

Kevin Baum, M.Sc., M.A.

🏠 Martin-Luther-Str. 28, 66111 Saarbrücken, Saarland, Germany

📅 28.07.1986, Schwetzingen

📞 +49 173 7888 380

✉ academia@kevinbaum.de

🌐 <https://kevinbaum.de>

👤 Scholar

🌐 LinkedIn

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

- 12/2023 – today 📌 **Head of the Center for European Research in Trusted AI (CERTAIN).**
German Research Center for Artificial Intelligence (DFKI), Saarbrücken, Saarland.
- 01/2023 – today 📌 **Deputy Head and Manager of the Research Department for Neuro-Mechanistic Modeling (NMM).**
German Research Center for Artificial Intelligence (DFKI), Saarbrücken, Saarland.
- 02/2015 – 12/2022 📌 **Research Assistant and Lecturer.**
Group for Dependable Systems and Software (Prof. Hermanns), since 2018 in project Explainable Intelligent Systems, funded by VolkswagenStiftung, Saarland University (UdS), Saarbrücken, Saarland.
- 📌 **Research Assistant and Lecturer.**
Chair for Practical Philosophy (Prof. Wessels, Prof. Fehige), since 2018 in project Explainable Intelligent Systems, funded by VolkswagenStiftung, Saarland University (UdS), Saarbrücken, Saarland.

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.* Defended, publication pending.
- 2014 📌 **M.A. Philosophy** at Saarland University 1.1
Thesis title: *Vom Bezug singularer Terme in Aussagen über propositionale Einstellungen.*
- 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**

Teaching

Summer 2024	<p>»Machine Ethics. A Tutorial for Prospective Researchers« (upcoming) with Marija Slavkovik and Louise Dennis. <i>33rd International Joint Conference on Artificial Intelligence (IJCAI 2024)</i>, Jeju Island, South Korea.</p> <p>»Practical AI Ethics« (upcoming) with Sarah Sterz. KI Campus. (online course, in preparation)</p>
Summer 2023	<p>»AI for the Social Good« (Seminar) with Dr. Gerrit Großmann, Lisa Dargasz, Sarah Sterz, Prof. Verena Wolf. Saarland Informatics Campus, UdS. 7 ECTS</p> <p>»Ethics for Nerds« (Lecture) with Sarah Sterz, Prof. Holger Hermanns. Saarland Informatics Campus, UdS. 6 ECTS</p>
Summer 2022	<p>»Computer Ethics« (Lecture) with Sarah Sterz. LL.M. <i>Informationstechnologie und Recht</i>, UdS. 3 ECTS</p> <p>»Ethics for Nerds« (Lecture) with Sarah Sterz, Prof. Holger Hermanns. UdS. 6 ECTS</p>
Winter 2021/2022	<p>»Ethische Fragen des Technologieeinsatzes« (Seminar). Bucerius Law School, Hamburg. 2 ECTS</p>
Summer 2021	<p>»Computer Ethics for IT & Law« (Lecture) with Sarah Sterz. LL.M. <i>Informationstechnologie und Recht</i>, UdS. 3 ECTS</p> <p>»Ethics for Nerds« (Lecture) with Sarah Sterz, Prof. Holger Hermanns. Saarland Informatics Campus, UdS. 6 ECTS</p>
Winter 2020/2021	<p>»Ethische Fragen des Technologieeinsatzes« (Seminar). Bucerius Law School, Hamburg. 2 ECTS</p>
Summer 2020	<p>»Ethics for Nerds« (Lecture) with Sarah Sterz, Prof. Holger Hermanns, Saarland Informatics Campus, UdS. 6 ECTS</p>
Summer 2019	<p>»Ethics for Nerds« (Lecture) with Sarah Sterz, Prof. Holger Hermanns, Saarland Informatics Campus, UdS. 6 ECTS</p>
Winter 2018/2019	<p>»Computer sagt: „wahrscheinlich“. Wie KI und Algorithmen unsere Welt verändern« (Seminar). College of Fine Arts (Universität der Künste), Berlin. 2 ECTS</p> <p>»Weapons of Math Destruction – Wie Big Data Ungerechtigkeit fördert und Demokratie gefährdet« (Seminar). Philosophy Department, UdS. 6 ECTS</p>
Summer 2018	<p>»Ethics for Nerds« (Lecture) with Prof. Holger Hermanns. Saarland Informatics Campus, UdS. 6 ECTS</p> <p>»Derek Parfits Praktische Philosophie in Reasons & Persons« (Seminar). Philosophy Department, UdS. 6 ECTS</p>
Winter 2017/2018	<p>»Das Problem gemeinschaftlichen Handels« (Seminar). Philosophy Department, UdS. 6 ECTS</p>
Summer 2017	<p>»Moral und Algorithmen« (Seminar). Philosophy Department, UdS. 6 ECTS</p> <p>»Ethics for Nerds« (Lecture) with Prof. Holger Hermanns. Saarland Informatics Campus, UdS. 6 ECTS</p>
Winter 2017/2018	<p>»Von der Entscheidungstheorie zum Konsequentialismus« (Seminar). Philosophy Department, UdS. 6 ECTS</p>

Teaching (continued)

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

Supervision and Review of Student Theses, Student Advising

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

- 2024 [A thesis on machine ethics, upcoming]. Master's Thesis of Lisa Dargasz (**M.Sc.**).
- 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.**).
- 2018 *Another Take on Newcomb's Problem*. Bachelor's Thesis of Tim Dirk Metzler (**B.A.**).

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

- 2018-2023 *Building Bridges for Better Machines: From Machine Ethics to Machine Explainability and Back*. Doctoral Thesis of Timo Speith (**Dr. Phil**).
- 2020 *Gibt es Umstände, unter denen Abtreibung moralisch erlaubt ist?* State examination thesis of Alisha Monika Mateas (**StEx**).
- 2018 *A Framework of Verifiable Machine Ethics and Machine Explainability*. Master's Thesis of Timo Speith (**M.Sc.**).
- 2017-2018 *Schuld der Massen – Kollektive Verantwortung und das Überdeterminiertheitsproblem*. Master's Thesis of Yannik Bast (**M.A.**).
- 2017 *Klimawandel und Spieltheorie: Ist kooperatives Verhalten der Staaten in der Klimapolitik rational?* Bachelor's Thesis of Tessy Aulner (**B.A.**).
- Warum ich einen Unterschied mache. Zur Ethik individuellen Handelns im gesellschaftlichen Handeln*. Bachelor's Thesis of Sarah Maurer (**M.A.**).

Co-Supervised Research Immersion Labs

- 2024 Lena Marie Budde: [A research immersion lab on mechanistic interpretability and explainable AI, upcoming].
- 2023 Janine Lohse: *Predicting and Manipulating Game Rules through Weight Analysis in Neural Networks?*

Research Publications

Journal Articles

- 1 S. Biewer, K. Baum, S. Sterz, *et al.*, “Software Doping Analysis for Human Oversight,” *Formal Methods in System Design*, 2024. [DOI: 10.1007/s10703-024-00445-2](#).
- 2 K. Baum, J. Bryson, F. Dignum, *et al.*, “From Fear to Action: AI Governance and Opportunities for All,” *Frontiers in Computer Science*, vol. 5, 2023. [DOI: 10.3389/fcomp.2023.1210421](#).
- 3 K. Baum and S. Sterz, “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](#).
- 4 K. Baum, S. Mantel, E. Schmidt, and T. Speith, “From Responsibility to Reason-Giving Explainable Artificial Intelligence,” *Philosophy & Technology*, vol. 35, no. 1, 2022. [DOI: 10.1007/s13347-022-00510-w](#).
- 5 N. Schlicker, M. Langer, S. K. Ötting, K. Baum, C. J. König, and D. Wallach, “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](#).
- 6 M. Langer, D. Oster, T. Speith, *et al.*, “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](#).
- 7 M. Langer, K. Baum, C. J. König, V. Hähne, D. Oster, and T. Speith, “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 S. Sterz, K. Baum, S. Biewer, *et al.*, “On the Quest for Effectiveness in Human Oversight: Interdisciplinary Perspectives,” accepted for the *7th ACM Conference on Fairness, Accountability, and Transparency (FAccT ’24)*, 2024. [DOI: 10.48550/arXiv.2404.04059](#).
- 2 D. Baum, K. Baum, T. P. Gros, and V. Wolf, “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](#).
- 3 M. Langer, K. Baum, K. Hartmann, S. Hessel, T. Speith, and J. Wahl, “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](#).
- 4 S. Sterz, K. Baum, A. Lauber-Rönsberg, and H. Hermanns, “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](#).

- 5 M. A. Köhl, K. Baum, M. Langer, D. Oster, T. Speith, and D. Bohlender, “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](https://doi.org/10.1109/RE.2019.00046).
- 6 K. Baum, H. Hermanns, and T. Speith, “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](https://doi.org/10.4204/EPTCS.286.4).
- 7 K. Baum, N. Kirsch, K. Reese, P. Schmidt, L. Wachter, and V. Wolf, “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](https://doi.org/10.18420/INFOS2019-B1).
- 8 A. Sasing and K. Baum, “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](http://olwir.de/?content=reihen/uebersicht&sort=tb&isbn=978-3-95599-061-9).
- 9 K. Baum, H. Hermanns, and T. Speith, “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](https://isaim2018.cs.ou.edu/papers/ISAIM2018_Ethics_Baum_etal.pdf).
- 10 K. Baum, M. Köhl, and E. Schmidt, “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](https://doi.org/10.18653/v1/W17-3701).
- 11 K. Baum, “What the Hack Is Wrong with Software Doping?” In *Proceedings of the 7th International Symposium on Leveraging Applications of Formal Methods: ISOFA 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](https://doi.org/10.1007/978-3-319-47169-3_49).
- 12 A. Krekhov, J. Grüniger, K. Baum, D. McCann, and J. Krüger, “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) in co-operation with EUROGRAPHICS*, ser. Computer Science Research Notes (CSRN), Václav Skala-UNION Agency, 2016, pp. 299–308. [URL: http://wscg.zcu.cz/WSCG2016/!!_CSRN-2602.pdf](http://wscg.zcu.cz/WSCG2016/!!_CSRN-2602.pdf).

Chapters

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









Other Academic Activities (Selection)

Renowned Workshops and Seminars (by Invitation Only)

- 1 Workshop on Artificial Intelligence at the UK-German Science, Innovation and Technology Dialogue, Imperial College White City Campus, London, UK, March 12, 2024.
- 2 Dagstuhl Seminar 23371: *Roadmap for Responsible Robotics*, Leibniz Center for Informatics Schloss Dagstuhl, Sep 10 – Sep 15, 2023.
(The actual *Roadmap for Responsible Roadmaps* as a joint Dagstuhl publication under the direction of Prof. Michael Fisher is currently being finalized.)
- 3 Dagstuhl Seminar 23151: *Normative Reasoning for AI*, Leibniz Center for Informatics Schloss Dagstuhl, Apr 10 – Apr 14, 2023.
- 4 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 K. Baum, *Erklärbarkeit und Effective Human Oversight im Art. 14 des EU AI Acts*, 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.
- 2 K. Baum, *On the Quest for Effectiveness in Human Oversight*, talk at the *Institute for Ethics in Technology* at Hamburg University of Technology (TUHH), Hamburg, Germany, Apr. 2024.
- 3 K. Baum, *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](https://aisola.org/papers/baum.pdf).
- 4 K. Baum, S. Biewer, H. Hermanns, et al., *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](https://aisola.org/papers/hermanns-lauber-roensberg-langer-baum-hetmank-sterz-meinel-biewer.pdf).
- 5 N. Schlicker, K. Baum, A. Uhde, S. Sterz, M. C. Hirsch, and M. Langer, *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](https://aisola.org/papers/schlicker-baum-uhde-sterz-hirsch-langer.pdf).
- 6 K. Baum, *Machine Ethics ⇔ Machine Explainability*, guest lecture and expert panel participation on invitation by Rune Nystrup 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.
- 7 K. Baum, M. Langer, and S. Sterz, *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-explain/](http://lcfi.ac.uk/news-and-events/events/issues-explainable-ai-2-understanding-and-explain/).

- 8 K. Baum and S. Sterz, *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/digitaletics/events/symposiaarchive/2021/>.
- 9 K. Baum, *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.
- 10 K. Baum, *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>.
- 11 K. Baum, *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/>.
- 12 K. Baum, L. Kästner, and E. Schmidt, *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.
- 13 K. Baum and F. Bräuer, *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>.
- 14 K. Baum and E. Schmidt, *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.
- 15 K. Baum, H. Hermanns, and T. Speith, *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/>.
- 16 K. Baum, *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.
- 17 K. Baum, *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.
- 18 K. Baum and E. Schmidt, *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>.
- 19 K. Baum, *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, Invitation Accepted

- 1 S. Sterz, K. Baum, and P. Meinel, *On the Quest for Effectiveness in Human Oversight: Interdisciplinary Perspectives*, accepted for the 7th ACM Conference on Fairness, Accountability, and Transparency (FAccT 2024), Rio de Janeiro, Brazil, Jun. 2024. [URL: https://facctconference.org/2024/](https://facctconference.org/2024/).
- 2 K. Baum and L. Dargasz, *A Reason-Responsiveness Approach to Machine Ethics for Reinforcement-Learning Agents*, accepted for 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/](https://www.wiko-greifswald.de/formal-ethics-2024/).

Publications and Presentations Aimed at Public Audience

Publications (Selection)

- 1 K. Baum and S. Sterz, "Ethics for Nerds: Verantwortung für die Informatiker:innen von Morgen," *Mitteilungen des Fachverbandes Philosophie*, vol. 63, pp. 72–83, 2023. [URL: https://www.fv-philosophie.de/wp-content/uploads/2024/02/Mitteilungen-2023-Hompape-mit-kommentiertem-Teilabdruck-der-Beitraege.pdf](https://www.fv-philosophie.de/wp-content/uploads/2024/02/Mitteilungen-2023-Hompape-mit-kommentiertem-Teilabdruck-der-Beitraege.pdf).
- 2 K. Baum, "Sicherheitsrisiko KI," *IMPULSE – Das Magazin der VolkswagenStiftung*, pp. 38–39, 2022. [URL: https://www.volkswagenstiftung.de/de/news/aktuelles/sicherheitsrisiko-ki](https://www.volkswagenstiftung.de/de/news/aktuelles/sicherheitsrisiko-ki).
- 3 K. Baum, "Moral für Maschinen?" *OPUS Kulturmagazin*, no. 85, 2021. [URL: https://opus-kulturmagazin.de/moral-fuer-maschinen/](https://opus-kulturmagazin.de/moral-fuer-maschinen/).
- 4 M. Schillo and K. Baum, "»Have you ever questioned the nature of your reality?« — Westworld aus der Perspektive der Philosophie," in *Folge um Folge. Multiple Perspektiven auf die Fernsehserie*, Universitätsverlag Hildesheim, 2020, ch. 13. [DOI: 10.18442/160](https://doi.org/10.18442/160).
- 5 K. Baum, T. Feldkamp, S. Henning, and N. Käfer, "Algorithmen und die Grenzen von Fairness," in *Unternehmensverantwortung im digitalen Wandel*, Bertelsmann Stiftung and Wittenberg-Zentrum für Globale Ethik, Ed., Verlag Bertelsmann Stiftung, 2019, ch. 4. [DOI: 10.11586/2020063](https://doi.org/10.11586/2020063).
- 6 M. Langer, K. Baum, and C. J. König, "Die (un-)nachvollziehbarkeit algorithmenbasierter entscheidungen: Implikationen und empfehlungen für die zukunft," in *Mensch und Gesellschaft im digitalen Wandel*, Deutscher Psychologen Verlag, 2018, pp. 32–38, ISBN: 978-3-942761-47-5. [DOI: 10.22028/D291-31275](https://doi.org/10.22028/D291-31275).
- 7 M. Langer, K. Baum, and C. J. König, "Algorithmen bei der Personalauswahl – eine kritische und hoffnungsvolle Betrachtung," *Wirtschaftspsychologie aktuell*, 1st ser., vol. 25, pp. 36–42, 2018, ISSN: 1611-9207. [DOI: 10.22028/D291-31272](https://doi.org/10.22028/D291-31272).

Talks and Panel Appearances (Selection)

- 1 K. Baum, *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.
- 2 K. Baum, *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.

- 3 K. Baum, *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.
- 4 K. Baum, *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==](https://app.swapcard.com/event/mission-germany/exhibitors/RXZlbnRWaWV3XzY2Nzg3MQ==).
- 5 K. Baum, *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](https://www.rechtsschutzsaal.de/aktuelles-termine/meldung-1/technologiekonferenz-best-ev-im-rechtsschutzsaal-bildstock).
- 6 K. Baum, *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/](https://modus-zad.de/publikation/blog/online-veranstaltung-ki-fluch-oder-segen/).
- 7 K. Baum, *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](https://www.law-school.de/news-artikel/expertengespraech-zu-ki-und-die-bedeutung-fuer-wissenschaft-und-hochschulen).
- 8 K. Baum, *Künstliche Intelligenz und Verantwortung*, keynote at the *Paul Fritsche Stiftung Wissenschaftliches Forum*, Homburg, Germany, Dec. 2022. [URL: https://www.uniklinikum-saarland.de/de/lehre/dekanat/wissenschaftliche_foren_gastvortraege](https://www.uniklinikum-saarland.de/de/lehre/dekanat/wissenschaftliche_foren_gastvortraege).
- 9 K. Baum, *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/](https://2022.boosterconf.no/talk/panel-ai-friend-foe-or-fad/).
- 10 K. Baum, *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](https://www.onlinerlandsaar.de/wp-content/uploads/2021/01/safer_internet_day_2021_X4.pdf).
- 11 K. Baum, *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](https://www.youtube.com/watch?v=PM8r70hJZnI).
- 12 K. Baum, *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](https://www.onlinerlandsaar.de/wp-content/uploads/2021/01/safer_internet_day_2021_X4.pdf).
- 13 K. Baum, *Fünf ethische Herausforderungen im Zeitalter der digitalisierten Medizin*, keynote (»Festvortrag«) at the *Eröffnung des Fortbildungsjahres 2020 / 2021* of the *Ärzttekammer des Saarlandes*, Homburg, Germany, Sep. 2020. [URL: https://www.aerztekammer-saarland.de/index/news/News-20200824-Eroeffnung-Fortbildungsjah/](https://www.aerztekammer-saarland.de/index/news/News-20200824-Eroeffnung-Fortbildungsjah/).
- 14 K. Baum, *Wenn Computer die Lebenszeit vorhersagen: Autonomie, Verantwortung und Erklärbarkeit*, talk at the »Hospizgespräch« at *St. Jakobus Hospiz*, Jun. 2019. [URL: https://www.st-jakobus-hospiz.de/](#)

https://www.kinderhospizdienst-saar.de/uploads/media/18-05-28_Einladung_JUNI_01.pdf.

- 15 K. Baum, *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](https://theater.ingolstadt.de/fileadmin/doc/dokumentation_futurologischer_kongress/futurologischer_kongress_2018.pdf).

(Co-)Organizer of Events (Selection)

- 1 *Automatisierte Aufmerksamkeit: Wie KI Journalismus, Öffentlichkeit und Meinungsbildung prägt*, 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](https://eveeno.com/automatisierte_aufmerksamkeit).
- 2 *Entmystifizierung eines Hypes: ChatGPT zum Anfassen*, 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., May 2023. [URL: https://eveeno.com/ChatGPT_zum_anfassen](https://eveeno.com/ChatGPT_zum_anfassen).
- 3 *Synergien oder Konflikt? Forschung, Regulierung und Innovation im Spannungsverhältnis am Beispiel Künstlicher Intelligenz*, expert panel, May 2023. [URL: https://www.saarnews.com/podiumsdiskussion-in-saarbruecken-kuenstliche-intelligenz-an-der-schnittstelle-von-wissenschaft-politik-und-wirtschaft/](https://www.saarnews.com/podiumsdiskussion-in-saarbruecken-kuenstliche-intelligenz-an-der-schnittstelle-von-wissenschaft-politik-und-wirtschaft/).

Upcoming, Invitation Accepted

- 1 K. Baum, *Trusted AI – An Opportunity for German-French Startups*, expert panel participation at the Viva Technology 2024, Paris, France, May 2024.
- 2 K. Baum, *Digitalisierung und KI in der Medizin - Chancen und Risiken für Arzt und Patient*, keynote at the 16. Saarländischer Fachärztetag, Saarbrücken, Deutschland, Jun. 2024.
- 3 *KI und Bildung: Der Stand der Dinge im Frühjahr 2024*, 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.

Scientific Services & Science Infrastructure

(Co-)Organizing of Workshops, & Special Tracks, Tutorials

- 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** of »Machine Ethics« at the *33rd International Joint Conference on Artificial Intelligence (IJCAI 2024)*, 3 August – 9 August, 2024, Jeju, South Korea.

Scientific Services & Science Infrastructure (continued)

- 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 2024), 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

Program Committee Memberships

- 2024
- **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 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 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 workshop on »Contesting AI« at the *33rd International Joint Conference on Artificial Intelligence* (IJCAI 2024), 3 August – 9 August, 2024, Jeju, South Korea.
 - **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, Philosophy & Technology, Big Data & Society, Communications of the ACM

Ethics Committees, Commissions, Advisory Functions






- 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.

Scientific Services & Science Infrastructure (continued)

- 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

Memberships

- since 2019  **Algoright e.V.** (non-profit/»gemeinnützig«), *founding member*
- 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