Dmitry Kobak

CV

April 2026

Ghent University
Technologiepark-Zwijnaarde 75, 9052 Gent, Belgium
dmitry.kobak@ugent.be

https://dkobak.github.io
https://scholar.google.com/citations?user=BUQbD5kAAAAJ

Married, with three kids (2011, 2013, 2016).

Academic positions

from 2026 Full professor at Ghent University, Dept. of Mathematics, Computer Science, and Statistics

from 2026 Principal investigator at VIB Center for AI and Computational Biology (VIB.AI)

2021–2026 Group leader at the Hertie Institute for AI in Brain Health, University of Tübingen

2023–2024 Temporary professor (Vertretungsprofessor) at Heidelberg University

Academic affiliations and memberships

from 2022 Member of ELLIS (European Laboratory for Learning and Intelligent Systems)

2024–2026 Privatdozent at the Department of Computer Science, University of Tübingen

2024–2026 Member of the Cluster of Excellence "Machine Learning in Science"

2024–2026 Member of the education board for Computational Neuroscience MSc program

2023–2026 Associated scientist at the International Max Planck Research School for Intelligent Systems (IMPRS-IS)

Post-doctoral experience

2017–2021 University of Tübingen, lab of Philipp Berens

2013–2016 Champalimaud Centre for the Unknown (Lisbon), lab of Christian Machens

Education

2024 University of Tübingen, Dept. of Computer Science, Habilitation

2012 Imperial College London, Dept. of Bioengineering, PhD (Supervisor: Carsten Mehring)
2007–2010: at University of Freiburg, Bernstein Center for Computational Neuroscience

2007 St. Petersburg State University
Physics Department, High Energy Physics group, MSc (summa cum laude)

2004 St. Petersburg University of Information Technologies, Mechanics, and Optics
Department of Information Technologies and Computer Science, BSc (summa cum laude)

2000 St. Petersburg Classical Gymnasium #610

Third-party funding

2022–2025 €162 610. Deutsche Forschungsgemeinschaft (DFG), KO6282/2-1. "Trustworthy multi-scale manifold learning for genomic and transcriptomic data". Principal investigator.

2021–2025 €260 000. CyberValley Research Fund, CyVy-RF-2020-14. "Contrastive learning for dimensionality reduction and visualization of transcriptomic data". Principal investigator.

2015 €40 500. Bial Foundation Grant, 389/14. "Demixing and visualizing neural population activity in higher cortical areas". Principal investigator.

Organising scientific meetings

2026 VIB conference "AI and Computational Biology", Leuven, Belgium

2024 The main organiser of the Dagstuhl seminar "Low-Dimensional Embeddings of High-Dimensional Data: Algorithms and Applications", Germany. Co-organizers: Fred Hamprecht, Smita Krishnaswamy, Gal Mishne.

Reviewing and Chairing

- Area chair at ICLR (from 2026), ICML (from 2026), and NeurIPS (from 2026). Action editor at TMLR (from 2025).
- Reviewer at NeurIPS (2020–2025), ICLR (2021–2025), ICML (2022–2025), TMLR (2023–2025), ECML (2020–2023), AISTATS (2022–2023), JMLR (from 2020). *Highlighted Reviewer* at ICLR 2022 and *Expert Reviewer* at TMLR.
- Reviewer for *Bioinformatics*, *Genome Biology*, *Nature Communications*, *Nature Methods*, *Nature Biotechnology*, and *PLoS Computational Biology*.
- Reviewer of grant applications for ERC (Europe), NSERC (Canada), and NWO (The Netherlands).

Teaching

2024/25 Introduction to Machine Learning, Tübingen University. MSc course, compulsory for MSc programs *Computational Neuroscience* and *Quantitative Data Science*. 4 SWS.

2020/21 lectures were offered online due to the Covid pandemic and are available at https://www.youtube.com/playlist?list=PL05umP7R6ij35ShKLDqccJSDntugY4FQT.

2023/24 Transformers, large language models, and applications in physics. Heidelberg University, MSc seminar. 2 SWS.

2023/24 Einführung ins Machinelle Lernen, Heidelberg University (in German). BSc course in the Physics department. 4 SWS.

2022/23 Introduction to Machine Learning, Tübingen University. 4 SWS.

2021/22 Introduction to Machine Learning, Tübingen University. 4 SWS.

2020/21 Introduction to Machine Learning, Tübingen University. 4 SWS.

Publications

1. Ahlmann-Eltze C, Barkamm F, Lause J, Boeva V, Kobak D (2026) Representation learning of single-cell RNA-seq data. *RNA*, 080889.125.

2. González-Márquez R, Berens P, Kobak D (2026) Cropping outperforms dropout as an augmentation strategy for training self-supervised text embeddings. *Transactions of Machine Learning Research*.

3. Kury N, Kobak D, Damrich S (2026) DREAMS: preserving both local and global structure in dimensionality reduction. *Transactions of Machine Learning Research*.

4. Böhm JN, Keute M, Guzmán A, Damrich S, Draganov D, Kobak D (2025) Node embeddings via neighbor embeddings. *Transactions of Machine Learning Research*.

5. Schmors L, Gonschorek D, Böhm JN, Qiu Y, Zhou N, Kobak D, Tolias A, Sinz F, Reimer J, Franke K, Damrich S, Berens P (2025) TRACE: Contrastive learning for multi-trial time-series data in neuroscience. *Conference on Neural Information Processing Systems*.

6. Draganov D, Vadgama S, Damrich S, Böhm JN, Maes L, Kobak D*, Bekkers E* (2025) On the importance of embedding norms in self-supervised learning. *International Conference for Machine Learning*.

7. Kobak D, González-Márquez R, Horvát EA, Lause J (2024) Delving into LLM-assisted writing in biomedical publications through excess vocabulary. *Science Advances*, 11 (27).

8. Bernaerts Y, Deistler M, Goncalves PJ, Beck J, Stimberg M, Scala F, Tolias AS, Macke JH, Kobak D, Berens P (2025) Combined statistical-mechanistic modeling links ion channel genes to physiology of cortical neuron types. *Patterns*, 6 (10), 101323.

9. Kobak D, Bessudnov A, Ershov A, Mikhailova T, Raksha A (2025) War fatalities in Russia in 2022–2023 estimated via excess male mortality: a research note. *Demography*, 62(2): 335–347.

10. Damrich S, Berens P, Kobak D (2024) Persistent homology for high-dimensional data based on spectral methods. *Conference on Neural Information Processing Systems*.

11. Lause J, Berens P, Kobak D (2024) The art of seeing the elephant in the room: 2D embeddings of single-cell data do make sense. *PLoS Computational Biology* 20(10): e1012403.

12. Greydanus S, Kobak D (2024) Scaling down deep learning with MNIST-1D. *International Conference for Machine Learning*.

13. González-Márquez R, Schmidt L, Schmidt BM, Berens P, Kobak D (2024) The landscape of biomedical research. *Patterns*, 5 (6), 100968.

14. González-Márquez R, Kobak D (2024) Learning representations of learning representations. *Data-centric Machine Learning Research (DMLR) workshop at ICLR 2024*.

15. Damrich S, Böhm JN, Hamprecht FA, Kobak D (2023) From t-SNE to UMAP with contrastive learning. *International Conference for Learning Representations*.

16. Böhm JN, Berens P, Kobak D (2023) Unsupervised visualization of image datasets using contrastive learning. *International Conference for Learning Representations*.

17. Tamazian G, Komissarov AB, Kobak D, Polyakov D, Andronov E, Nechaev S, Kryzhevich S, Porozov Y, Stepanov E (2023) t-SNE Highlights Phylogenetic and Temporal Patterns of SARS-CoV-2 Spike and Nucleocapsid Protein Evolution. *International Symposium on Bioinformatics Research and Applications*, 255–262.

18. Bachmann F, Hennig P, Kobak D (2023) Wasserstein t-SNE. *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*.

19. Böhm JN, Berens P, Kobak D (2022) Attraction-repulsion spectrum in neighbor embeddings. *Journal of Machine Learning Research*, 23(95), 1–32.

20. Shen S, Jiang X, Scala F, Fu J, Fahey P, Kobak D, Tan Z, Zhou N, Reimer J, Sinz F, Tolias AS (2022) Distinct organization of two cortico-cortical feedback pathways. *Nature Communications* 13, 6389.

21. González-Márquez R, Berens P, Kobak D (2022) Two-dimensional visualization of large document libraries using t-SNE. *Geometrical and Topological Representation Learning (GTRL) workshop at ICLR 2022*.

22. Bashford L*, Kobak D*, Diedrichsen J, Mehring C (2022) Motor skill learning decreases movement variability and increases planning horizon. *Journal of Neurophysiology* 127(4), 995–1006.

23. Kobak D (2022) Underdispersion: A statistical anomaly in reported Covid data. *Significance* 19(2), 10–13.

24. Scala F*, Kobak D*, Bernabucci M, Bernaerts Y, Cadwell CR, Castro JR, Hartmanis L, Jiang X, Laturnus S, Miranda E, Mulherkar S, Tan ZH, Yao Z, Zeng H, Sandberg R, Berens P, Tolias AS (2021) Phenotypic variation of transcriptomic cell types in mouse motor cortex. *Nature* 598, 144–150.

25. BRAIN Initiative Cell Census Network (BICCN) (2021) A multimodal cell census and atlas of the mammalian primary motor cortex. *Nature* 598, 86–102.

26. Ilanchezian I, Kobak D, Faber H, Ziemssen F, Berens P, Ayhan MS (2021) Interpretable gender classification from retinal fundus images using BagNets. *International Conference on Medical Image Computing and Computer-Assisted Intervention*.

27. Kobak D, Linderman GC (2021) Initialization is critical for preserving global data structure in both t-SNE and UMAP. *Nature Biotechnology* 39, 156–157.

28. Kobak D, Bernaerts Y, Weis MA, Scala F, Tolias A, Berens P (2021) Sparse reduced-rank regression for exploratory visualization of multimodal data sets. *Journal of the Royal Statistical Society, Series C (Applied Statistics)* 70 (4), 980–1000.

29. Karlinsky A, Kobak D (2021) Tracking excess mortality across countries during the COVID-19 pandemic with the World Mortality Dataset. *eLife* 10:e69336.

30. Lause J, Berens P, Kobak D (2021) Analytic Pearson residuals for normalization of single-cell RNA-seq UMI data. *Genome Biology* 22: 258.

31. Kobak D (2021) Excess mortality reveals Covid's true toll in Russia. *Significance* 18 (1), 16–19.

32. Lipovsek M, Bardy C, Cadwell CR, Hadley K, Kobak D, Tripathy SJ (2021) Patch-seq: past, present, and future. *Journal of Neuroscience* 41 (5), 937–946.

33. Kobak D, Lomond J, Sanchez B (2020) The optimal ridge penalty for real-world high-dimensional data can be zero or negative due to the implicit ridge regularization. *Journal of Machine Learning Research* 21(169): 1–16.

34. Kobak D, Shpilkin S, Pshenichnikov MS (2020) Suspect peaks in Russia's "referendum" results. *Significance* 17 (5), 8–9.

35. Laturnus S, Kobak D, Berens P (2020) A systematic evaluation of interneuron morphology representations for cell type discrimination. *Neuroinformatics* 18: 591–609.

36. Cadwell CR, Scala F, Fahey PG, Kobak D, Mulherkar S, Sinz FH, Papadopoulos S, Tan ZH, Johnsson P, Hartmanis L, Li S, Cotton RJ, Tolias KF, Sandberg R, Berens P, Jiang X, Tolias A (2020) Cell type composition and circuit organization of clonally related excitatory neurons in the juvenile mouse neocortex. *eLife* 9, e52951.

37. Kobak D, Linderman G, Steinerberger S, Kluger Y, Berens P (2019) Heavy-tailed kernels reveal a finer cluster structure in t-SNE visualisations. *Joint European Conference on Machine Learning and Knowledge Discovery in Databases* 124–139.

38. Kobak D, Berens P (2019) The art of using t-SNE for single-cell transcriptomics. *Nature Communications* 10(1): 5416.

39. Scala F*, Kobak D*, Shan S, Bernaerts Y, Laturnus S, Cadwell CR, Hartmanis L, Froudarakis E, Castro JR, Tan ZH, Papadopoulos S, Patel SS, Sandberg R, Berens P, Jiang X, Tolias AS (2019) Layer 4 of mouse neocortex differs in cell types and circuit organization between sensory areas. *Nature Communications* 10(1), 4174.

40. Kobak D*, Pardo-Vazquez JL*, Valente M, Machens CK, Renart A (2019) State-dependent geometry of population activity in rat auditory cortex. *eLife* 8, e44526.

41. Kobak D, Shpilkin S, Pshenichnikov MS (2018) Putin's peaks: Russian election data revisited. *Significance* 15 (3), 8–9.

42. Kobak D*, Brendel W*, Constantinidis C, Feierstein CE, Kepecs A, Mainen ZF, Qi X-L, Romo R, Uchida N, Machens CK (2016) Demixed principal component analysis of neural population data. *eLife* 5, e10989.

43. Tian J, Huang R, Cohen JY, Osakada F, Kobak D, Machens CK, Callaway EM, Uchida N, Watabe-Uchida M (2016) Distributed and mixed information in monosynaptic inputs to dopamine neurons. *Neuron* 91 (6), 1374–1389.

44. Kobak D, Shpilkin S, Pshenichnikov MS (2016) Integer percentages as electoral falsification fingerprints. *Annals of Applied Statistics* 10 (1), 54–73.

45. Kobak D, Shpilkin S, Pshenichnikov MS (2016) Statistical fingerprints of electoral fraud? *Significance* 13 (4), 20–23.

46. Kobak D and Mehring C (2012) Adaptation paths to novel motor tasks are shaped by prior structure learning. *Journal of Neuroscience* 32 (29), 9898–9908.

Preprints

47. de Bodt C*, Diaz-Papkovich A*, Bleher M, Bunte K, Coupette C, Damrich S, Sanmartin EF, Hamprecht FA, Horvát EA, Kohli D, Krishnaswamy S, Lee JA, Lelieveldt BPF, McInnes L, Nabney IT, Noichl M, Poličar PG, Rieck B, Wolf G, Mishne G^{+}, Kobak D^{+} (2025) Low-dimensional embeddings of high-dimensional data. *arXiv*, 2508.15929.

48. Stanley JS, Yang J, Li R, Lindenbaum O, Kobak D, Landa B, Kluger Y (2025) Principled PCA separates signal from noise in omics count data. *bioRxiv*, 2025.02.03.636129v1.

49. Damrich S, Klockow MV, Berens P, Hamprecht FA, Kobak D (2024) Visualizing single-cell data with the neighbor embedding spectrum. *bioRxiv*, 2024.04.26.590867.

50. Nwabufo IV, Böhm JN, Berens P, Kobak D (2024) Self-supervised visualisation of medical image datasets. *arXiv*, 2402.14566.

51. Lause J, Ziegenhain C, Hartmanis L, Berens P, Kobak D (2023) Compound models and Pearson residuals for normalization of single-cell RNA-seq data without UMIs. *bioRxiv*, 2023.08.02.551637.

52. Bernaerts Y, Berens P, Kobak D (2020) Sparse bottleneck networks for exploratory analysis and visualization of neural Patch-seq data. *arXiv* 2006.10411.

Invited talks at scientific meetings

2025 Keystone Symposium "AI in Molecular Biology", Santa Fe, USA. All embeddings are wrong but some are useful.

2025 SIAM Annual Meeting workshop "Recent Theoretical and Computational Advances in Geometry-Driven Analysis of Complex Data", Montreal, Canada. Neighbor embeddings & persistent homology for high-dimensional data and abstract graphs.

2024 Ascona Workshop "Spatial and temporal statistical modeling in molecular biology", Ascona, Switzerland. All embeddings are wrong but some are useful

2024 Dagstuhl Seminar "Low-Dimensional Embeddings of High-Dimensional Data: Algorithms and Applications", Dagstuhl, Germany. Neighbour embeddings meet contrastive learning.

2023 Network Seminar (online). Contrastive and neighbour embedding methods for data visualisation. https://www.youtube.com/watch?v=A2HmdO8cApw

2022 International Conference on Learning Representations (ICLR), Geometrical and Topological Representation Learning workshop (online). What are 2D neighbour embeddings of scRNA-seq data actually useful for? https://www.youtube.com/watch?v=qt8z-MzIwpg

2021 Society for Neuroscience Annual Conference (SfN), Minisymposium "Patch-seq — Everything You Always Wanted to Know, But Were Afraid to Ask" (online). Mapping Patch-seq cells to a reference atlas.

2021 ELLIS Life Data Science Seminar (online). Neighbour embeddings for scientific visualization. https://www.youtube.com/watch?v=CsUqmug7ZMc

2020 Meeting "Emerging Perspectives in Clinical Brain Research" (online), Statistical analysis of transcriptomic and multi-omic neural data.

Invited talks at departmental seminars

2025 Institute of Stochastics, Ulm University

2024 Institute for Machine Learning, ETH Zurich

Supervision

Postdocs

2023–2025 Sebastian Damrich

PhD students

from 2022 Rita González Márquez

2025 Jan Niklas Böhm, "Visualization beyond Riemannian manifolds"

MSc students

from 2025 Lena Holzwarth

2025 Moritz Christ: "kNN-graph coarsening from the perspective of dimensionality reduction"

2023 Marius Keute: "Graph embeddings with t-SNE and contrastive learning"

2021 Rita González Márquez: "Two-dimensional visualizations of document libraries"

2021 Fynn Bachmann: "Wasserstein t-SNE and applications to social science"

2020 Jan Niklas Böhm: "Dimensionality reduction with neighborhood embeddings"

Lab rotation students

2026 Ananya Albrecht-Buehler

2025 Artemii Schlychkov: "t-SNE visualisations with adaptive degrees of freedom parameter"

2025 Lena Holzwarth: "MTEB for TF-IDF: Using a modern benchmark to evaluate old-school embeddings"

2023 Alica Guzmán: "Graph layouts and contrastive learning using neighbour embedding algorithms"

2023 Tian Jin & Natsumi Omura: "Modeling the relationship between extreme weather and excess deaths based on local European data"

2022 Luca Schmidt: "Visualizing biomedical text corpora in 2D using pre-trained language models"

2021 Rita González Márquez: "Two-dimensional embedding of the entire biomedical literature"

2021 Weiyi Xiao: "t-SNE visualization of large genomic cohorts"