

BEBA CIBRALIC, PH.D.

Areas of Specialization: Philosophy of AI, Social and Political Philosophy, Applied Ethics

Areas of Competence: Ethics, Epistemology, Philosophy of Language, Public Policy

EMPLOYMENT & AFFILIATIONS

- 2024 - Researcher, RAND
2024 Visiting scholar, Leverhulme Centre for the Future of Intelligence, Cambridge, Summer
2022-24 Executive Director in Global Technology, JPMorgan Chase & Co. Responsible AI lead for design and deployment of AI/ML. Tasks included: Using socio-technical methods for ethics risk mapping across AI development lifecycles; providing ethical advising on data and AI use cases; and establishing responsible AI best practices.
2022 - Associate fellow, Leverhulme Centre for the Future of Intelligence, University of Cambridge
2022 Affiliate, Machine Intelligence and Normative Theory Lab, Australian National University
2022 Visiting predoctoral researcher, Northeastern University, Summer
2022 Visiting predoctoral researcher, Leverhulme Centre for the Future of Intelligence, University of Cambridge, Spring

EDUCATION

- 2022 Georgetown University, Ph.D. in Philosophy
Dissertation: *Reframing Foreign Influence: War, Technology, and Words*
Committee: David Luban (chair), John Greco, Milton Regan, Shannon Vallor (external)
2018 Peking University, Master of Law at Yenching Academy
2016 Wellesley College, Bachelor of Arts with Honors in Philosophy and Political Science
Visiting student at the University of Oxford, 2014-15

PUBLICATIONS

BOOKS

-
- Mattingly, James & **Cibralic, Beba** (co first authors), *Machine Agency* (MIT Press, Feb 5, 2025)
<https://mitpress.mit.edu/9780262549981/philosophy-agency-and-ai/>

PEER-REVIEWED PUBLICATIONS

-
- Nyrup, Rune & **Cibralic, Beba** (co first authors) (2024). "Idealism, realism, pragmatism: three modes of theorising within secular AI ethics", in Barry Solemain & I. Glenn Cohen (eds.), *Research Handbook on Health, AI and the Law*. Edward Edgar Publishing. pp. 203-2018.
Cibralic, Beba (2024). "A Topography of Information-Based Foreign Influence", in Mitt Regan & Aurel Sari (eds.), *Hybrid Threats and Grey Zone Conflict: The Challenge to Liberal Democracies*. New York, NY: Oxford University Press. pp. 157-178.
Cibralic, Beba (2024). "Influence, War, and Ethics", *Journal of National Security Law and Policy* 14 (1):29-54.
Cibralic, Beba (2023). "Impermissible Targeting of Human Shields", *ARSP (Archiv für Rechts- und Sozialphilosophie)* 109 (2):171-194.
Cibralic, Beba & Mattingly, James (co first authors) (2022). "Machine agency and representation", *AI and Society* 39 (1):345-352.

BOOK REVIEWS

Cibralic, Beba, 2022. Review of Kate Darling's *New Breed. Essays in Philosophy*.

Cibralic, Beba, 2021. Review of German A. Duarte and Justin Michael Battin's *Reading Black Mirror*.
London Schools of Economics Review of Books blog.

POLICY REPORTS

Cibralic, Beba (first author) & Bogen, Miranda. "Building Blocks in Responsible AI." *Center for Democracy and Technology* (CDT). (Forthcoming, 2024).

Toner et al, 2024. "Through the Chat Window and Into the Real World: Preparing for AI Agents." Center for Security and Emerging Technology,
<https://cset.georgetown.edu/publication/through-the-chat-window-and-into-the-real-world-preparing-for-ai-agents/>

Cibralic, Beba, 2020. "Climate Colonialism Workshop Report." *Available upon request*.

Cibralic, Beba (first author) & Connelly, Aaron, 2018. "Russia's Disinformation Game in Southeast Asia." *Lowy Interpreter*,
<https://www.lowyinstitute.org/the-interpreter/russias-disinformation-game-southeast-asia>

Cibralic, Beba (first author) & Flitton, Daniel, 2018. "Trump-Putin: Beyond Election Meddling." *Lowy Interpreter*, <https://www.lowyinstitute.org/the-interpreter/trump-putin-beyond-election-meddling>

PUBLIC PHILOSOPHY

Leverhulme Centre (Cambridge) Group Submission to Future of Life Institute WorldBuilder Contest on AI Futures, 2022 (2nd place). <https://worldbuild.ai/core-central/>

Táíwò, Olúfẹ́mi & **Cibralic, Beba** (co first authors), 2022. "If the west can harbour Ukrainians, it can accept the many climate refugees to come." *Guardian*.
<https://www.theguardian.com/commentisfree/2022/apr/01/ukraine-war-west-immigration-climate-refugees>

Cibralic, Beba, 2021. "Astrology: Informative, Harmful, or Just Plain Fun?" *Aesthetics for Birds*,
<https://aestheticsforbirds.com/2021/08/19/astrology-informative-harmful-or-just-plain-fun/>

Cibralic, Beba. 2021. "Epistemic Norms and Failures of Reporting." *Social Epistemology Review and Reply Collective* 10 (5): 1-5. <https://wp.me/p1Bfg0-5NTi>.

Cibralic, Beba. 2020. "'Caliphate' and the Problem of Testimony." *Social Epistemology Review and Reply Collective* 9 (12): 33-36. <https://wp.me/p1Bfg0-5zi>.

Táíwò, Olúfẹ́mi & **Cibralic, Beba** (co first authors), 2020. "The Case for Climate Reparations." *Foreign Policy*,
<https://foreignpolicy.com/2020/10/10/case-for-climate-reparations-crisis-migration-refugees-in-equality/>

Cibralic, Beba & Lang, JJ, 2020. "Why do we get sad when robots die?" *The Outline*,
<https://theoutline.com/post/8514/why-am-i-sad-robot-died>

ACADEMIC ARTICLES IN PREPARATION

Cibralic, Beba. "The Nefarious Actor Problem: Disinformation, Political Speech, and Online Manipulation." (R&R from *Social Epistemology*).

Fleisher W, **Cibralic B**, Basl J, Ricks V, and Smith M. "Responsibility and Accountability in an Algorithmic Society." (Under review).

Cibralic, Beba. "Morality in War: Relevance of *Intentions in Jus Ad Bellum*". (Under review.)

Cibralic, Beba, Cooper, Ned, Jones, Cameron, Moore, Jared, Overmark, Rasmus (alph order). "Mind Games". (In preparation).

Cibralic, Beba. "AI and Epistemic Responsibility Gaps." (In preparation).

ACADEMIC PRESENTATIONS

PEER-REVIEWED PRESENTATIONS

- Jan 2025 “The ‘left-overs’: Bureaucracy and AI”, Society for Business Ethics, Eastern APA, NYC
- Aug 2024 “AI and Epistemic Responsibility Gaps”, Epistemic Accountability Workshop, Notre Dame University, Sydney
- July 2024 “Mind Games: Evaluating Theory of Mind and Persuasive Capacities in LLMs”, with Rasmus Overmark, Cameron Jones, Jared Moore, and Ned Cooper, Diverse Intelligence Summer Institute, St Andrews University
- Nov 2023 “Digital Ethics in Theory and Practice”, DevUP Conference, Texas
- Jan 2023 “Epistemic Responsibility and AI”, Philosophy, AI, and Society Workshop, Stanford
- Aug 2022 “Ethics of Machine Agents” with James Mattingly, Ethicomp, Turku University
- Feb 2022 “Impermissible Targeting of ‘Voluntary’ Human Shields”, American Philosophical Association (APA) Main Program, Central Division, Chicago, IL
- Feb 2022 “Rethinking Responsibility for Climate Refugees”, International Society for Environmental Ethics Session, APA Central Division, Chicago, IL
- Feb 2022 “The ‘How’ of Public Philosophy”, Public Philosophy Network, Central APA, Chicago, IL
- Feb 2022 “Human Shields in International Law”, Association for Practical and Professional Ethics International Conference, Cincinnati
- Nov 2021 “What is Owed: Special Duties and Climate Refugees”, International Studies Association, Northeast Division Meeting
- July 2021 “American Hegemony and Liberalism”, 38th International Social Philosophy Conference, North American Society for Social Philosophy, Neumann University
- July 2021 “Representation and Machine Agency” with James Mattingly, International Society for Ethics and Information Technology/International Association of Computing and Philosophy Joint Conference, Hamburg University
- Dec 2019 “The Ethics of Influence Operations”, Center for Ethics and the Rule of Law, University of Pennsylvania Law School
- Oct 2019 “Conceptual Boundaries of Fake News”, Bodaken Philosophy Symposium, Colorado State University
- Oct 2019 “The Epistemic Foundations of Misinformation”, International Studies Association/International Security Joint Annual Conference, University of Denver
- Sept 2018 “Challenges of Armed Conflict Classification”, 9th National Security Workshop, Judge Advocate General’s School, University of Virginia
- Feb 2018 “The Relevance of Intentions to Permissibility in *jus ad bellum*”, 2nd Australian Political Theory and Philosophy Conference, University of Sydney
- June 2017 “The Nature and Scope of LGBTQ+ counseling in China”, Peking University

PEER-REVIEWED TALKS NOT DELIVERED DUE TO COVID-19 PANDEMIC

- Nov 2021 “The Nefarious Actor Problems”, International Studies Association/International Security Joint Annual Conference, University of Bloomington
- Nov 2021 “Constructing Information Warfare”, International Studies Association/International Security Joint Annual Conference, University of Bloomington
- Nov 2021 “Language and Information Operations”, 71st Annual Northwest Philosophy Conference, Portland State University
- Sept 2021 “Rethinking Responsibility for Climate Refugees”, Western Canadian Philosophical Association, University of Victoria
- May 2020 “Mapping out Information Operations”, University of Virginia, Charlottesville
- May 2020 “Human Shielding”, Carnegie Council for Ethics in International Affairs

INVITED TALKS

- Feb 2025 “Legal Approaches to AI Agency”, Faculty of Law Seminar, Helsinki University
Spring ‘25 “Agency and the Problem of Biology”, Cognitive Science Symposium, Columbia University
May 2024 “On ‘AI Agents’: Seeing and Unseeing Agency in Machines”, RAND
May 2024 “Cross-cutting issues in Agentic AI”, Center for Security and Emerging Technology, Georgetown University
June 2023 “AI Ethics”, National Society for Black Engineers Juneteenth Celebration
June 2023 “AI, Ethics, and Finance”, ACI’s 7th Annual Summit for Women Leaders in Financial Services Law
Mar 2022 “Autonomous Weapons and the Problem of Trust”, Technology Ethics Center, University of Notre Dame
Sept 2021 “Drawing the line between Permissible and Impermissible Online Influence Efforts”, Information Security Research group, University College London
June 2021 “The Politics of Scale”, Philosophy and Climate Crisis Conference, Temple University
Feb 2021 “Refugees and Climate Justice” with Olúfẹ̀mi Táíwò, University of Pennsylvania
Feb 2020 “Attacking Words: Illocutions and Information Campaigns”, Center for Security and Emerging Technology, Georgetown University
May 2019 “Autonomous Weapons – Ethical Challenges and Opportunities”, The Interdisciplinary Computational Visionary Laboratory, Ben Gurion University
July 2018 “Russia’s new role in the Indo-Pacific”, European Studies Centre, Australian National University

INVITED COMMENTS

- Mar 2023 Comment on Tae Wan Kim, “Why AI chatbots should practice deceit less than humans”, Penn-Georgetown Digital Ethics Workshop
Jan 2022 Comment on A.G. Holdier, "On the Epistemic Threats of Counterfeit Memorials", Eastern American Philosophical Association

SELECTED FELLOWSHIPS, AWARDS, HONORS

- 2024 *100 Brilliant Women in AI Ethics* list, <https://womeninaethics.org/the-list/of-2024/>
2024 Diverse Intelligences Summer Institute (DISI) Fellowship, St. Andrews University (\$850 and accommodation)
2022-23 Charlotte W. Newcombe Fellowship, awarded for dissertation research (\$30,000) (declined funding; recognized as honorary fellow)
2022 Selected Participant and Scholarship Recipient for FAccT/ACM Doctoral Consortium, South Korea (\$2,000)
2022 Selected Participant and Scholarship Recipient for Summer Institute in Technology Ethics (SITE), Santa Clara College (\$2,750)
2022 2nd place in IVR Young Scholar Prize, awarded by the International Association for Philosophy of Law and Social Philosophy every other year to an outstanding article produced by a scholar under 35 years of age
2021-22 Fritz Family Predoctoral Fellowship in Tech & Society, Georgetown University (\$40,000 awarded to sponsoring center; fellow receives stipend)
2021-22 Horton-Hallowell Fellowship, merit award for dissertation research, Wellesley College Graduate Fellowships Committee (\$11,000)
2021-22 Hans J. Morgenthau Fellowship, awarded for yearlong program in political science, Notre Dame University (\$10,000)
2021-22 Language Scholarship for Russian, Georgetown University (\$13,000 for 2 semesters)

- 2021 Dissertation Research Travel Grant, Graduate School of Arts and Sciences, Georgetown University (\$5000)
- 2020-21 Gender+ Justice Fellowship, for research on climate migration, Georgetown University (\$1,500)
- 2020 *Young Woman to Watch in International Affairs* list, <https://www.youngausint.org.au/post>
- 2019 Graduate Student Teaching Assistant Award Nomination, Georgetown University
- 2019 Laura Fleming Grant, used for public philosophy project at D.C. Jail, Georgetown University (\$1,200)
- 2019 Selected Participant and Stipend Recipient, Certificate in the Engaged and Public Humanities, Georgetown University (\$1,000)
- 2019 Harvard Project in Asia and International Relations (HPAIR) Scholarship, to attend Kazakhstan program (\$1,200)
- 2019 Selected Participant and Scholarship Recipient, Yenching Symposium, China (\$1,500)
- 2018-19 Selected Participant and Scholarship Recipient for Stanford U.S.-Russian Forum
- 2018-19 Selected Participant, Young Professionals Program on foreign policy, Jamestown FoundationM
- 2018 Presidential Fellowship, awarded to highest-ranked doctoral admittees on Selection Committee's list, Brown University (\$35,750 p.a. plus tuition costs) (declined)
- 2017 Selected Participant and Scholarship Recipient, Future Leaders Program, China (\$1,000)
- 2017 Selected Participant and Scholarship Recipient, ToChina Summer Program, Italy (\$800 and accommodation)
- 2017 Selected Participant and Scholarship Recipient (\$1,000), Baku Global Forum, Azerbaijan
- 2016-18 Yenching Fellowship, two-year merit program at Peking University (\$20,000 p.a.)
- 2016 Phi Beta Kappa, Wellesley College
- 2016 Departmental Honors in Philosophy, Wellesley College
- 2016 Si Sigma Alpha Political Science Honors Society, Wellesley College
- 2012-16 Davis International Scholarship, one of two seniors to receive scholarship (\$20,000 p.a.)
- 2010-12 Australia United World Colleges Scholarship, one of two recipients (\$40,000 p.a.)

OTHER RESEARCH EMPLOYMENT

- 2024 Research contractor leading report on AI governance for Center for Democracy and Technology (think tank), D.C., Summer
- 2020 Research analyst working on AI and national security at the Center for Security and Emerging Technology (think tank), D.C., Summer
- 2020 Research assistant for Ben Buchanan on *The New Fire*, a book on AI and national security, Georgetown University, Spring
- 2020 Research assistant for Olúfẹ́mi Táíwò, Philosophy Dept., Georgetown University, Spring
- 2019-20 Co-Director of Research for Stanford U.S.-Russia Forum (research community), overseeing 10 multidisciplinary projects between Russian and American participants
- 2019 Ethics Lab graduate assistant focused on technology ethics, Kennedy Institute of Ethics, Georgetown University, Summer
- 2018 Research assistant on foreign policy at the Lowy Institute (think tank), Sydney, Summer
- 2016-17 Associate on international affairs at ThinkINChina (research community), Beijing
- 2016 Research assistant for director at Justitia (legal think tank), Copenhagen, Summer
- 2016 Research assistant for Paul MacDonald, Political Science Dept., Wellesley College, Spring
- 2015 Research assistant for Corinne Gartner, Philosophy Dept., Wellesley College. Summer

POLICY AND INDUSTRY WORKSHOPS

- June 2024 Invited participant at “Ethics of Advanced AI Assistants” Workshop, Google DeepMind and Stanford McCoy Family Center for Ethics in Society
- June 2024 Invited participant at “Documentation for AI Governance” Workshop, Center for Democracy and Technology (CDT)
- Oct 2023 Invited participant at “Operationalizing the Measure Function of NIST AI RMF” Workshop, Center for Advancing Safety of Machine Intelligence (CASMI), Northwestern University
- Oct 2022 Invited participant at “Responsible Innovation Practices Together” Workshop, Salesforce, SF

TEACHING EXPERIENCE

INSTRUCTOR

- 2025 “Philosophy of Computing”, Columbia University, Spring
- 2021 “Ethical Challenges of Artificial Intelligence”, Georgetown University, Fall
- 2021 “Ethics and the Environment”, Georgetown University, Summer
- 2021 “Global Justice”, Georgetown University, *Prisons and Justice Initiative*, Spring
- 2019 “Digital Ethics”, Georgetown University *Prisons and Justice Initiative*, Summer

ETHICS TEACHING IN INDUSTRY

- 2023 Developed inaugural firmwide beginner’s training in AI Ethics, JPMorgan Chase, Spring
- 2023 Designed Responsible AI practices workshops (400+ attendees), JPMorgan Chase, Spring
- 2022-23 Led AI Ethics for New Data Scientists global trainings, JPMorgan Chase

TEACHING ASSISTANT

- 2020-21 “Political and Social Thought”, Georgetown University
- 2019 “Introduction to Ethics”, Georgetown University, Fall
- 2019 “Philosophy of Friendship”, Georgetown *Prisons and Justice Initiative*, Spring
- 2019 “Journalism”, Georgetown *Prisons and Justice Initiative*, Substitute Instructor, Spring

GUEST LECTURES

- 2022 “The Responsibility Gap and Moral Agency”, *Minds and Machines*, Georgetown University, Spring
- 2022 “Speech and Technology”, *Critical Thinking, Knowledge, and Identity*, Georgia College and State University, Fall
- 2022 “Epistemic Security”, Grand Strategy, SAIS, Summer
- 2020 “Responsibility and Blame”, *Crime and Punishment*, Georgetown *Prisons and Justice Initiative*, Spring
- 2019 “Consent and Involuntary Human Shields”, *Morality and War*, University of Haifa, Summer

OTHER TEACHING ROLES

- 2022 - Volunteer philosophy tutor and discussant at Rikers Island Prison through Columbia
- 2021 Instructor on “International Humanitarian Law”, American Red Cross, Spring
- 2017 “Ethics” teacher for 5th grade, New South Wales Primary Ethics Program, Fall
- 2014-15 Writing instructor, Gatehouse drop-in center for the homeless and poorly housed, Oxford
- 2014 Program coordinator, ARAVA Valley of Peace, Israel/Palestine, Summer
- 2012-14 English tutor, St Stephens’ B-SAFE after-school program, Boston

MENTORSHIP

- 2021 Co-organizer of “Global Justice Dialogues” with Berkley Center for Religion, Peace and World Affairs, Georgetown University
- 2021 Founder and graduate director of “Symposium”, an undergraduate philosophy club and mentorship group, Georgetown University

TEACHING CERTIFICATES AND PEDAGOGICAL DEVELOPMENT

- 2022 Participated in teacher training through Morgenthau Fellowship, Notre Dame
- 2021 Participated in three-part “Doyle Conversations on Anti-racism in Higher Education”, Berkeley Center for Religion, Peace and World Affairs
- 2021 Participated in the “Anti-Racist Pedagogy in Philosophy” webinar, American Philosophical Association
- 2020 Completed “Apprenticeship in Teaching” Certificate, Georgetown Center for New Designs in Learning and Scholarship (CNDLS)
- 2020 Completed the Political and Social Philosophy Summer Teaching Institute, Georgetown University (\$5000)

SERVICE

ACADEMIC SERVICE

- 2023 - Co-founder and co-director, AI Ethics and Policy Discussion Group (industry-wide)
- 2024 Panel speaker on philosophy careers in industry, Georgetown University
- 2024 Referee for Conference on Association of Computing conference on Fairness, Accountability, and Transparency (ACM/FAccT)
- 2023 Referee for Conference on Association of Computing conference on Fairness, Accountability, and Transparency (ACM/FAccT)
- 2021 Panel speaker on academic success for philosophy graduate students, Georgetown University
- 2021-22 Philosophy Dept. Graduate Representative of Academic Integrity Standing Committee
- 2021 Providing curricular design support on teaching “Global Justice” to the School of Foreign Service, Georgetown University, Summer
- 2021 Organizer of “Constructing Information Warfare” workshop, MANCEPT Political Theory
- 2021 Panel speaker on applying to philosophy graduate school, Wellesley College
- 2021 Organizer of “Climate Crisis and Migration” workshop with Olúfẹ̀mi Táíwò, Georgetown University
- 2020 Panel speaker on studying at the Yenching Academy in China, Wellesley College
- 2019 Panel speaker on academic success for first-year graduate students, Georgetown University

CIVIC ENGAGEMENT

- 2022-24 LGBTQ+ Executive Council, JPMorgan Chase, New York
- 2022 Half-Marathon run in support of Cambridge Rape Crisis Centre, UK
- 2020 Volunteer, Democratic National Committee Voter Registration Hotline, remote
- 2017-18 Board of Directors, helping select scholarship recipients, United World Colleges, Australia
- 2015-16 Wellesley chapter developer, ALLIES (Civ-Mil education), Boston

LANGUAGES

Bosnian (fluent), Russian (high intermediate), Spanish (low intermediate)

REFERENCES

David Luban (dissertation committee chair), University Professor, Professor of Law and Philosophy, Georgetown University
luband@georgetown.edu

Kate Withy (teaching), Professor of Philosophy, Georgetown University
Kate.Withy@georgetown.edu

Milton Regan Jr., McDevitt Professor of Jurisprudence, Director of the Center on the Legal Profession, Co-Director on the Center on National Security and the Law, Georgetown University
regan@georgetown.edu

John Greco, Robert L. McDevitt, K.S.G., K.C.H.S. and Catherine H. McDevitt L.C.H.S Chair in Philosophy, Georgetown University
John.Greco@georgetown.edu

Shannon Vallor, Baillie Gifford Chair in the Ethics of Data and Artificial Intelligence and Professor of Philosophy, University of Edinburgh
svallor@exseed.ed.ac.uk

Deborah Johnson, Professor Emeritus of Applied Ethics, University of Virginia
dgj7p@virginia.edu

Seth Lazar, Professor of Philosophy, Australian National University and lead of Machine Intelligence and Moral Theory (MINT) Lab
seth.lazar@anu.edu.au

John Basl, Associate Director, AI and Data Ethics Initiatives at Northeastern Ethics Institute and Associate Professor of Philosophy, Northeastern University
j.basl@northeastern.edu

Rune Nyrup, Associate Professor, Center for Science Studies, Department of Mathematics, Aarhus University
rune.nyrup@css.au.dk

James Mattingly, Professor of Philosophy, Georgetown University
jmm67@georgetown.edu

Byce Huebner, Professor of Philosophy, Georgetown University
Byce.Huebner@georgetown.edu

Ben Buchanan, White House Special Advisor for AI
blb43@georgetown.edu