



## COURSE UNIT DESCRIPTION

Course Unit Title	Code
<b>WORKSHOP: AI ETHICS</b>	-

Lecturer(s)	Department(s)
<b>Coordinator:</b> j. assit. Goda Klumbytė <b>Other(s):</b>	Institute of International Relations and Political Science, Vilnius university, Vokiečių str. 10, LT-01130, Vilnius, tel. +370 52514130, e-mail: tspimi@tspmi.vu.lt

Study cycle	Type of the course unit
First	Elective

Mode of delivery	Course unit delivery period	Language (s) of instruction
Face-to-face	5 (spring) semester	English

Requirements for students	
<b>Pre-requisites:</b> -	<b>Co-requisites (if any):</b> -

Number of credits allocated	Total student's workload	Contact hours	Self-study hours
5	130	32	98

Purpose of the course unit: programme competences to be developed		
<p>Aim of this course is to introduce students into the field of AI ethics, algorithmic accountability and fairness, key social, cultural, and political issues surrounding the deployment of AI and algorithmic systems. Students will learn about ethical; to develop their knowledge about key ethical dilemmas in AI, key ethical principles currently proposed and applied, the concepts of fairness, accountability and transparency, their conceptualization and application in algorithmic systems; also to develop their abilities to identify, interpret, and analyse questions around ethics, fairness and accountability in the field of AI and algorithmic systems while working in an interdisciplinary environment.</p>		
Learning outcomes of the course unit	Teaching and learning methods	Assessment methods
Students will understand and be able to explain the main concepts of ethics, accountability, and fairness as they are conceptualized and addressed in AI and algorithmic systems.	Problem-oriented lectures, seminars (analysis of most recent studies of relevant topics, discussions about important cases which had a great impact to the society), individual studies (critical literature studies, the analysis of theoretical debates and practical cases)  Introductory lectures, seminars (analysis of the leading literature and completed projects in the field), flipped classrooms, assignments and creative tasks, hands-on activities  Problem-oriented lectures, seminars (analysis of most recent studies of relevant topics, discussions about important cases which had a great impact to the society), individual studies (critical literature studies, the analysis of theoretical debates and practical cases)  Problem-oriented lectures, seminars (text analysis, comparative assessment and systemic analysis of practical issues, case study, oral presentations on assigned topics, group discussions), individual studies	Participation in seminars, presentation of a practical case study, individual project (academic conference simulation)
Students will be able to interpret, analyze and present research on ethics, accountability and fairness in AI and algorithmic systems.		
Students will be able to develop skills in critical thinking, informed and rational discussion, conceptual creativity, to engage in productive dialogue with representatives from non-social-science disciplines in order to address joint challenges and seek for solutions.		
Students will be able to conduct a scientifically sound analysis of ethical challenges related to AI and come up with evidence-based recommendations in various professional settings (especially in inter-disciplinary and innovative teams).		
Students will professionally communicate orally and in written, unambiguously and reasonably convey well-grounded ideas, arguments and conclusions based on		

theoretical knowledge and will be able to trigger or contribute to the discussion.	(individual search of information, critical literature studies and the analysis of theoretical and practical problems), individual project
Students will be able to identify and explain ethical challenges that emerge in contemporary AI.	Problem-oriented lectures, seminars (analysis of most recent studies of relevant topics, discussions about important cases which had a great impact to the society), individual studies (critical literature studies, the analysis of theoretical debates and practical cases)

Content: breakdown of the topics	Contact hours							Self-study: hours and assignments	
	Lectures	Consultations	Seminars	Practical sessions	Laboratory activities	Internship/work placement	Contact hours	Self-study hours	Assignments
1. Introduction: What is AI and what does it have to do with ethics? Introducing concepts of ethics, accountability and fairness in AI and algorithmic systems.	2		4				6	13	<p><b>Reading and preparation for in-class discussion:</b></p> <ul style="list-style-type: none"> <li>- Joanna J. Bryson. 2020. The Artificial Intelligence of the Ethics of Artificial Intelligence. In <i>The Oxford Handbook of Ethics of AI</i>, Markus D. Dubber, Frank Pasquale, Sunit Das, Markus D. Dubber, Frank Pasquale, Sunit Das and Joanna J. Bryson, Eds. Oxford University Press, 1-25. DOI: <a href="https://doi.org/10.1093/oxfordhb/9780190067397.013.1">https://doi.org/10.1093/oxfordhb/9780190067397.013.1</a>.</li> <li>- Andrew D. Selbst, Danah Boyd, Sorelle A. Friedler, Suresh Venkatasubramanian, and Janet Vertesi. 2019. Fairness and Abstraction in Sociotechnical Systems. In <i>Proceedings of the Conference on Fairness, Accountability, and Transparency - FAT* '19</i>. ACM Press, New York, New York, USA, 59-68. DOI: <a href="https://doi.org/10.1145/3287560.3287598">https://doi.org/10.1145/3287560.3287598</a>.</li> <li>- Cave, Stephen (2020): The Problem with Intelligence. In Annette Markham, Julia Powles, Toby Walsh, Anne L. Washington (Eds.): <i>Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society</i>. AIES '20: AAAI/ACM Conference on AI, Ethics, and Society. New York NY USA, 07 02 2020 09 02 2020. New York, NY, USA: ACM, pp. 29-35. <a href="http://lcfi.ac.uk/media/uploads/files/aiesfp12-Cave - approved_k12Qmfp.pdf">http://lcfi.ac.uk/media/uploads/files/aiesfp12-Cave - approved_k12Qmfp.pdf</a></li> <li>- Thomas M. Powers and Jean-Gabriel Ganascia. 2020. The Ethics of the Ethics of AI. In <i>The Oxford Handbook of Ethics of AI</i>, Markus D. Dubber, Frank Pasquale, Sunit Das, Thomas M. Powers and Jean-Gabriel Ganascia, Eds. Oxford University Press, 25-51. DOI: <a href="https://doi.org/10.1093/oxfordhb/9780190067397.013.2">https://doi.org/10.1093/oxfordhb/9780190067397.013.2</a>.</li> </ul>
2. Key socio-political issues in AI today: structural inequalities, bias and discrimination, ethical dilemmas			4				4	13	<p><b>Reading and preparation for in-class discussion:</b></p> <ul style="list-style-type: none"> <li>- Alexander G. Mirnig and Alexander Meschtscherjakov. 2019. Trolled by the Trolley Problem. In <i>Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems - CHI '19</i>. ACM Press, New York, New York, USA, 1-10. DOI: <a href="https://doi.org/10.1145/3290605.3300739">https://doi.org/10.1145/3290605.3300739</a>.</li> </ul>

								<ul style="list-style-type: none"> <li>- James Zou and Londa Schiebinger. 2018. AI can be sexist and racist - it's time to make it fair. <i>Nature</i> 559, 7714, 324-326. DOI: <a href="https://doi.org/10.1038/d41586-018-05707-8">https://doi.org/10.1038/d41586-018-05707-8</a>.</li> <li>- Sorelle A. Friedler, Carlos Scheidegger, and Suresh Venkatasubramanian. 2021. The (Im)possibility of Fairness. Different Value Systems Require Different Mechanisms For Fair Decision Making. <i>Commun. ACM</i> 64, 4, 136-143. DOI: <a href="https://doi.org/10.1145/3433949">https://doi.org/10.1145/3433949</a>.</li> <li>- Alex Campolo, Madelyn Sanfilippo, Meredith Whittaker, and Kate Crawford. 2017. AI Now 2017 Report. AI Now Institute, New York. <a href="https://ainowinstitute.org/AI_Now_2017_Report.pdf">https://ainowinstitute.org/AI_Now_2017_Report.pdf</a></li> </ul>
3. Implementing ethical perspectives in AI: defining fairness, accountability and transparency			6			6	8	<p><b>Reading and preparation for in-class discussion:</b></p> <ul style="list-style-type: none"> <li>- Han Yu, Zhiqi Shen, Chunyan Miao, Cyril Leung, Victor R. Lesser, and Qiang Yang. 2018. Building Ethics into Artificial Intelligence. <i>Proceedings of the 27th International Joint Conference on Artificial Intelligence (IJCAI'18)</i>, 5527-5533.</li> <li>- OECD. 2021. <i>State of Implementation of the OECD AI Principles: Insights from National AI Policies</i> 311. DOI: <a href="https://doi.org/10.1787/1cd40c44-en">https://doi.org/10.1787/1cd40c44-en</a>.</li> <li>- Christian H. Hoffmann and Benjamin Hahn. 2020. Decentered ethics in the machine era and guidance for AI regulation. <i>AI &amp; Soc</i> 35, 3, 635-644. DOI: <a href="https://doi.org/10.1007/s00146-019-00920-z">https://doi.org/10.1007/s00146-019-00920-z</a>.</li> <li>- Paula Boddington. 2020. Normative Modes. In <i>The Oxford Handbook of Ethics of AI</i>, Markus D. Dubber, Frank Pasquale, Sunit Das and Paula Boddington, Eds. Oxford University Press, 123-140. DOI: <a href="https://doi.org/10.1093/oxfordhb/9780190067397.013.7">https://doi.org/10.1093/oxfordhb/9780190067397.013.7</a>.</li> </ul>
4. From AI ethics to just AI: feminist, decolonial, anti-racist and disability-based perspectives to AI ethics and fairness			6			6	13	<p><b>Reading and preparation for in-class discussion:</b></p> <ul style="list-style-type: none"> <li>- Timnit Gebru. 2020. Race and Gender. In <i>The Oxford Handbook of Ethics of AI</i>, Markus D. Dubber, Frank Pasquale, Sunit Das and Timnit Gebru, Eds. Oxford University Press, 251-269. DOI: <a href="https://doi.org/10.1093/oxfordhb/9780190067397.013.16">https://doi.org/10.1093/oxfordhb/9780190067397.013.16</a>.</li> <li>- Alexandra R. Givens and Meredith R. Morris. 2020. Centering disability perspectives in algorithmic fairness, accountability, &amp; transparency. In <i>Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency</i>. ACM, New York, NY, USA, 684. DOI: <a href="https://doi.org/10.1145/3351095.3375686">https://doi.org/10.1145/3351095.3375686</a>.</li> <li>- Alex Hanna, Emily Denton, Andrew Smart, and Jamila Smith-Loud. 2020. Towards a Critical Race Methodology in Algorithmic Fairness. In <i>Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency</i>. ACM, New York, NY, USA, 501-512. DOI: <a href="https://doi.org/10.1145/3351095.3372826">https://doi.org/10.1145/3351095.3372826</a>.</li> <li>- Shakir Mohamed, Marie-Therese Png, and William Isaac. 2020. Decolonial AI: Decolonial</li> </ul>

									<p>Theory as Sociotechnical Foresight in Artificial Intelligence. <i>Philos. Technol.</i> 33, 4, 659-684. DOI: <a href="https://doi.org/10.1007/s13347-020-00405-8">https://doi.org/10.1007/s13347-020-00405-8</a>.</p> <p>-Abeba Birhane. 2021. Algorithmic injustice: a relational ethics approach. <i>Patterns</i> 2, 2, 100205. DOI: <a href="https://doi.org/10.1016/j.patter.2021.100205">https://doi.org/10.1016/j.patter.2021.100205</a>.</p>
5. AI as a critical technical practice: approaching ethics, accountability and fairness through design			6			6	13		<p><b>Reading and preparation for in-class discussion:</b></p> <p>- Phillip E. Agre. 1997. Toward a Critical Technical Practice: Lessons Learned in Trying to Reform AI. In <i>Social science, technical systems, and cooperative work. Beyond the great divide</i>, Geoffrey C. Bowker, Susan L. Star, Les Gasser and William Turner, Eds. Computers, cognition, and work. Psychology Press, New York, 131-157.</p> <p>- Claude Draude, Goda Klumbyte, Phillip Lücking, and Pat Treusch. 2019. Situated algorithms: a sociotechnical systemic approach to bias. <i>OIR</i> 44, 2, 325-342. DOI: <a href="https://doi.org/10.1108/OIR-10-2018-0332">https://doi.org/10.1108/OIR-10-2018-0332</a>.</p> <p>- Jessica Morley, Anat Elhalal, Francesca Garcia, Libby Kinsey, Jakob Mökander, and Luciano Floridi. 2021. Ethics as a Service: A Pragmatic Operationalisation of AI Ethics. <i>Minds and machines</i> 31, 2, 239-256. DOI: <a href="https://doi.org/10.1007/s11023-021-09563-w">https://doi.org/10.1007/s11023-021-09563-w</a>.</p> <p>- Anuradha Reddy, Iohanna Nicenboim, James Pierce, and Elisa Giaccardi. 2020. Encountering ethics through design: a workshop with nonhuman participants. <i>AI &amp; Soc.</i> DOI: <a href="https://doi.org/10.1007/s00146-020-01088-7">https://doi.org/10.1007/s00146-020-01088-7</a>.</p>
Presentation of a practical case study							4		Preparation of presentation of practical case study
Individual project							28		Preparation of individual project related to an ethical dilemma selected by the student that expands on and incorporates the case study.
Academic conference simulation (individual project presentation)			4			4	6		Preparation for academic conference simulation
<b>Total</b>	<b>2</b>		<b>26</b>	<b>4</b>		<b>32</b>	<b>98</b>		

Assessment strategy	Weight, percentage	Assessment period	Assessment criteria
Participation in seminars	40	During semester	Active and effective participation in seminars assessment consists of: <ul style="list-style-type: none"> <li>- Quality of comments, insights and relevant remarks (10%);</li> <li>- Ability to base answers on academic literature (10%);</li> <li>- Ability to raise questions for the discussion (20%).</li> </ul>
Presentation of a practical case study	10	During semester	For each session (except for the first three introductory sessions) one or two students will be asked to prepare a brief (max 10 minutes) presentation of a practical case study related to the reading material and/or topic of the session. <p>Assessment of presentation consists of:</p> <ul style="list-style-type: none"> <li>- Quality of presentation (5%) -</li> <li>- Ability to answer questions related with presented case (5%)</li> </ul>
Final individual project	50	At the end of semester	The individual project has to be around 6-10 pages, i.e. 3000 - 6000 words, excluding bibliography. At the end of the course a mini-conference will be organised where students will present their projects and get feedback from each other. The aim of this is to train students to write and present academic papers and simulate an academic conference setting.

			The assessment consists of correspondence to methodological requirements of research (10%), quality of analysis (10%), comprehensiveness and consistency of argumentation (10%), coherent presentation during the mini student conference (20%).
--	--	--	--

Author	Year of publication	Title	Issue of periodical or volume of publication	Publishing place and house or web link
<b>Compulsory reading</b>				
Markus D. Dubber, Frank Pasquale, Sunit Das (eds.)	2020	<i>The Oxford Handbook of Ethics of AI</i>		Oxford University Press
Andrew D. Selbst, Danah Boyd, Sorelle A. Friedler, Suresh Venkatasubramanian, and Janet Vertesi	2019	Fairness and Abstraction in Sociotechnical Systems		<i>Proceedings of the Conference on Fairness, Accountability, and Transparency - FAT* '19.</i> <a href="https://doi.org/10.1145/3287560.3287598">https://doi.org/10.1145/3287560.3287598</a>
Stephen Cave	2020	The Problem with Intelligence		<i>Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society</i> <a href="http://cfi.ac.uk/media/uploads/files/aiesfp12-Cave_-_approved_k12Qmfp.pdf">http://cfi.ac.uk/media/uploads/files/aiesfp12-Cave_-_approved_k12Qmfp.pdf</a>
Alexander G. Mirnig and Alexander Meschtscherjakov	2019	Trolled by the Trolley Problem.		<i>Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems - CHI '19.</i> DOI: <a href="https://doi.org/10.1145/3290605.3300739">https://doi.org/10.1145/3290605.3300739</a> .
James Zou and Londa Schiebinger.	2018	AI can be sexist and racist - it's time to make it fair	Vol. 599, No. 7714	Nature, <a href="https://doi.org/10.1038/d41586-018-05707-8">https://doi.org/10.1038/d41586-018-05707-8</a>
Sorelle A. Friedler, Carlos Scheidegger, and Suresh Venkatasubramanian	2021	The (Im)possibility of Fairness. Different Value Systems Require Different Mechanisms For Fair Decision Making	Vol. 64, No. 6	<i>Commun. ACM</i> , DOI: <a href="https://doi.org/10.1145/3433949">https://doi.org/10.1145/3433949</a>
Alex Campolo, Madelyn Sanfilippo, Meredith Whittaker, and Kate Crawford	2017	AI Now 2017 Report		<a href="https://ainowinstitute.org/AI_Now_2017_Report.pdf">https://ainowinstitute.org/AI_Now_2017_Report.pdf</a>
Han Yu, Zhiqi Shen, Chunyan Miao, Cyril Leung, Victor R. Lesser, and Qiang Yang.	2018	Building Ethics into Artificial Intelligence		<i>Proceedings of the 27th International Joint Conference on Artificial Intelligence (IJCAI'18)</i>
OECD	2021	State of Implementation of the OECD AI Principles: Insights from National AI Policies	Vol. 311	DOI: <a href="https://doi.org/10.1787/1cd40c44-en">https://doi.org/10.1787/1cd40c44-en</a> .
Christian H. Hoffmann and Benjamin Hahn	2020	Decentered ethics in the machine era and guidance for AI regulation	Vol. 35, Issue 3	<i>AI &amp; Soc</i> DOI: <a href="https://doi.org/10.1007/s00146-019-00920-z">https://doi.org/10.1007/s00146-019-00920-z</a> .
Alexandra R. Givens and Meredith R. Morris	2020	Centering disability perspectives in algorithmic fairness, accountability, & transparency		<i>Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency.</i> DOI: <a href="https://doi.org/10.1145/3351095.3375686">https://doi.org/10.1145/3351095.3375686</a> .
Scheurman, M. K., Jiang, J. A., Spiel, K., & Brubaker, J.	2021	Revisiting Gendered Web Forms: An Evaluation of Gender Inputs with (Non-) Binary People		<a href="https://www.morgan-klaus.com/pdfs/pubs/Scheurman2021-gender-forms.pdf">https://www.morgan-klaus.com/pdfs/pubs/Scheurman2021-gender-forms.pdf</a>

Alex Hanna, Emily Denton, Andrew Smart, and Jamila Smith-Loud	2020	Towards a Critical Race Methodology in Algorithmic Fairness		Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency. DOI: <a href="https://doi.org/10.1145/3351095.3372826">https://doi.org/10.1145/3351095.3372826</a> .
Shakir Mohamed, Marie-Therese Png, and William Isaac	2020	Decolonial AI: Decolonial Theory as Sociotechnical Foresight in Artificial Intelligence	Vol. 33, Issue 4	<i>Philos. Technol.</i> DOI: <a href="https://doi.org/10.1007/s13347-020-00405-8">https://doi.org/10.1007/s13347-020-00405-8</a> .
Abeba Birhane	2021	Algorithmic injustice: a relational ethics approach	Vol. 2, Issue 2	<i>Patterns</i> , DOI: <a href="https://doi.org/10.1016/j.patter.2021.100205">https://doi.org/10.1016/j.patter.2021.100205</a> .
Phillip E. Agre	1997	Toward a Critical Technical Practice: Lessons Learned in Trying to Reform AI		<i>Social science, technical systems, and cooperative work. Beyond the great divide</i> , Geoffrey C. Bowker, Susan L. Star, Les Gasser and William Turner, Eds. Computers, cognition, and work. Psychology Press, New York, 131-157.
Claude Draude, Goda Klumbyte, Phillip Lücking, and Pat Treusch	2019	Situated algorithms: a sociotechnical systemic approach to bias	Vol. 44, Issue 2	<i>Online Information Review</i> , DOI: <a href="https://doi.org/10.1108/OIR-10-2018-0332">https://doi.org/10.1108/OIR-10-2018-0332</a>
Jessica Morley, Anat Elhalal, Francesca Garcia, Libby Kinsey, Jakob Mökander, and Luciano Floridi.	2021	Ethics as a Service: A Pragmatic Operationalisation of AI Ethics	Vo. 31, Issue 2	<i>Minds and machines</i> . DOI: <a href="https://doi.org/10.1007/s11023-021-09563-w">https://doi.org/10.1007/s11023-021-09563-w</a> .
Anuradha Reddy, Iohanna Nicenboim, James Pierce, and Elisa Giaccardi	2020	Encountering ethics through design: a workshop with nonhuman participants		<i>AI &amp; Soc.</i> DOI: <a href="https://doi.org/10.1007/s00146-020-01088-7">https://doi.org/10.1007/s00146-020-01088-7</a> .
<b>Recommended reading</b>				
Angie Abdilla, Noelani Arista, Kaipulaumakaniolono Baker, Scott Benesiinaabandan, Michelle Brown, Melanie Cheung, Meredith Coleman, Ashley Cordes, Joel Davison, Kūpono Duncan, Sergio Garzon, D. F. Harrell, Peter-Lucas Jones, Kekuhi Kealiikanakaoleohaililani, Megan Kelleher, Suzanne Kite, Olin Lagon, Jason Leigh, Maroussia Levesque, Jason E. Lewis, Keoni Mahelona, Caleb Moses, Isaac Nahuewai, Kari Noe, Danielle Olson, 'Ōiwi Parker Jones, Caroline Running Wolf, Michael Running Wolf, Marlee Silva, Skawennati Fragnito, and Hēmi Whaanga	2020	Indigenous Protocol and Artificial Intelligence Position Paper		DOI: <a href="https://doi.org/10.11573/spectrum.library.concordia.ca.00986506">https://doi.org/10.11573/spectrum.library.concordia.ca.00986506</a> .

Mathieu d'Aquin, Pinelopi Troullinou, Noel E. O'Connor, Aindrias Cullen, Gráinne Faller, and Louise Holden	2018	Towards an "Ethics by Design" Methodology for AI Research Projects		<i>Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society (AIES '18).</i> DOI: <a href="https://doi.org/10.1145/3278721.3278765">https://doi.org/10.1145/ 3278721.3278765</a>
---	------	--	--	--