

Lehrveranstaltung		LV-Kurzbezeichnung
Responsible and Trustworthy Artificial Intelligence Systems		KTAI
Verantwortliche/r	Fakultät	
Prof. Dr. Shakkeera Liaquath	Informatik und Mathematik	
Lehrende/r / Dozierende/r	Angebotsfrequenz	
Prof. Dr. Shakkeera Liaquath		
Lehrform		
Seminaristischer Unterricht mit Übungen		

Studiensemester gemäß Studienplan	Lehrumfang [SWS oder UE]	Lehrsprache	Arbeitsaufwand [ECTS-Credits]
	4 SWS	Englisch	5

Zeitaufwand:

Präsenzstudium	Eigenstudium
60	90

Studien- und Prüfungsleistung
Studienarbeit mit Präsentation
Zugelassene Hilfsmittel für Leistungsnachweis
keine

Inhalte und Qualifikationsziele
<ul style="list-style-type: none"> <li>• Foundations of Responsible and Ethical Artificial Intelligence</li> <li>• Ethical principles in the design, development and deployment of AI systems</li> <li>• Bias, fairness, inclusivity and non-discrimination in AI</li> <li>• Transparency, accountability and human oversight in AI-based decision-making</li> <li>• Explainable Artificial Intelligence techniques and methods</li> <li>• Privacy, security and trust in AI systems</li> <li>• AI governance, policies, standards and regulatory frameworks</li> <li>• Global AI regulations and compliance requirements</li> <li>• Evaluation of AI systems with regard to societal, organisational and technical impact</li> <li>• Case studies in responsible and trustworthy AI deployment across real-world applications</li> </ul>

Inhalte und Qualifikationsziele (Fortsetzung)
<b>Lernziele: Fachkompetenz</b>
After successful completion of the module, students are able to, <ul style="list-style-type: none"><li>• understand and explain the foundations, concepts and principles of responsible and trustworthy Artificial Intelligence (1).</li><li>• identify ethical, legal and technical challenges in the design and deployment of AI systems (2).</li><li>• analyse bias, fairness, inclusivity and transparency issues in AI models and AI-supported decision-making processes (2).</li><li>• select and apply suitable methods for explainable Artificial Intelligence and accountable AI system design (2).</li><li>• evaluate AI systems with regard to privacy, security, robustness, trustworthiness and regulatory compliance (3).</li><li>• understand and compare global AI governance approaches, policies and regulatory frameworks (2).</li><li>• design AI systems and AI-based applications that are aligned with ethical principles, organisational requirements and regulatory standards (3).</li></ul>
<b>Lernziele: Persönliche Kompetenz</b>
After successful completion of the module, students are able to, <ul style="list-style-type: none"><li>• reflect on the societal, ethical and organisational implications of AI systems and AI-based decision-making (3).</li><li>• assess their own responsibility in the design, development, deployment and evaluation of trustworthy AI systems (3).</li><li>• communicate ethical, legal and technical risks of AI systems to different stakeholder groups in a clear and structured manner (2).</li><li>• work collaboratively on responsible AI concepts, case studies and solution approaches in interdisciplinary teams (3).</li><li>• critically evaluate the impact of AI systems on individuals, organisations and society (3).</li></ul>
<b>Angeborene Lehrunterlagen</b>
Slides
<b>Lehrmedien</b>
<b>Literatur</b>
Foundations of Responsible and Ethical Artificial Intelligence <ul style="list-style-type: none"><li>• Butvinik, D., Foundations of Ethical AI: Concepts and Principles of Explainability and Trust, <a href="https://www.niceactimize.com/blog/foundations-of-ethical-ai-concepts-and-principles-of-explainability-and-trust">https://www.niceactimize.com/blog/foundations-of-ethical-ai-concepts-and-principles-of-explainability-and-trust</a></li><li>• Papagiannidis, E., Mikalef, P., &amp; Conboy, K. (2025). Responsible artificial intelligence governance: A review and research framework. <i>The Journal of Strategic Information Systems</i>, 34(2), 101885. <a href="https://doi.org/10.1016/j.jsis.2024.101885">https://doi.org/10.1016/j.jsis.2024.101885</a></li></ul> Ethical Principles in the Design, Development and Deployment of AI Systems <ul style="list-style-type: none"><li>• UNESCO. (2021). <i>Recommendation on the ethics of artificial intelligence</i>. <a href="https://www.unesco.org/en/artificial-intelligence/recommendation-ethics">https://www.unesco.org/en/artificial-intelligence/recommendation-ethics</a></li><li>• Transcend. (n.d.). <i>Key principles for ethical AI development</i>. <a href="https://transcend.io/blog/ai-ethics">https://transcend.io/blog/ai-ethics</a></li><li>• Shukla, S. (2024). Principles Governing Ethical Development and Deployment of AI. <i>International Journal of Engineering, Business and Management</i>, 8(2), 26–46. <a href="https://doi.org/10.22161/ijeem.8.2.5">https://doi.org/10.22161/ijeem.8.2.5</a></li><li>• McKinsey &amp; Company. (n.d.). <i>Responsible AI (RAI) principles</i>. <a href="https://www.mckinsey.com/capabilities/quantumblack/how-we-help-clients/generative-ai/responsible-ai-principles">https://www.mckinsey.com/capabilities/quantumblack/how-we-help-clients/generative-ai/responsible-ai-principles</a></li></ul> Bias, Fairness, Inclusivity and Non-Discrimination in AI <ul style="list-style-type: none"><li>• GeeksforGeeks. (n.d.). <i>Fairness and Bias in Artificial Intelligence</i>. <a href="https://www.geeksforgeeks.org/artificial-intelligence/fairness-and-bias-in-artificial-intelligence/">https://www.geeksforgeeks.org/artificial-intelligence/fairness-and-bias-in-artificial-intelligence/</a></li><li>• Inter-Parliamentary Union. (n.d.). <i>Ethical principles: Fairness and non-discrimination</i>. In <i>AI guidelines for parliaments</i>. <a href="https://www.ipu.org/ai-guidelines/ethical-principles-fairness-and-non-discrimination">https://www.ipu.org/ai-guidelines/ethical-principles-fairness-and-non-discrimination</a></li><li>• BluePrism (2024), Fairness and Bias in AI Explained, <a href="https://www.blueprism.com/resources/blog/bias-fairness-ai/">https://www.blueprism.com/resources/blog/bias-fairness-ai/</a></li><li>• Chapman University. (n.d.). <i>Bias in AI</i>. <a href="https://www.chapman.edu/ai/bias-in-ai.aspx">https://www.chapman.edu/ai/bias-in-ai.aspx</a></li></ul> Transparency, Accountability and Human Oversight in AI-Based Decision-Making <ul style="list-style-type: none"><li>• James, M. (2026). Balancing Automation and Accountability: Human Oversight in AI-Based Planning and Forecasting Systems. [Preprint]. ResearchGate. <a href="https://www.researchgate.net/publication/399795396">https://www.researchgate.net/publication/399795396</a></li><li>• Dancy, T., &amp; Zalnierute, M. (2026). AI and transparency in judicial decision making. <i>Oxford journal of legal studies</i>, 46(1), 1-34. <a href="https://doi.org/10.1093/ojls/laaf030">https://doi.org/10.1093/ojls/laaf030</a></li><li>• Judjianto, L., &amp; Rustiyana, R. (2026). AI-GOVERNANCE AND ALGORITHMIC ACCOUNTABILITY: RETHINKING LEGAL STANDARDS IN THE AGE OF AUTONOMOUS DECISION-MAKING. <i>INJOSS 5(1)</i>, 11-25. <a href="https://injoqast.net/index.php/JOSS/article/view/762745">https://injoqast.net/index.php/JOSS/article/view/762745</a></li><li>• Langer, M., Dachselt, R., Liao, Q. V., Miller, T., &amp; Tintarev, N. (2026). Challenges of Human Oversight: Achieving Human Control of AI-Based Systems (Dagstuhl Seminar 25272). <i>Dagstuhl Reports</i>, 15(6), 189-204. <a href="https://drops.dagstuhl.de/storage/04dagstuhl-reports/volume15/issue06/25272/DagRep.15.6.189/DagRep.15.6.189.pdf">https://drops.dagstuhl.de/storage/04dagstuhl-reports/volume15/issue06/25272/DagRep.15.6.189/DagRep.15.6.189.pdf</a></li></ul> Explainable Artificial Intelligence Techniques and Methods <ul style="list-style-type: none"><li>• Moraes, I. A., Arrighi, L., Barbon Junior, S., Cunha, R. L., &amp; Barbin, D. F. (2026). Explainable artificial intelligence (XAI) applied to deep computer vision for the assessment and classification of oleogels with varying oleogelator types and concentrations. <i>Microchemical Journal</i>, 221, 116821. <a href="https://doi.org/10.1016/j.microc.2026.116821">https://doi.org/10.1016/j.microc.2026.116821</a></li><li>• Buñay-Guisñan, P., Lara, J.A., Cano, A. et al. Towards accessible AI for addressing students' academic dropout: an auto machine learning and explainable artificial intelligence approach. <i>Univ Access Inf Soc</i> 25, 19 (2026). <a href="https://doi.org/10.1007/s10209-025-01278-4">https://doi.org/10.1007/s10209-025-01278-4</a></li><li>• Bordekar, H., Koord, J., Völkerink, O., &amp; Hühne, C. (2026). Microstructure characterization of fiber composite laminate using eXplainable Artificial Intelligence. <i>Journal of Composite Materials</i>, 60(1), 3-25. <a href="https://doi.org/10.1177/00219983251345559">https://doi.org/10.1177/00219983251345559</a></li></ul>

