Dr. phil. Kevin Baum, M.Sc., M.A.

♂ 28.07.1986, Schwetzingen
academia@kevinbaum.de https://kevinbaum.de Scholar LinkedIn

(Academic) Employment History (since the end of my Master's studies)

09/2024 – present
Independent Research Group Leader.
Responsible AI and Machine Ethics (RAIME)
Research Department for Neuro-Mechanistic Modeling (NMM)
German Research Center for Artificial Intelligence (DFKI), Saarbrücken, Saarland.

01/2023 – present
Deputy Head.
Research Department for Neuro-Mechanistic Modeling (NMM)
German Research Center for Artificial Intelligence (DFKI), Saarbrücken, Saarland.

01/2023 – 05/2025
Lab Manager.
Research Department for Neuro-Mechanistic Modeling (NMM)
German Research Center for Artificial Intelligence (DFKI), Saarbrücken, Saarland.

04/2019 – 12/2022
Research Associate.
Project *Explainable Intelligent Systems*, funded by VolkswagenStiftung
PIs Wessels (Ethics) and Hermanns (Computer Science)
Saarland University (UdS), Saarbrücken, Saarland.

02/2015 – 04/2019
Research Associate and Lecturer.
Group for Dependable Systems and Software (Prof. Hermanns)
Saarland University (UdS), Saarbrücken, Saarland.

Research Associate and Lecturer.
Chair for Practical Philosophy (Prof. Wessels, Prof. Fehige)
Saarland University (UdS), Saarbrücken, Saarland.

Affiliations

04/2025 – 05/2025
Guest Associate Professor.
Department of Information Science and Media Studies, University of Bergen, Norway

09/2024 – present
Member (Associated Research since 2018).
Collaborative Research Center "Perspicuous Computing" (CRC 240, DFG).

12/2023 – present
Executive Board Member (Head until 04/2025).
Center for European Research in Trusted AI (CERTAIN).
German Research Center for Artificial Intelligence (DFKI), Saarbrücken, Saarland.

01/2023 – present
Associated Researcher.
Volkswagen Foundation research project Explainable Intelligent Systems (EIS).

12/2019 – present
Founding Member.
Algoright e.V.

Education

2024
Doctor of Philosophy at TU Dortmund *summa cum laude*
Thesis title: *Doing Wrong with Others – Multi-Agent Consequentialism as a Solution for the Collective Action Problem.*

2014
M.A. Philosophy at Saarland University *1.1*
Thesis title: *Vom Bezug singulärer Terme in Aussagen über propositionale Einstellungen.*

Education (continued)

2013 **M.Sc. Computer Science** at Saarland University *1.4*
Thesis title: *GPGPU-gestützte diffusionsbasierte naive Videokompression.*

2011 **B.Sc. Computer Science, (Minor: Mathematics)** at Saarland University *2.0*
Thesis title: *Stützstellenauswahl für diffusionsbasierte Bildkompression unter Berücksichtigung einer Quadrixel-Substruktur-Restriktion.*

2006 **Abitur** at Gymnasium Süderelbe, Hamburg *1.5*

Awards and Achievements

2020 **Hochschulperle of the Year for lecture »Ethics for Nerds«**, German Stifterverband.

2019 **Hochschulperle of the Month (January) for lecture »Ethics for Nerds«**, German Stifterverband.

2011, 2012 **Deutschlandstipendium**

University Teaching

Summer 2025 **»Einführung in die KI-Ethik (und, spezifischer, warum Verzerrungen in Daten ein Problem sein können)«**, individual lecture as part of the lecture series *»Grundkurs Künstliche Intelligenz«* by Prof. Verena Wolf. Saarland Informatics Campus, UdS.

Summer 2023 **»AI for the Social Good« (Seminar)** with Dr. Gerrit Großmann, Lisa Dargasz, Sarah Sterz, Prof. Verena Wolf. Saarland Informatics Campus, UdS. *7 ECTS*

»Ethics for Nerds« (Lecture) with Sarah Sterz, Prof. Holger Hermanns. Saarland Informatics Campus, UdS. *6 ECTS*

Summer 2022 **»Computer Ethics« (Lecture)** with Sarah Sterz. LL.M. *Informationstechnologie und Recht*, UdS. *3 ECTS*

»Ethics for Nerds« (Lecture) with Sarah Sterz, Prof. Holger Hermanns. UdS. *6 ECTS*

Winter 2021/2022 **»Ethische Fragen des Technologieeinsatzes« (Seminar).** Bucerius Law School, Hamburg. *2 ECTS*

Summer 2021 **»Computer Ethics for IT & Law« (Lecture)** with Sarah Sterz. LL.M. *Informationstechnologie und Recht*, UdS. *3 ECTS*

»Ethics for Nerds« (Lecture) with Sarah Sterz, Prof. Holger Hermanns. Saarland Informatics Campus, UdS. *6 ECTS*

Winter 2020/2021 **»Ethische Fragen des Technologieeinsatzes« (Seminar).** Bucerius Law School, Hamburg. *2 ECTS*

Summer 2020 **»Ethics for Nerds« (Lecture)** with Sarah Sterz, Prof. Holger Hermanns, Saarland Informatics Campus, UdS. *6 ECTS*

Summer 2019 **»Ethics for Nerds« (Lecture)** with Sarah Sterz, Prof. Holger Hermanns, Saarland Informatics Campus, UdS. *6 ECTS*

Winter 2018/2019 **»Computer sagt: „wahrscheinlich“. Wie KI und Algorithmen unsere Welt verändern« (Seminar).** College of Fine Arts (Universität der Künste), Berlin. *2 ECTS*

University Teaching (continued)

	»Weapons of Math Destruction – Wie Big Data Ungerechtigkeit fördert und Demokratie gefährdet« (Seminar). Philosophy Department, UdS. *6 ECTS*
Summer 2018	**»Ethics for Nerds« (Lecture)** with Prof. Holger Hermanns. Saarland Informatics Campus, UdS. *6 ECTS*
	»Derek Parfits Praktische Philosophie in *Reasons & Persons*« (Seminar). Philosophy Department, UdS. *6 ECTS*
Winter 2017/2018	**»Das Problem gemeinschaftlichen Handels« (Seminar).** Philosophy Department, UdS. *6 ECTS*
Summer 2017	**»Moral und Algorithmen« (Seminar).** Philosophy Department, UdS. *6 ECTS*
	»Ethics for Nerds« (Lecture) with Prof. Holger Hermanns. Saarland Informatics Campus, UdS. *6 ECTS*
Winter 2017/2018	**»Von der Entscheidungstheorie zum Konsequentialismus« (Seminar).** Philosophy Department, UdS. *6 ECTS*
Summer 2016	**»Extending Morals: Moralisch handelnde Roboter, Roboter moralisch behandeln« (Seminar).** Philosophy Department, UdS. *6 ECTS*
	»Ethics for Nerds« (Lecture) with Prof. Holger Hermanns. Saarland Informatics Campus, UdS. *6 ECTS*
Winter 2015/2016	**»Was kommt jetzt? Die Zukunft der Menschheit, Transhumanismus und die technologische Singularität« (Seminar).** Philosophy Department, UdS. *6 ECTS*
Summer 2015	**»Es macht doch keinen Unterschied! Oder: Was wäre, wenn jeder das täte?« (Seminar).** Philosophy Department, UdS. *6 ECTS*
	»Ethik für Nerds« (Proseminar) with Prof. Holger Hermanns. Saarland Informatics Campus, UdS. *5 ECTS*
Summer 2013	**»Prädikatenlogik erster Stufe« (Seminar).** Philosophy Department, UdS. *6 ECTS*

Conference Tutorials, Summer School Lectures, & Online Teaching

Winter 2025*	**»KI und Moral: Der Basiskurs«** with Sarah Sterz, Andre Steingrüber, and Laura Stenzel. KI Campus. *(online course, in publication)*
September 2025*	**»Ethical Reasoning in the Dark: How to Make Justifiable Decisions When You Don't Know What's Right«** at the *Summer Institutes of Computational Social Science (SICSS) Saarbrücken*. Saarbrücken, 8 – 19 September 2025.
July 2025	**»The Ethics of Humans *in* and *on* the Loops«** at the *Europaeum Summer School: AI and the Digital Future* at the University of Luxembourg. Luxembourg, 7 – 9 July 2025.
July 2024	**»Machine Ethics. A Tutorial for Prospective Researchers«** with Marija Slavkovik (University of Bergen) and Louise Dennis (University of Manchester). *33rd International Joint Conference on Artificial Intelligence* (IJCAI 2024), Jeju Island, South Korea.

Supervision and Review of Student Theses, Student Advising

Independently Supervised and Reviewed Bachelor's and Master's Theses

- 2025 *Causal Explanations and Epistemic Relativization.* Master's Thesis of Yannic Muskalla **(M.A., Philosophy; 2nd Reviewer)**.
- 2024-2025 *Integrating Reason-Based Moral Decision-Making in the Reinforcement Learning Architecture.* Master's Thesis of Lisa Dargasz **(M.Sc., Computer Science; Supervisor; 2nd Reviewer)**.
- 2023 *Outputs as Evidence: Exploring the Explanatory Role of XAI Methods through Thomas Bartelborth's Theory on Nomic Patterns.* Bachelor's Thesis of Lena Marie Budde **(B.A., Philosophy; 1st Reviewer, Supervisor)**.
- 2018 *Another Take on Newcomb's Problem.* Bachelor's Thesis of Tim Dirk Metzler **(B.A., Philosophy; 1st Reviewer, Supervisor)**.

Co-Supervised Bachelor's and Master's Theses, Advisory Service, External Reviewer

- 2025 *Lexicographic Modeling of Preferences in the Moral Machine Experiment. Role: External Reviewer.* Master's Thesis by Victoria Ronæss **(M.Sc., Computer Science)**, University of Oslo. Role: External Reviewer.
- 2018-2023 *Building Bridges for Better Machines: From Machine Ethics to Machine Explainability and Back.* Doctoral Thesis by Timo Speith **(Dr. phil.)**, Saarland University. Role: Advisory Services.
- 2020 *Gibt es Umstände, unter denen Abtreibung moralisch erlaubt ist?* State examination **(StEx)** thesis by Alisha Monika Mateas. Role: Advisory Services.
- 2018 *A Framework of Verifiable Machine Ethics and Machine Explainability.* Master's Thesis of Timo Speith **(M.Sc., Computer Science)**. Role: Advisory Services.
- 2017-2018 *Schuld der Massen – Kollektive Verantwortung und das Überdeterminiertheitsproblem.* Master's Thesis of Yannik Bast **(M.A., Philosophy)**. Role: Advisory Services.
- 2017 *Klimawandel und Spieltheorie: Ist kooperatives Verhalten der Staaten in der Klimapolitik rational?* Bachelor's Thesis of Tessy Aulner **(B.A., Philosophy)**. Role: Advisory Services.
- *Warum ich einen Unterschied mache. Zur Ethik individuellen Handelns im gemeinschaftlichen Handeln.* Bachelor's Thesis of Sarah Maurer **(B.A., Philosophy)**. Role: Advisory Services.

Co-Supervised Research Immersion Labs

- 2023 Janine Lohse: *Predicting and Manipulating Game Rules through Weight Analysis in Neural Networks?*

Scientific Services & Science Infrastructure

(Co-)Organizing of Conference, Workshops, Special Tracks, Summer Schools, Tutorials

- 2025* **Scientific Committee Member** of the *5th Inria-DFKI European Summer School on AI* (IDESSAI 2025), October 2025, Paris, France.
- **Special Track Co-Organizer** of »Responsible and Trusted AI: An Interdisciplinary Perspective« at the *3rd Artificial Intelligence Symposium On Leveraging Applications of Formal Methods, Verification and Validation* (AISoLA 2025), end of October, 2025, Greece.

Scientific Services & Science Infrastructure (continued)

2024

- **Special Track Co-Organizer** of »Responsible and Trusted AI: An Interdisciplinary Perspective« at the *2nd Artificial Intelligence Symposium On Leveraging Applications of Formal Methods, Verification and Validation* (AISoLA 2024), 30 October – 3 November, 2024, Crete, Greece.
- **Tutorial Co-Organizer** on »Machine Ethics« at the *33rd International Joint Conference on Artificial Intelligence* (IJCAI 2024), 3 August – 9 August, 2024, Jeju, South Korea.

2023

- **Special Track Co-Organizer** of »Responsible and Trustworthy AI: Normative Perspectives on and Societal Implications of AI Systems« at the *1st Artificial Intelligence Symposium On Leveraging Applications of Formal Methods, Verification and Validation* (AISOLA 2023), 23 – 28 October 2023, Crete, Greece.
- **Special Track Co-Organizer** of »Interdisciplinary Perspectives on XAI« at the *1st World Conference on eXplainable Artificial Intelligence* (xAI 2023), 26 – 28 July 2023, Lisboa, Portugal

2019

- **Workshop Co-Organizer** of »Issues in Explainable AI: Blackboxes, Recommendations, and Levels of Explanation«, 30 September – 2 October, 2019, Saarbrücken, Germany.

Program Committee Memberships

2025

- **Program Committee Member** of the *Special Track on AI Alignment* at the *40th Annual AAAI Conference on Artificial Intelligence* (AAAI-26-AIA), January 20 – 27 2026, Singapore.
- **Program Committee Member** of the workshop *2nd Workshop on Formal Ethical Agents and Robots* (FEAR), November 2025, Manchester, UK.
- **Program Committee Member** of the *8th AAAI/ACM Conference on AI, Ethics, and Society* (AIES 2025), 20 – 22 October, 2025, Madrid, Spain.
- **Program Committee Member** of the *34th International Joint Conference on Artificial Intelligence* (IJCAI 2025), 16 – 22 August, 2025, Montreal, Canada.
- **Program Committee Member** of the *3rd TRR 318 Conference on Contextualizing Explanations* (ContEx25), 17 – 18 June, 2025, Bielefeld, Germany.
- **Program Committee Member** of the *24rd International Conference on Autonomous Agents and Multi-Agent Systems* (AAMAS 2025), 19 May – 23 May, 2025, Detroit, Michigan, United States.
- **Reviewer** at the *First Workshop on Sociotechnical AI Governance: Opportunities and Challenges for HCI* at the ACM *Conference on Human Factors in Computing Systems* (CHI) (STAIG@CHI'25).

2024

- **Program Committee Member** of the workshop *Formal Ethical Agents and Robots* (FEAR) co-located with *The 19th International Conference on Integrated Formal Methods* (iFM 2024), 13 November – 15 November, 2024, Manchester, UK.
- **Program Committee Member** of the *2nd Artificial Intelligence Symposium On Leveraging Applications of Formal Methods, Verification and Validation* (AISoLA 2024), 30 October – 3 November, 2024, Crete, Greece.
- **Program Committee Member** of the *7th AAAI/ACM Conference on AI, Ethics, and Society* (AIES 2024), 21 – 23 October, 2024, Santa Clara, California, USA.
- **Program Committee Member** of the workshop track on *Value Engineering in AI* (VALE) at the *International Workshop on AI Value Engineering and AI Compliance Meechanisms* (VECOMP) affiliated with the *27th European Conference on Artificial Intelligence* (ECAI 2024), 19 – 24 October, 2024, Santiago de Compostela, Spain.

Scientific Services & Science Infrastructure (continued)

- **Program Committee Member** of (the track on »Agents in Ethics« at) the *21st European Conference on Multi-Agent Systems* (EUMAS 2024), 26 – 28 August 2024, Dublin, Ireland.
- **Program Committee Member** of the *23rd International Conference on Autonomous Agents and Multi-Agent Systems* (AAMAS 2024), 6 – 10 May 2024, Auckland, New Zealand.
- 2023 **Program Committee Member** of the *1st Artificial Intelligence Symposium On Leveraging Applications of Formal Methods, Verification and Validation* (AISoLA 2023), 23 – 28 October 2023, Crete, Greece.

Reviewer Activities for Academic Journals

- Philosophical Studies, Synthese, Artificial Intelligence, Philosophy & Technology, AI & Society, Nature Communications, Communications of the ACM, Big Data & Society, AI and Ethics, Journal of Artificial Intelligence Research

Ethics Committees, Commissions, Advisory, and Board Functions

- 2025–* Science Mentor at **AI Grid**, a BMBF-funded national mentoring and networking program for outstanding AI students in Germany.
- 2024-* Member of the **Strategy Board of the Saarland de:hub on *Artificial Intelligence***, responsible for the topic of trustworthy AI in the context of sustainable transformation.
- 2024 Member of the **delegation of Federal Minister of Education and Research Bettina Stark-Watzinger** for the high-level **UK-German Science, Innovation and Technology Dialogue**, Imperial College White City Campus in London, March 12.
- 2023-* **Ethics Advisor**. Ethics Team, German Research Center for Artificial Intelligence (DFKI).
- DFKI Representative in the ***East Side Fab e.V.* board** (on behalf of Prof. Verena Wolf).
- 2021-* Member of the **Ethical Review Board**. Faculty of Mathematics and Computer Science, Saarland University.
- 2020-2022 **Permanent expert on digital ethics** at the **Enquête Commission »Digitalisierung im Saarland – Bestandsaufnahme, Chancen und Maßnahmen**« of the Saarland State Parliament.
- 2019-2022 Member and Deputy Chairman of the »**Kommission für die Ethik sicherheitsrelevanter Forschung**«. Saarland University.
- 2018-2022 **Ethics Advisor**. Undisclosed Horizon 2020 Project (together with Sarah Sterz and Timo Speith).

Miscellaneous Experience

Further Experiences

Media

I have been involved in a range of media productions. Below is a selection of highlights:

-Frankfurter Allgemeine Zeitung: Co-author of an opinion piece advocating for the establishment of a German AI Safety Institute, together with Jonas Andrulis (Founder and CEO of Aleph Alpha), Kristian Kersting (TU Darmstadt, DFKI), Sebastian Vollmer (RPTU, DFKI), Annika von Mutius (Co-Founder and Co-CEO of the HR tech startup Empion, board member of the *KI Bundesverband*), and Patrick Schramowski (TU Darmstadt, DFKI).
- Deutschlandfunk: Das war der Tag: Interview guest on Pope Francis's 2024 G7 Summit address concerning AI, ethics, and lethal autonomous weapons systems.
- ARD Panorama: Interviewee on the ethical implications of *Aspire Health*'s life expectancy prediction model.

Additional appearances range from regional media—such as *Saarländischer Rundfunk* (SR, television as well as radio broadcasts, including feature segments), and the Saarbrücker Zeitung, including interviews)—to national platforms such as Handelsblatt, DLF Kultur: Breitband, DLF Kultur: Studio 9, and DLF: Campus & Karriere.
I have also appeared in a variety of digital formats—including heise online, Spektrum der Wissenschaft, podcasts such as this one, radio features like this episode on Bayern 2, YouTube vlogs, and other online platforms.
Coming from a journalism family, I was practically raised around recording equipment. I also serve as the University of Saarland's point of contact for topics including *machine ethics, explainable AI (XAI), technology assessment, and the ethics of digitalization*, as part of its expert network for journalists.

Politics and Regulation

In addition to serving as a permanent expert on the ethics of digitalization for an *Enquête Commission* of the *Saarland State Parliament* (see above), I contributed to several political and regulatory processes at both the state and federal levels.
This includes preparing expert opinions for the *German Bundestag*—for example, on the draft legislation on autonomous driving—and for the Saarland State Parliament, including topics such as computer-assisted procedures used by the Saarland police in accident investigations and the use of digital technologies in schools (*Committee on Education*, March 2023).
My recommendations have also been cited by the *Saarland State Commissioner for Data Protection and Freedom of Information*. Additionally, I have taken part in background consultations with members of both the German Bundestag and the European Parliament, contributing to discussions on the structure and regulatory classification of two drafts of the EU AI Act.

Miscellaneous Experience (continued)

Startups, SMEs, Industry

- I have planned and conducted several large-scale, paid consultations and workshops for corporate clients such as Villeroy & Boch and Accso. I also bring startup experience, having been part of a team funded by the *EXIST-Gründungsstipendium* from 2016 to 2017.
  In recent years, I have established strong working relationships with a range of medium-sized enterprises and industrial partners, including ZF and eurodata.

Memberships

since 2025
- **International Association for Safe and Ethical AI** (IASEAI), a non-profit organization founded by Stuart Russell to address the risks and opportunities associated with rapid advances in AI, inaugural conference at the OECD *La Muette* Headquarters and Conference Centre in Paris, ahead of the Paris *AI Action Summit*.

since 2018
- **Gesellschaft für Informatik** (GI)
- **Universitätsgesellschaft des Saarlandes e.V.**

since 2016
- **Gesellschaft für Utilitarismusstudien** (GUS), *founding member*

since 2015
- **Gesellschaft für Analytische Philosophie** (GAP)

Language Skills & Hobbies

Languages

- German (native), English (fluent)

Hobbies

- streetball, bouldering, hiking, reading (fiction), chess, cats

Research Publications

Journal Articles

1. Schlicker, Nadine, **Baum**, **Kevin**, Uhde, Alarith, Sterz, Sarah, Hirsch, Martin C, and Langer, Markus, "How Do We Assess the Trustworthiness of AI? Introducing the Trustworthiness Assessment Model (TrAM)," *Computers in Human Behavior*, vol. 170, p. 108 671, 2025. DOI: 10.1016/j.chb.2025.108671.

2. Langer, Markus, **Baum**, **Kevin**, and Schlicker, Nadine, "Effective Human Oversight of AI-Based Systems: A Signal Detection Perspective on the Detection of Inaccurate and Unfair Outputs," *Minds and Machines*, vol. 35, no. 1, pp. 1–30, 2024. DOI: 10.1007/s11023-024-09701-0.

3. Biewer, Sebastian, **Baum**, **Kevin**, Sterz, Sarah, Hermanns, Holger, Hetmank, Sven, Langer, Markus, Lauber-Rönsberg, Anne, and Lehr, Franz, "Software Doping Analysis for Human Oversight," *Formal Methods in System Design*, 2024. DOI: 10.1007/s10703-024-00445-2.

4. **Baum**, **Kevin**, Bryson, Joanna, Dignum, Frank, Dignum, Virginia, Grobelnik, Marko, Hoos, Holger, Irgens, Morten, Lukowicz, Paul, Muller, Catelijne, Rossi, Francesca, Shawe-Taylor, John, Theodorou, Andreas, and Vinuesa, Ricardo, "From Fear to Action: AI Governance and Opportunities for All," *Frontiers in Computer Science*, vol. 5, 2023. DOI: 10.3389/fcomp.2023.1210421.

5. **Baum**, **Kevin**, and Sterz, Sarah, "Ethics for Nerds," *The International Review of Information Ethics*, vol. 31, no. 1, 2022, special issue on »Ethics in the Age of Smart Systems«. DOI: 10.29173/irie484.

6. **Baum**, **Kevin**, Mantel, Susanne, Schmidt, Eva, and Speith, Timo, "From Responsibility to Reason-Giving Explainable Artificial Intelligence," *Philosophy & Technology*, vol. 35, no. 1, 2022. DOI: 10.1007/s13347-022-00510-w.

7. Schlicker, Nadine, Langer, Markus, Ötting, Sonja K, **Baum**, **Kevin**, König, Cornelius J, and Wallach, Dieter, "What to Expect from Opening Up 'Black Boxes'? Comparing Perceptions of Justice Between Human and Automated Agents," *Computers in Human Behavior*, vol. 122, 2021. DOI: 10.1016/j.chb.2021.106837.

8. Langer, Markus, Oster, Daniel, Speith, Timo, Hermanns, Holger, Kästner, Lena, Schmidt, Eva, Sesing, Andreas, **Baum**, **Kevin**, "What Do We Want from Explainable Artificial Intelligence (XAI)? – A Stakeholder Perspective on XAI and a Conceptual Model Guiding Interdisciplinary XAI Research," *Artificial Intelligence*, vol. 296, 2021. DOI: 10.1016/j.artint.2021.103473.

9. Langer, Markus, **Baum**, **Kevin**, König, Cornelius J, Hähne, Viviane, Oster, Daniel, and Speith, Timo, "Spare Me the Details: How the Type of Information About Automated Interviews Influences Applicant Reactions," *International Journal of Selection and Assessment*, vol. 29, no. 2, pp. 154–169, 2021. DOI: 10.1111/ijsa.12325.

Conference Proceedings

1. Steingrüber, Andre, **Baum**, **Kevin**, "Justifications for Democratizing AI Alignment and Their Limits," in *Proceedings of the 3rd Artificial Intelligence Symposium on Large-Scale Applications (AISoLA 2025), Rhodes, Greece, Nov 1–5, 2025*, accepted for on-site proceedings, to appear, Springer, Lecture Notes in Computer Science, Nov. 2025. DOI: 10.48550/arXiv.2507.19548.

2. Langer, Markus, Lazar, Veronika, **Baum**, **Kevin**, "On the Complexities of Testing for Compliance with Human Oversight Requirements in AI Regulation," in *Proceedings of the 3rd Artificial Intelligence Symposium on Large-Scale Applications (AISoLA 2025), Rhodes, Greece, Nov 1–5, 2025*, accepted for on-site proceedings, to appear, Springer, Lecture Notes in Computer Science, Nov. 2025. DOI: 10.48550/arXiv.2504.03300.

3 **Baum**, **Kevin**, and Slavkovik, Marija, "Aggregation Problems in Machine Ethics and AI Alignment," in *Proceedings of the 8th AAAI/ACM Conference on AI, Ethics, and Society (AIES 2025), Madrid, Spain, Oct 20–22, 2025*, accepted, to appear, Madrid, Spain: ACM, Oct. 2025.

4 **Baum**, **Kevin**, "Disentangling AI Alignment: A Structured Taxonomy Beyond Safety and Ethics," in *Post-Proceedings of the 2nd Artificial Intelligence Symposium on Large-Scale Applications (AISoLA 2024), Crete, Greece, Oct 30–Nov 3, 2024*, Crete, Greece, Oct. 2025. 🔗 DOI: 10.48550/arXiv.2506.06286.

5 **Baum**, **Kevin**, Dargasz, Lisa, Jahn, Felix, Gros, Timo P., and Wolf, Verena, "Acting for the Right Reasons: Creating Reason-Sensitive Artificial Moral Agents," in *Proceedings of the Formal Ethical Agents and Robots Workshop at the 19th International Conference on Integrated Formal Methods (iFM 2024), Manchester, UK, Nov 11–13, 2024*, Manchester, UK, Nov. 2024. 🔗 DOI: 10.48550/arXiv.2409.15014.

6 Baum, Deborah, **Baum**, **Kevin**, Zamani, Sasha, Bennoit, Christian, and Werth, Dirk, "Transparent Transparency: Developing a Scheme for Understanding Transparency Requirements," in *Bridging the Gap Between AI and Reality*, B. Steffen, Ed., ser. Lecture Notes in Computer Science, Published in the post-proceedings of the 2nd Artificial Intelligence Symposium on Leveraging Applications of Formal Methods, Verification and Validation (AISoLA 2024), Crete, Greece, Oct 30–Nov 3, 2024, vol. 15217, Cham: Springer Nature Switzerland, 2025, pp. 55–73, ISBN: 978-3-031-75434-0. 🔗 DOI: 10.1007/978-3-031-75434-0_5.

7 **Baum**, **Kevin**, Biewer, Sebastian, Hermanns, Holger, Hetmank, Sven, Langer, Markus, Lauber-Rönsberg, Anne, and Sterz, Sarah, "Taming the AI Monster: Monitoring of Individual Fairness for Effective Human Oversight," in *Model Checking Software*, T. Neele and A. Wijs, Eds., ser. Lecture Notes in Computer Science (LNCS), Proceeding of the *30th International Symposium on Model Checking Software* (SPIN 2024) colocated with the *27th European Joint Conferences on Theory and Practice of Software* (ETAPS 2024), vol. 14624, Cham: Springer Nature Switzerland, 2025, pp. 3–25, ISBN: 978-3-031-66149-5. 🔗 DOI: 10.1007/978-3-031-66149-5_1.

8 Sterz, Sarah, **Baum**, **Kevin**, Biewer, Sebastian, Hermanns, Holger, Lauber-Rönsberg, Anne, Meinel, Philip, and Markus, Langer, "On the Quest for Effectiveness in Human Oversight: Interdisciplinary Perspectives," FAccT '24: Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency, Jun. 2024. 🔗 DOI: 10.1145/3630106.3659051.

9 Baum, Deborah, **Baum**, **Kevin**, Gros, Timo P., and Wolf, Verena, "XAI Requirements in Smart Production Processes: A Case Study," in *Explainable Artificial Intelligence. Proceedings of the World Conference on eXplainable Artificial Intelligence (xAI 2023)*, L. Longo, Ed., ser. Communications in Computer and Information Science (CCIS), vol. 1901, Cham: Springer Nature Switzerland, 2023, pp. 3–24. 🔗 DOI: 10.1007/978-3-031-44064-9_1.

10 Langer, Markus, **Baum**, **Kevin**, Hartmann, Kathrin, Hessel, Stefan, Speith, Timo, and Wahl, Jonas, "Explainability Auditing for Intelligent Systems: A Rationale for Multi-Disciplinary Perspectives," in *29th IEEE International Requirements Engineering Conference Workshops (RE 2021 Workshops), Notre Dame, Indiana, USA*, T. Yue and M. Mirakhorli, Eds., IEEE, 2021, pp. 164–168. 🔗 DOI: 10.1109/REW53955.2021.00030.

11 Sterz, Sarah, **Baum**, **Kevin**, Lauber-Rönsberg, Anne, and Hermanns, Holger, "Towards Perspicuity Requirements," in *29th IEEE International Requirements Engineering Conference Workshops (RE 2021 Workshops), Notre Dame, Indiana, USA*, T. Yue and M. Mirakhorli, Eds., IEEE, Sep. 2021, pp. 159–163. 🔗 DOI: 10.1109/REW53955.2021.00029.

12 Köhl, Maximilian A, **Baum**, **Kevin**, Langer, Markus, Oster, Daniel, Speith, Timo, and Bohlender, Dimitri, "Explainability as a Non-Functional Requirement," in *27th IEEE International*

Requirements Engineering Conference (RE 2019), Jeju Island, South Korea, IEEE, 2019, pp. 363–368. DOI: 10.1109/RE.2019.00046.

13 **Baum**, **Kevin**, Hermanns, Holger, and Speith, Timo, “Towards a framework combining machine ethics and machine explainability,” in *Proceedings of the 3rd Workshop on Formal Reasoning about Causation, Responsibility, and Explanations in Science and Technology (CREST 2018), Thessaloniki, Greece, 21st April 2018*, B. Finkbeiner and S. Kleinberg, Eds., 2019. DOI: 10.4204/EPTCS.286.4.

14 **Baum**, **Kevin**, Kirsch, Nadine, Reese, Kerstin, Schmidt, Pascal, Wachter, Lukas, and Wolf, Verena, “Informatikunterricht in der Grundschule? Erprobung und Auswertung eines Unterrichtsmoduls mit Calliope mini,” in *Proceedings of the Informatik für alle, 18. GI-Fachtagung Informatik und Schule (INFOS 2019) in the* GI-Edition: Lecture Notes in Informatics (*LNI*), A. Pasternak, Ed., vol. P-288, Gesellschaft für Informatik, 2019, pp. 49–58. DOI: 10.18420/INFOS2019-B1.

15 Sesing, Andreas, **Baum**, **Kevin**, “Anforderungen an die Erklärbarkeit maschinengestützter Entscheidungen,” in *Die Macht der Daten und der Algorithmen – Regulierung von IT, IoT und KI. Tagungsband DSRI-Herbstakademie 2019*, J. Taeger, Ed., 2019, pp. 435–449. URL: http://olwir.de/?content=reihen/uebersicht&sort=tb&isbn=978-3-95599-061-9.

16 **Baum**, **Kevin**, Hermanns, Holger, and Speith, Timo, “From Machine Ethics To Machine Explainability and Back,” in *International Symposium on Artificial Intelligence and Mathematics (ISAIM 2018), Fort Lauderdale, Florida, USA*, 2018. URL: https://isaim2018.cs.ou.edu/papers/ISAIM2018_Ethics_Baum_etal.pdf.

17 **Baum**, **Kevin**, Köhl, Maximilian, and Schmidt, Eva, “Two Challenges for CI Trustworthiness and How to Address Them,” in *Proceedings of the 1st Workshop on Explainable Computational Intelligence (XCI 2017)*, M. Pereira-Fariña and C. Reed, Eds., Dundee, United Kingdom: Association for Computational Linguistics, Sep. 2017. DOI: 10.18653/v1/W17-3701.

18 **Baum**, **Kevin**, “What the Hack Is Wrong with Software Doping?” In *Proccedings of the 7th International Symposium on Leveraging Applications of Formal Methods: ISoLA 2016: Leveraging Applications of Formal Methods, Verification and Validation: Discussion, Dissemination, Applications*, T. Margaria and B. Steffen, Eds., ser. Lecture Notes in Computer Science (LNCS), Springer International Publishing, vol. 9953, 2016, pp. 633–647. DOI: 10.1007/978-3-319-47169-3_49.

19 Krekhov, Andrey, Grüninger, Jürgen, **Baum**, **Kevin**, McCann, David, and Krüger, Jens, “Morphableui: A hypergraph-based approach to distributed multimodal interaction for rapid prototyping and changing environments,” in *Proceedings of the 24th International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision (WSCG 2016)*, ser. Computer Science Research Notes (CSRN), In cooperation with EUROGRAPHICS, Plzeň, Czech Republic: Václav Skala - UNION Agency, 2016, pp. 299–308. URL: http://wscg.zcu.cz/WSCG2016/!!_CSRN-2602.pdf.

Chapters

1 **Baum**, **Kevin**, “Utilitarismus und das Problem kollektiven Handelns,” in *Handbuch Utilitarismus*, V. Andrić and B. Gesang, Eds., *in print*, J.B. Metzler.

Workshop Abstracts and Extended Abstracts

1 Langer, Markus, Lazar, Veronika, **Baum**, **Kevin**, *How to Test for Compliance with Human Oversight Requirements in AI Regulation?* Accepted abstract at the *Workshop on Sociotechnical AI Governance* (CHI-STAIG 2025) at the *ACM Conference on Human Factors in Computing Systems* (CHI 2025), Honolulu, USA. Presented by Markus Langer. Collaboration with the German Federal Office for Information Security (BSI), 2025. DOI: 10.48550/arXiv.2504.03300.

Preprints and Idea Papers (Unreviewed)

1. Ostermann, Simon, **Baum**, **Kevin**, Endres, Christoph, Masloh, Julia, and Schramowski, Patrick, *Soft Begging: Modular and Efficient Shielding of LLMs against Prompt Injection and Jailbreaking based on Prompt Tuning*, 2024. 🔗 DOI: 10.48550/arXiv.2407.03391.

Submitted and Under Review

1. Jahn, Felix, Muskalla, Yannic, Dargasz, Lisa, **Baum**, **Kevin**, *Breaking Up with Monolithic Agency: A Neuro-Symbolic Reason-Based Architecture for Ethical AI Alignment*, Submitted to the *Special Track on AI Alignment* at the *40th Annual AAAI Conference on Artificial Intelligence* (AAAI-26-AIA), under review., 2025.

2. Illan, Dejanira Araiza, **Baum**, **Kevin**, Beebee, Helen, Chatila, Raja, Christensen, Sarah Moth-Lund, Coghlan, Simon, Collins, Emily Charlotte, Cunha, Alcino, Devitt, Susannah Kate, Dobrosovestnova, Anna, Hein, Duijf, Evers, Vanessa, Fisher, Michael, Kökciyan, Nadin, Hochgeschwender, Nico, Lemaignan, Séverin, Lera, Francisco Javier Rodríguez, Ljungblad, Sara, Magnusson, Martin, Mansouri, Masoumeh, Milford, Michael J., Moon, AJung, Powers, Thomas M., Salvini, Pericle, Scantamburlo, Teresa, Schuster, Nick, Slavkovik, Marija, Topcu, Ufuk, Vanegas, Daniel Fernando Preciado, Wasowski, Andrzej, and Yang, Yi, *A Roadmap for Responsible Robotics*, conditionally accepted at the *IEEE Robotics and Automation Magazine* for the special issue on *Robot Ethics*, 2025.

Other Academic Activities (Selection)

Renowned Workshops and Seminars (by Invitation Only)

1. Dagstuhl Seminar 25272: *Challenges of Human Oversight: Achieving Human Control of AI-Based Systems* , Leibniz Center for Informatics *Schloss Dagstuhl*, June 29–July 04, 2025.

2. Inaugural Conference of the *International Association for Safe and Ethical Artificial Intelligence* (IASEAI '25), OECD La Muette Headquarters and Conference Centre 2 Rue André Pascal, Paris, France, Feb. 6–7, 2025.

3. *Digital Democracy* of the *Interdisciplinary Institute for Societal Computing* (I^{2}SC), Saarland University, Saarbrücken, Oct. 9–10, 2024.

4. *Workshop on Artificial Intelligence* at the *UK-German Science, Innovation and Technology Dialogue*, Imperial College White City Campus, London, UK, March 12, 2024.

5. Dagstuhl Seminar 23371: *Roadmap for Responsible Robotics*, Leibniz Center for Informatics *Schloss Dagstuhl*, Sep. 10–15, 2023.

   (The actual *Roadmap for Responsible Roadmaps* as a joint Dagstuhl publication under the direction of Prof. Michael Fisher is currently being finalized.)

6. Dagstuhl Seminar 23151: *Normative Reasoning for AI*, Leibniz Center for Informatics *Schloss Dagstuhl*, Apr. 10–14, 2023.

7. Dagstuhl Seminar 16222: *Engineering Moral Agents – from Human Morality to Artificial Morality*, Leibniz Center for Informatics *Schloss Dagstuhl*, May 29 – Jun 03, 2016.

Academic Talks (Selection)

1. **Baum**, **Kevin**, *AI Certification as a Deep Socio-Technical Challenge: Toward a Responsible yet Realistic Approach*, Talk at the *International Workshop on AI certification* in context of the Summer School 2025 on IT Law and Legal Informatics, 1st August 2025, Saarbrücken, Germany, Aug. 2025.

2. **Baum**, **Kevin**, Bergs, Richard, Holger Hermanns Sophie Kerstan, Markus Langer, Lauber-Rönsberg, Anne, Meinel, Philip, Laura Stenzel, Sterz, Sarah, and Zhang, Hanwei, *The Principal's Principles: Actionable (Personalized) AI Alignment as Underexplored XAI Application Context*, Accepted talk for the 3rd TRR 318 Conference: Contextualizing Explanations (ContEx25) (presented by Laura Stenzel as backup), 17th and 18th June 2025, Bielefeld, Germany, Mar. 2025.

3. **Baum**, **Kevin**, *Trusted AI beyond Buzzwordiness: Interdisciplinary Approaches to Fairness, Effective Oversight, and Machine Ethics as Personalized Value Alignment*, invited talk at the department seminar of the Institutt for informatikk of the University of Bergen, May 2025. 🔗 URL: https://www.uib.no/ii/178047/trusted-ai-beyond-buzzwordiness-interdisciplinary-approaches-fairness-effective-oversight.

4. **Baum**, **Kevin**, *Making RL-Based AI Agents Do the Right Thing for the Right Reason: Toward Justifiable Moral Alignment*, invited talk to the 35th session of the seminar of the Logic & AI group at the Department of Information Science and Media Studies of the University of Bergen, May 2025.

5. **Baum**, **Kevin**, Dargasz, Lisa, Gros, Timo P., and Jahn, Felix, *Why Did It Do That? Trustworthy Artificial Agents via Justifiability*, talk accepted for the *2nd Artificial Intelligence Symposium On Leveraging Applications of Formal Methods, Verification and Validation* (AISoLA 2024), Crete, Greece, 30.10 – 03.11.2024, 2024.

6 **Baum**, **Kevin**, Schlüter, Maximilian, Schmidt, Eva, and Speith, Timo, *The Reasons of AI Systems*, talk accepted for the *2nd Artificial Intelligence Symposium On Leveraging Applications of Formal Methods, Verification and Validation* (AISoLA 2024), Crete, Greece, 30.10 – 03.11.2024, 2024.

7 **Baum**, **Kevin**, *Current Ethical Challenges of AI*, keynote at the *4th Inria-DFKI European Summer School on AI* (IDESSAI 2024) at *Saarland Informatics Campus*, Saarbrücken, Germany, Sep. 2024. 🔗 URL: https://idessai.eu/.

8 **Baum**, **Kevin**, and Dargasz, Lisa, *A Reason-Responsiveness Approach to Machine Ethics for Reinforcement-Learning Agents*, paper presentation at the international conference on *Formal Ethics 2024* at the *University of Greifswald*, Greifswald, Germany, Jul. 2024. 🔗 URL: https://www.wiko-greifswald.de/formal-ethics-2024/.

9 Sterz, Sarah, **Baum**, **Kevin**, and Meinel, Philip, *On the Quest for Effectiveness in Human Oversight: Interdisciplinary Perspectives*, paper presentation at the *7th ACM Conference on Fairness, Accountability, and Transparency* (FAccT 2024), Rio de Janeiro, Brazil, Jun. 2024. 🔗 URL: https://facctconference.org/2024/.

10 **Baum**, **Kevin**, *Erklärbarkeit und* Effective Human Oversight *im Art. 14 des EU AI Acts*, invited talk at the seminar *KI und der aufgeklärte Nutzer* of the *Law and AI Research Group* (LARG) at *Bucerius Law School*, Hamburg, Germany, Apr. 2024.

11 **Baum**, **Kevin**, *On the Quest for Effectiveness in Human Oversight*, invited talk at the *Institute for Ethics in Technology* at *Hamburg University of Technology* (TUHH), Hamburg, Germany, Apr. 2024.

12 **Baum**, **Kevin**, *Beyond Technicalities: Positive Normative Choices and the Ethical Dimensions of Responsible AI System Design*, at the *1st Artificial Intelligence Symposium On Leveraging Applications of Formal Methods, Verification and Validation* (AISoLA 2023), Crete, Greece, Oct. 2023. 🔗 URL: https://aisola.org/papers/baum.pdf.

13 **Baum**, **Kevin**, Biewer, Sebastian, Hermanns, Holger, Hermank, Sven, Langer, Markus, Lauber-Rönsberg, Anne, Meinel, Philip, and Sterz, Sarah, *Effective Human Oversight: Conditions and Implications of the Proposed EU AI Act from an Interdisciplinary Perspective*, at the *1st Artificial Intelligence Symposium On Leveraging Applications of Formal Methods, Verification and Validation* (AISoLA 2023), Crete, Greece, Oct. 2023. 🔗 URL: https://aisola.org/papers/hermanns-lauber-roensberg-langer-baum-hetmank-sterz-meinel-biewer.pdf.

14 Schlicker, Nadine, **Baum**, **Kevin**, Uhde, Alarith, Sterz, Sarah, Hirsch, Martin C., and Langer, Markus, *How Do We Assess System Trustworthiness? Introducing the Trustworthiness Assessment Model*, at the *1st Artificial Intelligence Symposium On Leveraging Applications of Formal Methods, Verification and Validation* (AISoLA 2023), Crete, Greece, Oct. 2023. 🔗 URL: https://aisola.org/papers/schlicker-baum-uhde-sterz-hirsch-langer.pdf.

15 **Baum**, **Kevin**, *Machine Ethics ⇔ Machine Explainability*, guest lecture and expert panel participation on invitation by Rune Nyrup in context of the seminar »Designing artificial moral agents – could we, and should we?« within the program »AI Ethics and Society« at the *Centre for the Future of Intelligence* (CFLI), Cambridge, UK, Jan. 2022.

16 **Baum**, **Kevin**, Langer, Markus, and Sterz, Sarah, *The Role of XAI for Meta-Trust and Trust Propagation - Psychological and Philosophical Perspectives*, at the *2nd Workshop on Issues in XAI: Understanding and Explaining in Healthcare* at the *Centre for the Future of Intelligence* (CFLI), Cambridge, UK (postponed and moved online due to COVID-19), May 2021. 🔗 URL: http://lcfi.ac.uk/news-and-events/events/issues-explainable-ai-2-understanding-and-explaini/.

17 **Baum**, **Kevin**, and Sterz, Sarah, *Ethics for Nerds: Best Practices for Teaching Ethics to Computer Scientists*, at the *10th Annual Symposium Ethics in the Age of Smart Systems* (online), Apr. 2021. 🔗 URL: https://www.luc.edu/digitalethics/events/symposiaarchive/2021/.

18 **Baum**, **Kevin**, *New Ground for Consequentialism: How to Solve the Coordination Problem*, paper presentation at Thomas Schmidt's *Colloquium in Practical Philosophy* at the Humboldt University (HU), Berlin, Germany (online due to COVID-19), Jan. 2021.

19 **Baum**, **Kevin**, *Verantwortung, Vertrauen, Rechte: Über die ethische Dimension erklärbarer KI*, at the workshop *Verantwortlichkeit digitalisierter Unternehmen – die ethischen und rechtlichen Auswirkungen des Einsatzes künstlicher Intelligenz, Universität Salzburg*, Salzburg, Austria, Nov. 2019. 🔗 URL: https://www.plus.ac.at/wp-content/uploads/2021/02/Programmentwurf.pdf.

20 **Baum**, **Kevin**, *Development or Regulation First*, at the *1st Workshop on Issues in XAI: Blackboxes, Recommendations, and Levels of Explanations*, Saarbrücken, Germany, Oct. 2019. 🔗 URL: https://explainable-intelligent.systems/workshop/.

21 **Baum**, **Kevin**, Kästner, Lena, and Schmidt, Eva, *Understanding Explainable AI: The EIS Project*, at the scientific colloquium of the »Science, Value, and the Future of Intelligence« project at the *Leverhulme Centre for the Future of Intelligence*, Cambridge, UK, Jul. 2019.

22 **Baum**, **Kevin**, and Bräuer, Felix, *Trusting Artificial Experts*, at the workshop *Evidence in Law and Ethics, Jagiellonian University*, Kraków, Poland, Apr. 2019. 🔗 URL: https://incet.uj.edu.pl/evidence-in-law-and-ethics.

23 **Baum**, **Kevin**, and Schmidt, Eva, *Moral Harmony vs. Supervenience: A New Dilemma for Consequentialism*, at the *10th International Congress of the Society for Analytical Philosophy* (GAP 10), Cologne, Germany, Sep. 2018. 🔗 URL: https://gap10.de/wp-content/uploads/2018/09/Programmheft.GAP_.10.Final_.NEU_-2.pdf.

24 **Baum**, **Kevin**, Hermanns, Holger, and Speith, Timo, *Towards a Framework Combining Machine Ethics and Machine Explainability*, 3rd Workshop on Formal Reasoning about Causation, Responsibility, and Explanations in Science and Technology (CREST2018) as part of the 28th European Joint Conferences on Theory and Practice of Software (ETAPS 2018), Thessaloniki, Greece, Apr. 2018. 🔗 URL: https://www.react.uni-saarland.de/crest2018/.

25 **Baum**, **Kevin**, *Obligation Overboard? Decision-Theoretic Objective Consequentialism to the Rescue*, talk as part of a research stay at the *Munich Center for Mathematical Philosophy* (MCMP) at the invitation of Prof. Dr. Stephan Hartmann, Munich, Germany, Jul. 2017.

26 **Baum**, **Kevin**, *What the Hack Is Wrong with Software Doping?* at the *7th International Symposium On Leveraging Applications of Formal Methods, Verification and Validation* (ISOLA 2016), Corfu, Greece, Oct. 2016.

27 **Baum**, **Kevin**, and Schmidt, Eva, *Kagans 'Lösung' des Problems gemeinschaftlichen Handelns: Wie eine verborgene Annahme Kagans Ansatz entwertet*, poster presentation at the *9th International Congress of the Society for Analytical Philosophy* (GAP 9), Osnabrück, Germany, Sep. 2015. 🔗 URL: https://gap9.de/wp-content/uploads/2015/08/Programmheft-GAP.9-final.pdf.

28 **Baum**, **Kevin**, *Asking the Right Questions*, at the seminar *Engineering Moral Agents – from Human Morality to Artificial Morality* on *Schloss Dagstuhl, Leibniz-Zentrum für Informatik*, Wadern, Germany, May 2015. 🔗 URL: https://drops.dagstuhl.de/entities/document/10.4230/DagRep.6.5.114.

Upcoming, Accepted Talks & Presentations

1. **Baum**, **Kevin**, *Doing Wrong with Others: Multi-Agent Consequentialism as a Solution to the Collective Action Problem*, Accepted talk for the 12th International Congress of the Society for Analytic Philosophy (GAP.12), September 2025, Düsseldorf, Germany, 2025.

Publications and Presentations Aimed at Public Audience

Talks and Panel Appearances (Selection)

1. *Zwischen Algorithmus und Abschied: KI, Ethik und die letzten Dinge*, Keynote at the 25th anniversary of the *Landesarbeitsgemeinschaft Hospiz Saarland e.V.*, themed "Gemeinsam in die Zukunft – Zwischen Menschlichkeit und Algorithmus", Jun. 2025. 🔗 URL: https://www.hospiz-saarland.de/aktuelles/veranstaltungen-und-termine/detailansicht-event/25-jahre-lag-hospiz-saarland-e-v-1.
2. *Gute KI, böse KI? Von Mensch, Maschine und Moral*, Talk at the lecture series of the *Volkshochschule Regionalverband Saarbrücken*, in cooperation with the *German Research Center for Artificial Intelligence* (DFKI), Jun. 2025. 🔗 URL: https://www.vhs-saarbruecken.de/programm/gesellschaft/kurs/Gute-KI-boese-KI-Von-Mensch-Maschine-und-Moral/AR4508.
3. **Baum**, **Kevin**, *Vertrauenswürdige KI: Aktuelle Herausforderungen als Chancen für europäische Innovation*, Keynote at the »ODDO BHF Tech & Taste: Eine Reise durch Gourmet und Technologie«, Dec. 2024.
4. **Baum**, **Kevin**, *Vertrauenswürdige KI: Europas (letzte?) Chance auf eine KI-Vorreiterrolle*, Keynote at the »Handelsblatt Summit Künstliche Intelligenz 2024«, Nov. 2024. 🔗 URL: https://live.handelsblatt.com/event/handelsblatt-summit-kuenstliche-intelligenz/referenten/.
5. **Baum**, **Kevin**, *Vertrauenswürdige KI: Was ist das und worauf kommt es an?* Keynote in the event series »AI Insights – Einblicke in die Vielfalt der Künstlichen Intelligenz« of the *Bundesamt für Sicherheit in der Informationstechnik* (BSI) and the *Landesmedienanstalt Saarland* (LMS), Oct. 2024.
6. **Baum**, **Kevin**, *Digitalisierung und KI in der Medizin - Chancen und Risiken für Arzt und Patient*, keynote at the *16. Saarländischer Fachärztetag*, Saarbrücken, Deutschland, Jun. 2024. 🔗 URL: https://www.facharztforum-saar.de/.cm4all/uproc.php/0/Einladung%20web.pdf.
7. **Baum**, **Kevin**, *Trusted AI: Von Grundlagenforschung zu echtem Impact*, keynote at the visit of the *Baden-Badener Unternehmer Gespräche e.V.* in Saarbrücken, Deutschland, Jun. 2024.
8. **Baum**, **Kevin**, *Trusted AI: Künstliche Intelligenz, der man vertrauen darf*, research keynote at the *Journalismuspreis Informatik 2023* of the *Ministerium für Wirtschaft, Innovation, Digitales und Energie des Saarlandes* at the *Saarland Informatics Campus*, Saarbrücken, Germany, May 2024.
9. **Baum**, **Kevin**, *Trusted AI as a Driving Force for Societally Beneficial Innovations*, keynote at the *1st Industry Collaboration and Transfer Exchange – From Research to Market*, Saarbrücken, Germany, May 2024.
10. **Baum**, **Kevin**, *Ethik und KI: Eine Balance zwischen Fortschritt und Verantwortung?* talk and expert panel participation at the *Afterwork-Event „Business-KI-Balance" @CFK/GymLodge* event in context of the *European Digital Innovation Hub Saarland* (EDIH), Spiesen-Elversberg, Germany, Mar. 2024.

11 **Baum**, **Kevin**, *AI Safety and Security in Practice*, talk as part of the visit of the »Trade Mission to Saarland« of the *Luxembourg Chamber of Commerce* at the invitation of the *Saarland Ministry of Economics, Innovation, Digital Affairs and Energy*, Mar. 2024. 🔗 URL: https://app.swapcard.com/event/mission-germany/exhibitors/RXZlbnRWaWV3XzY2Nzg3MQ==.

12 **Baum**, **Kevin**, *Künstliche Intelligenz im Betrieb: Ethische Fallstricke*, talk at »Technologiekonferenz« of the *Beratungsstelle für sozialverträgliche Technologiegestaltung e.V.* (BEST) and the *Arbeitskammer des Saarlandes*, Dec. 2023. 🔗 URL: https://www.rechtsschutzsaal.de/aktuelles-termine/meldung-1/technologiekonferenz-best-ev-im-rechtsschutzsaal-bildstock.

13 **Baum**, **Kevin**, *KI – Fluch oder Segen?* Expert panel participation in the context of the »Der Fabulant« project of *modus|zad*, Dec. 2023. 🔗 URL: https://modus-zad.de/publikation/blog/online-veranstaltung-ki-fluch-oder-segen/.

14 **Baum**, **Kevin**, *All Questions Answered (AQuA): Get ready for the AI Act. Implications for Researchers, AI, Data & Robotics.* expert panel participation organized by *Confederation of Laboratories for Artificial Intelligence Research in Europe* (CLAIRE), Oct. 2023. 🔗 URL: https://cairne.eu/portfolio-items/25-10-23-aqua-get-ready-for-the-ai-act-implications-for-researchers-ai-data-robotics/.

15 **Baum**, **Kevin**, *KI und deren Bedeutung für Wissenschaft und Hochschulen*, expert panel participation in the context of the »Studium Generale« at *Bucerius Law School*, Jun. 2023. 🔗 URL: https://www.law-school.de/news-artikel/expertengespraech-zu-ki-und-die-bedeutung-fuer-wissenschaft-und-hochschulen.

16 **Baum**, **Kevin**, *Künstliche Intelligenz und Verantwortung*, keynote at the *Paul Fritsche Stiftung Wissenschaftliches Form*, Homburg, Germany, Dec. 2022. 🔗 URL: https://www.uniklinikum-saarland.de/de/lehre/dekanat/wissenschaftliche_foren_gastvortraege.

17 **Baum**, **Kevin**, *AI. Friend, Foe or Fad?* Expert panel participation in the context of the *Booster Conference 2024* in Bergen, Norway, Apr. 2022. 🔗 URL: https://2022.boosterconf.no/talk/panel-ai-friend-foe-or-fad/.

18 **Baum**, **Kevin**, *Algorithmen, Fake News & Fragmentierung*, talk at »Safer Internet Day 2021«, Feb. 2021. 🔗 URL: https://www.onlinerlandsaar.de/wp-content/uploads/2021/01/safer_internet_day_2021_X4.pdf.

19 **Baum**, **Kevin**, *Nudging, Manipulation, Profiling*, talk at »Aktionstage Netzpolitik & Demokratie 2020« of the *Zentralen für politische Bildung (ZpB)*, Nov. 2020. 🔗 URL: https://www.youtube.com/watch?v=PM8r70hJZnI.

20 **Baum**, **Kevin**, *Moralische Hürden von Profiling und die Herausforderung der Erklärbarkeit*, talk at the event »Profiling 2.0« organized by the Thuringian State Commissioner for Data Protection and Freedom of Information (TLfDI), Oct. 2020. 🔗 URL: https://www.onlinerlandsaar.de/wp-content/uploads/2021/01/safer_internet_day_2021_X4.pdf.

21 **Baum**, **Kevin**, *Fünf ethische Herausforderungen im Zeitalter der digitalisierten Medizin*, keynote (»Festvortrag«) at the *Eröffnung des Fortbildungsjahres 2020 / 2021* of the *Ärztekammer des Saarlandes*, Homburg, Germany, Sep. 2020. 🔗 URL: https://www.aerztekammer-saarland.de/index/news/News-20200824-Eroeffnung-Fortbildungsjah/.

22 **Baum**, **Kevin**, *Wenn Computer die Lebenszeit vorhersagen: Autonomie, Verantwortung und Erklärbarkeit*, talk at the »Hospizgespräch« at St. Jakobus Hospiz., Jun. 2019. 🔗 URL: https://www.kinderhospizdienst-saar.de/uploads/media/18-05-28_Einladung_JUNI_01.pdf.

23 **Baum, Kevin**, *Aber warum, Computer? Erklär' es mir! Von rassistischen, undurchschaubaren künstlichen Intelligenzen, die vielleicht morgen schon über Ihr Leben bestimmen und von der Hilflosigkeit von Menschen in Schleifen*, talk at the »Futurologischer Kongress« of the *Stadttheater Ingolstadt.*, Jun. 2018. 🔗 URL: https://theater.ingolstadt.de/fileadmin/doc/dokumentation_futurologischer_kongress/futurologischer_kongress_2018.pdf.

(Co-)Organizer of Events (Selection)

1 *TEDxSaarbrigge: Digital Democracy*, first TEDxSaarbrigge event, organized by Ingmar Weber/the *Interdisciplinary Institute for Societal Computing* (I2SC), the non-profit association *Algoright e.V.*, and the *German Research Center for Artificial Intelligence* (DFKI), Jun. 2025. 🔗 URL: https://www.ted.com/tedx/events/62558.

2 *Gesundheitsversorgung: Was KI heute kann und morgen leisten könnte*, fourth joint event of the *Landeszentrale für politische Bildung des Saarlandes* (LpB), the non-profit association *Algoright e.V.*, the *German Research Center for Artificial Intelligence* (DFKI), this time together with the *Federal Office for Information Security* (BSI) and the Health.AI Hub and the Health.AI ELI project., Mar. 2025.

3 *Are AI Minds Minds?* Co-organizer and co-moderator of an expert panel (with Eva Schmidt, Karina Vold, Jan Broersen, and Hans-Johann Glock) at the *2nd Artificial Intelligence Symposium On Leveraging Applications of Formal Methods, Verification and Validation* (AISoLA 2024), Crete, Greece, 30.10 – 03.11.2024., Nov. 2024. 🔗 URL: https://programme.paperplane.services/session/102.

4 *Algorithmische Entscheidung und menschliche Selbstbestimmung*, co-organizer and expert panel participation at the first event of the series »I2SC Goes to Town« of the *Interdisciplinary Institute for Societal Computing* (I2SC), Jul. 2024. 🔗 URL: https://www.i2sc.net/events/i2sc-goes-to-town.

5 *Zukunft Bildung: KI im Klassenzimmer und Hörsaal*, co-organizer of the third joint event of the *Landeszentrale für politische Bildung des Saarlandes* (LpB), the non-profit association *Algoright e.V.*, the *German Research Center for Artificial Intelligence* (DFKI) with, among others, the *Saarland Minister of Education Christine Streichert-Clivot*, Jun. 2024. 🔗 URL: https://eveeno.com/ki-im-klassenzimmer.

6 *Automatisierte Aufmerksamkeit: Wie KI Journalismus, Öffentlichkeit und Meinungsbildung prägt*, co-organizer of the second joint event of the *Landeszentrale für politische Bildung des Saarlandes* (LpB), the non-profit association *Algoright e.V.*, the *German Research Center for Artificial Intelligence* (DFKI) and the *Saarland State Media Authority* (LMS)., Nov. 2023. 🔗 URL: https://eveeno.com/automatisierte_aufmerksamkeit.

7 *Entmystifizierung eines Hypes: ChatGPT zum Anfassen*, co-organizer of the first joint event of the *Landeszentrale für politische Bildung des Saarlandes* (LpB), the non-profit association *Algoright e.V.*, the *German Research Center for Artificial Intelligence* (DFKI) and the *General Studies Committee* (AStA) of *Saarland University*, including workshop on AI-powered disinformation together with Katharina Nocun., May 2023. 🔗 URL: https://eveeno.com/ChatGPT_zum_anfassen.

8 *Synergien oder Konflikt? Forschung, Regulierung und Innovation im Spannungsverhältnis am Beispiel Künstlicher Intelligenz*, co-organizer and co-organizer and expert panel participation, May 2023. 🔗 URL: https://www.saarnews.com/podiumsdiskussion-in-saarbruecken-kuenstliche-intelligenz-an-der-schnittstelle-von-wissenschaft-politik-und-wirtschaft/.

References

Prof. Dr. Marija Slavkovik

Professor for Artificial Intelligence

Postboks 7802
Universitetet i Bergen (UiB)
5020 Bergen, Norway

E-mail: marija.slavkovik@uib.no

Web: https://slavkovik.com/

Prof. Dr. Sebastian Vollmer

Professor for Applied Machine Learning

Fachbereich Informatik
Informatik in Kaiserslautern
Gottlieb-Daimler-Straße
Building 48
67663 Kaiserslautern

E-Mail: s.vollmer@rptu.de

Web: https://sebastian.vollmer.ms

Prof. Elijah Millgram

Distinguished Professor in Philosophy

Philosophy Department
215 Central Campus Dr, Room 467
University of Utah, Salt Lake City, USA

E-mail: Lije.Millgram@m.cc.utah.edu

Web: https://www.elijahmillgram.net/

Prof. Dr. Markus Langer

Professor of Work and Organizational Psychology

Albert-Ludwigs-Universität Freiburg
Institut für Psychologie
Engelbergerstraße 41
79085 Freiburg

E-Mail: Markus.Langer@psychologie.uni-freiburg.de

Web: https://www.psychologie.uni-freiburg.de/Members/langer/langer.html

Prof. Dr. Anne Lauber-Rönsberg

Professor of Civil Law, Intellectual Property Law, in particular Copyright Law, as well as Media and Data Protection Law

Institut für Internationales Recht, Geistiges Eigentum
und Technikrecht (IRGET)
Philosophische Fakultät der TU Dresden
01062 Dresden

E-Mail: office.lauber-roensberg@tu-dresden.de

Web: https://tu-dresden.de/gsw/phil/irget/jfbimd13/die-professur/prof-dr-anne-lauber-roensberg