Linacre College
OX1 3JA Oxford, UK
March 13th 1993
+44 77 297 586 15
soren.mindermann@cs.ox.ac.uk

Sören Mindermann

Academic career summary

4th year PhD student in machine learning at the University of Oxford.

In the last 3 years, wrote 14 peer-reviewed publications of which 7 as (equally contributing) first author and 2 as senior author.

High-profile venues such as *Science*, *Nature Communications*, *NeurIPS* (3x), *ICML* (2x), *PNAS*.

National-scale policy impacts · Interviews on major TV and news outlets · Invited talks at ML conferences, industry, academia, and policy forums.

Topics: alignment in ML, active learning, data selection for deep learning, causal inference, Bayesian modeling, COVID-19.

Education

Oct. 2019 – present
Ph.D. Student in Machine Learning, *University of Oxford.*
Supervised by Yarin Gal (machine learning) and Allan Dafoe (Centre for the Governance of AI).

Sept. 2016 – Sept. 2017
MSc Computational Statistics and Machine Learning, *University College London.*
- Distinction.
- Thesis with Peter Dayan on hierarchical Bayesian reinforcement learning.

Sept. 2013 – Sept. 2016
BSc Mathematics, *University of Amsterdam, GPA: **8** (equivalent to **4.0**).*
- Co-authored an (unpublished) research review on efficient Monte Carlo methods in year 2.

Sept. 2012 – Sept. 2016
BSc Future Planet Studies, *University of Amsterdam, GPA: **7.6** (equivalent to **3.6**).*
- Natural and social sciences degree on solving the current and future challenges facing humanity.
- Completed two 3-year degrees simultaneously in 4 years.
- In preparation, self-taught Dutch language from no skills to fluent level (C1) in 7 weeks.
- Focused on economics and governance of resources, water, food and energy.

Work experience

July 2019 – October 2019
AI Governance Fellow, *University of Oxford*, Centre for the Governance of AI.
Chose and led a project on the economics of the AI industry structure. To be submitted.

July – December 2018
Research intern, *University of Toronto*, Vector Institute.
Machine learning for open source game theory under Prof. D. Duvenaud and R. Grosse.

November 2017 – May 2018
Research intern, *UC Berkeley*, CHAI group.
Authored 'Active Inverse Reward Design'.

October 2017 – November 2017
Research intern, *University of Oxford*, Future of Humanity Institute.
Equal 1st author of NeurIPS theory paper on inverse RL with Dr Stuart Armstrong.

June–August 2016
Fellow, *Pareto Fellowship*, Oakland, California.
Managed interaction with attending organizations to the EA Global conference at UC Berkeley. Led a team of 3-4. The fellowship provided funding and a curriculum. 0.3% acceptance rate.

March 2009
School intern, *University of Bremen, Technical Mathematics department*.
Programmed LEGO robots in C, analyzed sensor data in Matlab, presented to department staff.

March 2007
School intern, *Regiodata*, Bremen.
School holiday internship in computer hardware.

Publications as lead author

* = equal contribution

○ = ordered by coin flip

✉ = corresponding author

2021
Brauner JM*✉, **S Mindermann*✉**, Sharma M*✉, Johnston D, Salvatier J, Gavenciak T, Stephenson AB, Leech G, Altman G, Mikulik V, Norman AJ, Monrad JT, Besiroglu T, Ge H, Hartwick MA, Teh YW, Chindelevitch L, Gal Y, Kulveit J. *Inferring the effectiveness of government interventions against COVID-19.* In **Science**.

2022 **S Mindermann*⊠**, Muhammed Razzak*, Winnie Xu*, Andreas Kirsch, Mrinank Sharma, Aidan Gomez, Sebastian Farquhar, Jan Brauner, Yarin Gal. *Prioritized Training on Points that are Learnable, Worth Learning, and Not Yet Learnt.* **International Conference on Machine Learning.**

2021 S Mishra*⊠, **S Mindermann***, M Sharma*, C Whittaker*, T Mellan, T Wilton, D Klapsa, R Mate, M Fritzsche, M Zambon, J Ahuja, A Howes, X Miscouridou, G Nason, O Ratmann, G Leech, J Fabienne Sandkuhler, C Rogers-Smith, M Vollmer, H Unwin, Y Gal, M Chand, A Gandy, J Martin, E Volz, N Ferguson, S Bhatt, J Brauner, S Flaxman. *Changing composition of SARS-CoV-2 lineages and rise of Delta variant in England.* In **EClinicalMedicine (The Lancet)**.

2021 **S Mindermann*⊠○**,Mrinank Sharma*⊠○, Charlie Rogers-Smith, Gavin Leech, Benedict Snodin, Janvi Ahuja, Jonas B Sandbrink, Joshua Teperowski Monrad, George Altman, Gurpreet Dhaliwal, Lukas Finnveden, Alexander John Norman, Sebastian B Oehm, Julia Fabienne Sandkühler, Thomas Mellan, Jan Kulveit, Leonid Chindelevitch, Seth Flaxman, Yarin Gal, Swapnil Mishra, Jan Markus Brauner⊠, Samir Bhatt⊠. *Understanding the effectiveness of government interventions against the resurgence of COVID-19 in Europe.* In **Nature Communications**.

2020 Mrinank Sharma*, **S Mindermann***, Jan Brauner*, Gavin Leech, Anna Stephenson, Tomas Gavenciak, Jan Kulveit, Yee Whye Teh, Leonid Chindelevitch, Yarin Gal. *How Robust are the Estimated Effects of Nonpharmaceutical Interventions against COVID-19?* In **NeurIPS (Spotlight talk)**.

2020 A Jesson*, **S Mindermann***, U Shalit, Y Gal. *Identifying Causal-Effect Inference Failure with Uncertainty-Aware Models.* In **NeurIPS**.

2018 **Mindermann*, S.** & Armstrong*, S., (2018). *Occam's razor is insufficient to infer the preferences of irrational agents.* In **NeurIPS**.

2018 **Mindermann*, S.**, Shah*, R., Gleave, A., Hadfield-Menell, D. (2018). *Active Inverse Reward Design*, AAMAS/ICML workshop on goals in RL.

Publications as senior author

* = equal contribution to senior authorship

2022 G Leech⊠, C Rogers-Smith, J Sandbrink, B Snodin, R Zinkov, B Rader, J Brownstein, Y Gal, S Bhatt*, M Sharma*, **S Mindermann***, J Brauner*, L Aitchison*. *Mask wearing in community settings reduces SARS-CoV-2 transmission.* **Proceedings of the National Academy of Sciences (PNAS)**.

2022 G Altman⊠, J Ahuja⊠, JT Monrad, G Dhaliwal, C Rogers-Smith, G Leech, B Snodin, JB Sandbrink, L Finnveden, AJ Norman, SB Oehm, JF SandkÃŒhler, J Kulveit, S Flaxman, Y Gal, S Mishra, S Bhatt, M Sharma*, **S Mindermann***, J Brauner*. *A dataset of non-pharmaceutical interventions on SARS-CoV-2 in Europe.* **Nature Scientific Data**.

Publications as co-author

2023 A Lison, N Banholzer, M Sharma, **S Mindermann**, H Juliette T Unwin, S Mishra, T Stadler, S Bhatt⊠, N Ferguson, J Brauner, and W Vach. *Effectiveness assessment of non-pharmaceutical interventions: lessons learned from the COVID-19 pandemic.* In **The Lancet Public Health**.

2022 R Ngo⊠, L Chan⊠, **S Mindermann**⊠. *The Alignment Problem from a Deep Learning Perspective.* Arxiv.

2021 A Jesson⊠, **S Mindermann**, Y Gal, U Shalit. *Quantifying Ignorance in Individual-Level Causal-Effect Estimates under Hidden Confounding.* In **International Conference on Machine Learning**.

2021 Gideon Meyerowitz-Katz⊠, Samir Bhatt, Oliver Ratmann, Jan Markus Brauner, Seth Flaxman, Swapnil Mishra, Mrinank Sharma, **S Mindermann**, Valerie Bradley, Michaela Vollmer, Lea Merone, Gavin Yamey. *Is the cure really worse than the disease? The health impacts of lockdowns during COVID-19.* In **BMJ Global**.

2021 Tomas Gavenciak*, Joshua Teperowski Monrad*⊠, Gavin Leech, Mrinank Sharma, **S Mindermann**, Jan Markus Brauner, Samir Bhatt, Jan Kulveit*. *Seasonal variation in SARS-CoV-2 transmission in temperate climates.* In **PLOS Computational Biology**.

Awards

2022 **MPLS Impact Award.** Awarded for impact of research on “Understanding effectiveness of interventions against Covid-19 using Bayesian models”. 1st prize among Oxford researchers in any career stage across all MPLS departments (STEM and life sciences). The award is the basis for the REF studies and has been given only 28 times historically, including for the largest contribution to creating the *R* language, and to Nobel laureate Roger Penrose. (£1000)

2021 Edward Chapman Research Prize 2021. Award for the best first-author paper in the natural sciences at Magdalen College, Oxford. (£1000)

2019 DeepMind-Oxford Scholarship. Funding for a PhD.

2016 Pareto Fellowship.

2014 Heinrich Böll Foundation scholarship for academic excellence and engagement. (€36,000)

2011 German Mathematical Society Abitur Prize, 1st among 142 students.

Policy impact

2021 Preprint cited in the **German federal bill** that decided the national lockdown in force as of May 2021.

2021 I presented work on mask-wearing at the UK Cabinet Office to support the UK’s plan for fall 2021. (I co-supervised this paper.)

2021 Some COVID-19 papers on which I was (equally contributing) first author have been presented at the **WHO**, the modeling groups of the **Africa CDC** and the UK’s Scientific Advisory Group for Emergencies (**SAGE**), and the **House of Representatives** of the Netherlands.

2020 I presented our work on interventions against COVID-19 transmission to the modelling group of the Africa CDC.

Interviews given on TV and newspapers

2021 Monitor TV magazine on ARD (German equiv. of BBC, ca. 3m viewers per episode). Talked about preprint covering COVID’s 2nd wave.

2021 ITV Peston (the flagship political program of ITV). Talked about paper covering COVID’s 2nd wave.

2021 Suddeutsche Zeitung. Interview about government interventions in COVID’s second wave.

2021 NRC Handelsblad, 2021. Talked about paper on government interventions in COVID’s first wave.

2021 Alan Turing Institute Podcast, 2021. Talked about government interventions in COVID’s first wave.

Invited talks

2023 **Philip Torr Vision Group**, *The Alignment Problem from a Deep Learning Perspective.*

2023 **Future of Life Institute**, *The Alignment Problem from a Deep Learning Perspective.*

2023 **University of Amsterdam**, *The Alignment Problem from a Deep Learning Perspective.*

2023 **University of Edinburgh**, *The Alignment Problem from a Deep Learning Perspective.*

2023 **UC Berkeley (CHAI)**, *The Alignment Problem from a Deep Learning Perspective.*

2022 **Meta AI**, *Prioritized training on points that are learnable, worth learning, and not yet learned.*

2022 **USC (Cutelab)**, *Prioritized training on points that are learnable, worth learning, and not yet learned.*

2022 **University of Oxford—AI for Agent-Based Modeling seminar**, *Inferring the Effectiveness of government interventions against COVID-19.*

2021 **Cohere.ai**, *Prioritized training on points that are learnable, worth learning, and not yet learned.*

2021 **ETH Zürich**, *Government interventions in the second wave.*

2021 **MRC Centre for Global Infectious Disease Analysis**, *Government interventions in the second wave*, Imperial College.

2020 **Africa CDC**, *Inferring the effectiveness of government interventions against COVID-19.*

2020 **German Centre for Infection Research**, *Inferring the effectiveness of government interventions against COVID-19.*

2020 **NeurIPS Spotlight**, *How Robust are the Estimated Effects of Nonpharmaceutical Interventions against COVID-19?.*

2020 **NeurIPS COVID Symposium Spotlight**, *How Robust are the Estimated Effects of Nonpharmaceutical Interventions against COVID-19?.*

Professional service

2023	Reviewer, ICML
2022	Reviewer, ICLR
2022	Reviewer, NeurIPS
2022	Reviewer, ICML
2021	Reviewer, ICLR
2021	Reviewer, NeurIPS
2021	Reviewer, ICML
2020	Reviewer, Nature Machine Intelligence
2020	Reviewer, NeurIPS
2020	Reviewer, ICML
2018	Reviewer, NeurIPS Smooth Games Optimization and Machine Learning Workshop.

Volunteer service

2014—present **Wikipedia contributor**.
Authoring articles such as *AI Alignment* to improve broad education.

April 2013– April 2014 **Volunteer**, *New Harvest.*
Built a database and map of scientists and identified grant-makers for alternative protein research.

2012–2015 **Board member / later treasurer**, *Interdisciplinary student association (Spectrum).*
Organized trips.