my intellectual property ☐ Unethical use of others intellectual property ☐ Unduly persuades or influences me/others. ☐ Other	**What potential harm could arise for customers?** • Providing customers with inaccurate order details or delivery times. • Recommending certain pizzas due to biased algorithms. • Sharing customers' private information: • Not considering dietary restrictions or preferences. • Limited accessibility for those without internet access • Unauthorized use of customer reviews or feedback.
Q5-16: How can potential harm be mitigated? ☐ Be aware and monitor for hallucinations. ☐ Be aware and monitor for biased AI outputs. ☐ Determine how an AI application protects my personal information. ☐ Refrain from sharing personal information. ☐ Include an AI disclosure statement on my assignment. ☐ Other	**How can potential harm be mitigated?** • Regularly review and update algorithms to ensure unbiased recommendations • Ensure customer information is protected. • Ensure staff are available to handle complex orders. • Regularly gather and act on customer feedback. • Ensure the application is user-friendly and accessible to all customers
Q22: What alternatives exist? ☐ Variations on this AI application ☐ Other AI applications ☐ Non-AI applications ☐ Other	**What alternatives exist?** • Different AI models or algorithms for order processing. • Non-AI applications: Traditional online ordering systems or phone-based orders. • Other: Manual order processing and delivery.

AI Human Rights Impact Assessment for Students: In Depth Guide

Artificial intelligence tools are everywhere! New applications are being released daily¸ but it can be hard to tell the good from the bad, or whether those distinctions even exist. AI has the potential to significantly impact our learning both positively and negatively, depending on how we choose to use it. When selected carefully and used thoughtfully, AI tools can help us engage more deeply with learning materials, enriching our learning experience, and fostering a deeper understanding of the subject matter. The questions in this guide are designed to help you think critically about the AI tool you are considering using for your course project. This will enable you to make an informed decision and explain that decision to others if needed.

STEP 1: Define the AI Tool's Purpose

Describe the AI tool and what it will do. How does it align with the learning goals of the assignment?

Q1
What is the name and primary function of the AI application?

Q2
Which of the assignment's learning outcomes will the AI application help you achieve?

STEP 2: Identify Potential Benefits

Consider how the AI tool may benefit you and your learning experience.

Q3
Why is this AI application needed or beneficial? (See the examples in 2.1 below)

2.1 Examples of potential benefits:

- **Personalized learning:** AI can adapt learning materials to meet your specific needs depending on your previous knowledge and experience. It can help make up for learning gaps resulting from the Covid-19 pandemic or other factors unique to your past experiences.
- **Accessibility:** AI can make education more accessible to those with disabilities through tools like speech-to-text, text-to-speech, and other assistive technologies.
- **24/7 support:** AI tutors and learning coaches can offer support outside of regular school hours, providing help when and where you need it.
- **Resource allocation:** When used to streamline routine tasks, AI tools can allow you to focus more time on learning.
- **Modality:** By increasing learning modalities available (text, audio, visual, interactive), AI can support Universal Design for Learning principles and create a more inclusive learning environment.

Q4
How will the AI application enhance your learning?

STEP 3: Identify Potential Harms

Consider potential harms or risks associated with the AI tool.

Review the potential harms listed in 3.1 below. Questions 5-16 examine these in more detail and help you plan to assess and mitigate each risk effectively.

3.1 Examples of potential harms:

- **Providing you with inaccurate** information can affect your academic performance and future understanding of the subject.
- **Providing you with biased information** can lead to a distorted view of the subject matter, which can perpetuate stereotypes and misinformation.
- **Sharing your private Information** is a breach of privacy which can lead to identity theft, cyberbullying, or other forms of exploitation.
- **Sharing of intellectual property –** Sharing your original ideas may make them available to others and jeopardize the originality or ownership of your work. Sharing other's original ideas (those of your instructor or classmates) could violate University of Waterloo Policy 71 and may be considered academic misconduct.
- **Unduly persuading or influencing opinions or actions**: Increasingly realistic AI applications can lead us to attribute inappropriate levels of trust in the technology and potentially undermine our capacity to think critically and exercise independent judgement.

Hallucinations and Inaccurate Information

Since AI does not actually understand language but simply predicts the next word in a sentence based on patterns it has learned from large amounts of text data, it has no way of determining whether its outputs are accurate. The AI model may produce flawed results if the data used to train it contains mistakes or if the training data is too specialized and the question asked falls outside that area of specialization. In addition, AI may struggle to interpret the nuances of human language.

Use Q5 and Q6 along with the examples in 3.2 below to create a plan for identifying and mitigating inaccuracies in the AI application you are hoping to use.

Q5
How can you determine the accuracy of AI outputs you receive?

Q6
What steps can you take to prevent inaccuracies from negatively impacting you and/or your assignment?

3.2 Sample Strategies for Identifying Inaccurate Outputs:

- **Cross-check information** with reliable sources such as trusted websites, textbooks, or academic papers to ensure accuracy.
- **Test with known data:** Use the AI tool with data or questions for which you already know the answers to confirm that it is producing correct and consistent results.
- **Look for logical consistency:** Evaluate whether the AI's responses make logical sense. Sometimes, AI can generate outputs that sound plausible but when examined closely, they are actually nonsensical or contradictory. For example, if an AI tool gives you a summary of a book, check if the summary aligns with the actual content of the book.
- **Use multiple prompts** to ask the same question in different ways or use different AI tools to see if the AI provides consistent answers.

Biased Outputs

AI models are developed and trained by individuals who have their own innate biases. The data used to train AI systems will reflect pre-existing societal biases as well. Consequently, when AI is used without proper assessment, monitoring, or mitigation strategies, AI can exacerbate these biases. It is important to monitor an AI application to determine whether it generates biased output that may negatively impact you or your assignment and take steps to mitigate identified biases.

Use Q7 and Q8 along with the in examples 3.3 below to create a plan for identifying and mitigating biases in the AI applications you are considering using.

Q7
How can you determine whether the AI application produces biased or discriminatory output?

Q8
What steps can you take to minimize risks that this will negatively impact you and/or your assignment?

3.3 Sample Strategies for Identifying Biased Outputs:

Check for Stereotypes: Look for any stereotypical representations in the AI's responses. For example, if the AI application is generating content about different professions, ensure it doesn't reinforce gender or racial stereotypes.

Use Diverse Prompts: Ask the AI the same question in different ways, including prompts that reflect diverse perspectives. This helps identify if the AI consistently favors certain viewpoints or if it can handle a variety of perspectives fairly.

Seek Feedback from Others: Share the AI's outputs with peers or mentors from diverse backgrounds and ask for their feedback on potential biases. This can provide additional insights and help identify biases you might have missed.

Privacy

AI tools often require access to large amounts of data to function effectively, which can include personal and sensitive information. This raises significant privacy concerns. The data collected by AI systems can be used in ways you did not intend or consent to, potentially compromising your privacy. Monitoring how your data is collected, stored, and used can help you mitigate privacy risks.

Use Q9 and Q10 along with the examples in 3.4 below to create a plan for identifying and addressing privacy concerns in the AI applications you are considering.

Q9
How can you determine whether the AI application compromises your privacy?

Q10
What steps can you take to minimize privacy-related risks?

3.4 Sample Strategies for Protecting Your Privacy When Using AI:

- **Review Privacy Policies:** Before using an AI tool, read its privacy policy to understand how the data you enter data is being stored and used.
- **Limit Personal Information:** Avoid sharing unnecessary personal information with AI tools. Only provide the minimum amount of data required for the tool to function effectively.
- **Use Anonymized Data:** Whenever possible, use anonymized or pseudonymized data. This means removing or masking any personally identifiable information before inputting it into the AI tool.
- **Enable Privacy Settings:** Check if the AI tool offers privacy settings or options to control data sharing. Enable these settings to restrict access to your data and enhance your privacy.
- **Be Cautious with Sensitive Data:** Avoid using AI tools for highly sensitive or confidential information e.g., financial, or medical information.

Intellectual Property

AI tools can also raise significant intellectual property concerns. These systems often rely on vast datasets that may include copyrighted material, and the content they produce can sometimes inadvertently infringe on intellectual property rights. To address these concerns, avoid using AI as a content producer. Instead, use it to enhance your own creative processes. Additionally, if you choose to share your (and others') original work with an AI application, understand that it could become part of that AI system's training data, putting that intellectual property at risk.

Note: Sharing the intellectual property of your instructor (slides, assignments, course materials, etc.) and/or your classmates may violate University of Waterloo Policy 71 and may be considered academic misconduct.

Use Q11 and Q12 along with the examples in 3.5 below to create a plan for identifying and mitigating intellectual property concerns in the AI applications you are considering.

Q11
How can you determine whether the AI application protects your and/or others' intellectual property?

Q12
What steps can you take to minimize the risk that intellectual property will be shared?

3.5 Sample Strategies for Protecting Your Intellectual Property When Using AI:

- **Avoid Using AI as a Content Producer:** Instead, use it to enhance your own creative processes.
- **Understand Terms of Service:** Before using an AI tool, carefully read its terms of service and privacy policy. Look for explicit statements about how your data will be used, particularly whether it will be used to train the AI model. If this information is not readily available, contact the provider for clarification.
- **Use Opt-Out Options:** Some AI tools offer options to opt out of data sharing for training purposes. Make sure to enable these settings if available.
- **Limit Sharing:** Be cautious about sharing your intellectual property with AI tools, especially if the tools are cloud-based or involve third-party services. Only share what is necessary and avoid uploading entire works if possible.

Undue Influence

As AI technologies become more complex and provide more sophisticated responses, their ability to persuade and influence will continue to grow. This is especially true when AI presents itself in the form of a realistic image, video, or avatar and when it presents itself as an expert, teacher, or mentor. Undue influence may subtly shape users' decisions and actions. This can happen when AI recommendations or outputs are taken at face value without critical

evaluation. For instance, if an AI tool suggests a particular approach to solving a problem, users might follow it without considering alternative methods, potentially limiting their creativity and independent thinking. Whether you are using or developing an AI tool, you need to be mindful of this risk to yourself and others. Use Q13 and Q14 along with the examples in 3.6 below to help you do so.

Q13
How can you determine whether the AI application might exert undue influence on you or others?

Q14
What steps can you take to minimize this risk?

3.6 Sample Strategies for Ensuring AI Does Not Unduly Persuade or Influence:

- **Think critically:** Question the information and recommendations provided by the AI and consider alternative viewpoints before forming an opinion or making a decision.
- **Set clear boundaries:** Be aware of the scope and limitations of the AI tool you are using. Understand that AI is a tool to assist you, not to make decisions for you.
- **Seek human input** when making important decisions. Consult with peers, mentors, or experts in the field.
- **Use multiple AI tools** developed with different training data to get a broader range of opinions and recommendations and avoid being unduly influenced by a single AI system.
- **Stay informed** about how AI works, including its strengths and limitations. Understanding the underlying technology can help you use AI more effectively and critically.

Other Potential Harms Specific to my AI Application

Q15
What other harms might be associated with this specific AI application?

Q16
How can these potential harms be mitigated?

STEP 4: Weight the Pros and Cons and Consider Alternatives

Consider whether this is the right technology to use or whether better alternative exist

After evaluating the potential harms associated with a specific AI application and identifying ways to mitigate these risks, it is important to consider whether alternate technologies are available that will better meet student needs. When considering alternatives, include AI and non-AI applications.

PROS	
CONS	

Q17
What alternatives exist?

STEP 5: Consider the Context

Consider the context in which an AI tool operates and the population it affects in order to assess whether it is likely to impact human rights significantly.

An educational context becomes high risk for human rights issues when AI tools are used to make decisions, assign a grade, or draw conclusions about a student. In these cases, there is a much higher potential for discrimination, inequality, and privacy violations[12]. Consider the following questions:

Q18	**Does the AI application make a decision, or provide information or a score that may influence a high stakes decision?** AI tools used to inform high-stakes decisions (e.g., grading, admission to a program, loan approval, funding eligibility) can have significant long-term impacts on an individual's future. If the AI system is biased or inaccurate, it could unfairly disadvantage certain groups and also leave the decision-maker in a precarious legal position.	**YES** ☐	**NO** ☐
Q19	**Does the AI application analyze individuals' face, facial expressions, fingerprints, voice, vocal patterns, movements, or behavior for the purpose of making predictions or drawing conclusions about the person?** AI applications that monitor individual's online activities or physical behaviors can infringe on their privacy rights and create a stressful, intrusive environment.	**YES** ☐	**NO** ☐
Q20	**Does the AI application make decisions on topics that impact vulnerable or protected populations or groups?** Could decisions made by the AI application negatively impact individuals in a protected category (age, race, sex, religion, disability, etc.) or those in a vulnerable circumstance (unwell, unemployed, or unhoused)? For example, an AI application that makes decisions impacting those with disabilities may fail to accommodate their specific needs or perpetuate existing biases.	**YES** ☐	**NO** ☐

Q21	**Does the AI present itself as a teacher, expert, mentor, or authority?** AI agents, especially embodied AI agents, have the potential to influence and persuade[13, 14]. Risks associated with this ability are higher when the agent is designed to instruct, direct, or mentor students.	**YES** ☐	**NO** ☐
Q22	**Does the AI application impersonate a real person (celebrity, influencer, medical or academic expert, politician, etc.)?** AI agents that impersonate real-life individuals not only violate the human rights of the person being impersonated but are deceptive and mislead those using the AI application.	**YES** ☐	**NO** ☐

Answering "yes" to any of Questions 18-22, indicates that AI use in this context may be at high risk for human rights concerns.

Note that the European Union (EU) AI Act describes all educational use of AI as high risk and outlines specific requirements for educational AI use in "Ethical Guidelines on the Use of Artificial Intelligence (AI) and Data in Teaching and Learning for Educators"[15]. Be sure to consult these guidelines if they are relevant to your context.

STEP 6: Ensure Transparency

Clearly communicate how AI informed your assignment

Just as you would want to know if your instructor was using AI to generate coursework or grade your assignment, it is important that you communicate to your instructor how AI informed your work. The Artificial Intelligence Disclosure (AID) Framework has been developed for this purpose by Kari Weaver, the University of Waterloo's Learning, Teaching, and Instructional Design Librarian. Learn how to use it and then use it regularly to ensure transparency.

Q23
Draft your Artificial Intelligence Disclosure Statement (See a sample AID Statement on p. 15 below).

STEP 7: Maintain Human Oversight: Review, and Revise Regularly

Review AI use and outputs regularly, and address concerns that arise.

Human-centered use of AI emphasizes the importance of designing and deploying AI systems that prioritize human values, needs, and rights. This approach ensures that AI tools are developed and used in ways that enhance human well-being and respect ethical principles.

As a student, this means using AI to support and augment your learning rather than replace critical thinking and creativity. It involves choosing to use AI tools responsibly, in ways that promote fairness, transparency, and the protection of human rights.

Since AI tools are rapidly evolving, it is also important to regularly review the questions in this guide and revise your choice to use or not use a specific tool based on new information that may become available.

Additional Resources

Online Learning | Ontario Human Rights Commission - Nine eLearning modules pertaining to human rights in Ontario.

European Union (EU) AI Act aims to foster the development of responsible and trustworthy AI in Europe and around the world. The Ethical Guidelines on the Use of Artificial Intelligence (AI) and Data in Teaching and Learning for Educators support achievement of these goals.

Framework for Accessible and Equitable Artificial Intelligence (AI) in Education by Jess Mitchell, Lorna Lo, Joseph Scheuhammer, Jutta Treviranus and Rachel Spence – A practical guide focusing on equity and accessibility of AI in educational settings.

Teaching AI Ethics: The Series – Leon Furze - Ten excellent blog posts pertaining to teaching AI ethics in an education context.

The Artificial Intelligence Disclosure (AID) Framework by Kari Weaver guides transparent attribution of the use of artificial intelligence in research or educational work

An Inquirer's Guide to Ethics in AI - Matthew Silk and Ian MacDonald explore the ethical dimensions of artificial intelligence through an interdisciplinary lens, covering themes such as bias, transparency, privacy, and the ethical use of AI.

The AI Consciousness Illusion: Why We Think AI Has Feelings - Deep Learning Dialogues
Dr. Clara Colombatto shares her 2024 study on public perceptions of AI consciousness, revealing that 67% of Americans attribute some level of consciousness to LLMs, highlighting that frequent use increases this perception.

AI through an EDI lens - The need for human-centered design – AI Dialogues - Soroush Sabbaghan, Director of EDI at the University of Calgary's Werklund School of Education, talks

about engaging with generative AI through an EDI lens, highlighting the importance of human-centered design to mitigate bias, empower educators, and maintain transparency.

The Ontario Human Rights Code | Ontario Human Rights Commission prohibits actions that discriminate against people based on any of fourteen protected grounds in any of five protected social areas.

Artificial Intelligence Disclosure Statement

Artificial Intelligence Tool: Microsoft Copilot (University of Waterloo institutional instance); Conceptualization: Microsoft Copilot was used to suggest previously unidentified benefits, risks, and mitigation strategies; Writing – Review & Editing: Microsoft Copilot was used to generate synonyms, refine wording, and improve flow.

Acknowledgements

In the AIHRIA for Educators, I advocate for seeking interdisciplinary input when selecting or developing AI applications for education, so it is only fitting that I did the same as I developed this tool. Thanks to the interdisciplinary experts who shared their thoughts and provided feedback:

- Christine Zaza, PhD: Online Learning Consultant, Centre for Extended Learning, University of Waterloo

Cite as:

Long, L. (2024). AI Human Rights Impact Assessment for Students. https://uwaterloo.ca/conflict-management-human-rights/ai-human-rights-impact-assessment-tools

AI Human Rights Impact Assessment

Summative Worksheet

Submit this worksheet with your assignment.

Assignment Title:

Course:

AI application used:

1. What is the purpose of the AI application you have chosen to use in your assignment?

2. How does its use support the learning goals of your assignment and enhance your learning?

3. Explain why the context in which you will use the AI application will or will not be high risk for human rights issues.

4. What potential harms have you identified as being associated with this AI application and what steps have you taken to mitigate them? Be sure to address inaccuracies, bias, privacy, and intellectual property concerns along with other issues specific to your chosen AI application.

5. The Artificial Intelligence Disclosure (AID) statement ensures transparency and explains how you have used the AI application in your assignment. Write your AID statement below.