Kempner INSTITUTE

For the Study of Natural & Artificial Intelligence at Harvard University

Research Roundup

October 2026

Discover our latest insights in natural and artificial intelligence research

Highlights

- Research Fellow David Clark and Associate Faculty Haim Sompolinsky **show** that disordered neural circuits can still support continuous attractors, with hidden symmetry in their connectivity helping explain how the brain may represent variables such as head direction.
- Associate Faculty Venkatesh Murthy and Farhad Pashakhanloo use LLM-derived descriptors to **study** the perceptual geometry of natural odor mixtures.
- A team including Associate Faculty Marinka Zitnik **introduces** CONCERT, a generative model for predicting how biological perturbations affect cells across spatial contexts.

Featured Formula

$$\frac{du_i}{d\tau} = -\kappa u_i + \frac{v_i}{c_* + \gamma}, \qquad \frac{dv_i}{d\tau} = \frac{u_i}{c_* + \gamma} - \left(\kappa + \frac{s_i}{2\gamma}\right) v_i$$

Formula from the preprint **"A theory of noisy linear recurrent neural network learning dynamics,"** by John J. Vastola and Kanaka Rajan that shows how noise and regularization shape the evolution of the network's read-in and readout weights.

Featured Figure

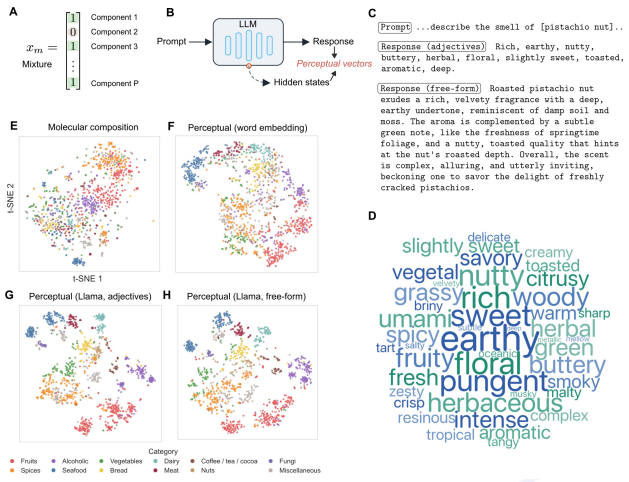


Figure from the preprint "**Learning Perceptual Geometry of Natural Odor Mixtures from LLM-Derived Descriptors,**" by Farhad Pashakhanloo and Venkatesh N. Murthy illustrating how molecular mixtures are mapped into LLM-based representation spaces.

Featured Artifacts

Latent Decision Flow Rewards for Clinical Guideline Reasoning
Bhatt, Gaurav, and Marinka Zitnik (2026).

Get the Code

Generative modeling of spatially resolved perturbation responses with CONCERT
Lin, Xiang, Zhenglun Kong, Boyang Fu, Soumya Ghosh, Manolis Kellis, and Marinka Zitnik (2026).

Get the Code

AI Innovation

Latent Decision Flow Rewards for Clinical Guideline Reasoning.
Bhatt, Gaurav, and Marinka Zitnik. arXiv:2610.38709 (2026).

Scaling Transformer Energy-based Models For High Quality Image Synthesis.
Lee, Jungsoo, Alexi Gladstone, and Yilun Du. arXiv:2610.31889 (2026).

Action Learning for Scientific Simulations.
Liaw, Sarah, Yuanqi Du, and Yilun Du. arXiv:2610.31585 (2026).

Motion Equivariant Convolutional LSTM: When Motion Becomes a Gauge Symmetry.
Hosseinzadeh, Vahid, T. Anderson Keller, and Michael Puthawala. arXiv:2610.28771 (2026).

Learning to Be Different: Component Specialization for Compositional Diffusion Policies.
Laprevotte, Alec, Haonan Chen, Xiaoshen Han, and Yilun Du. arXiv:2610.27194 (2026).

Weighted State Souping for Efficient Long-Context Retrieval.
Tong, William L., Aryo Lotfi, Emmanuel Abbe, Josh Susskind, Cengiz Pehlevan, and Eran Malach. arXiv:2610.25621 (2026).

Multimodal Filesystem Memory For Long-Horizon Embodied Agents.
Su, Chloe, Jiahan Li, Zhenting Qi, Chaoran Cheng, Haoyu Chen, Sejoon Jun, Weirui Ye, and Yilun Du. arXiv:2610.22686 (2026).

Are We Overthinking One-Step Generation?
Park, Woobin, Jaeyeon Kim, Yilun Du, Sitan Chen, and Bohyung Han. arXiv:2610.20331 (2026).

Model-Based Online Decision Making via Generative Trajectory Planning.
Balim, Haldun, Yilun Du, and Na Li. arXiv:2610.18677 (2026).

Releasing the Inner Jev: A Simple Recipe for Open Decision Models.
Geuter, Jonathan, Andrew Cheng, and David Alvarez-Melis. arXiv:2610.17205 (2026).

Scaling Multi-Agent Intelligence with Substrate-Centric Coordination.
Qi, Zhenting, Zhuokai Zhao, Rulin Shao, Kerem Oktar, Zhiyuan Yao, Himabindu Lakkaraju, Wen-tau Yih, Ansong Ni, Bhargavi Paranjape, and Yilun Du. arXiv:2610.16120 (2026).

Unsupervised Multimodal Reward Modeling.
Li, Yueying, Jingxuan Fan, Kianté Brantley, Sham M. Kakade, and Hanlin Zhang. arXiv:2610.12707 (2026).

Policy as Data: Replay-Based Policy Dual Averaging via Advantage Regression.
Peng, Nianli, Geoffrey J. Gordon, and Kianté Brantley. arXiv:2610.04638 (2026).

MoSE3: Learning World-Space SE (3) at Every Pixel.
Cheng, Jiahuan, Zhiyi Li, Tian Xia, Ruojin Cai, Yilun Du, and Qianqian Wang. arXiv:2610.03716 (2026).

Finetuning with Sampling: SFT Learns Better Than You Think.
Karan, Aayush, Sitan Chen, and Yilun Du. arXiv:2610.02140 (2026).

Interpretability and AI Theory

On the stability of learning.
Vastola, John J. arXiv:2610.35856 (2026).

Quantifying the failure of symmetry learning in linear diffusion models due to initialization and finite-sample effects.
Vastola, John J., and Kanaka Rajan. arXiv:2610.35720 (2026).

Effectiveness of Curriculum Learning Depends on Reward Sparsity and Competing Optima.
Vastola, John J., Ann Huang, Satpreet H. Singh, Samuel J. Gershman, and Kanaka Rajan. arXiv:2610.31884 (2026).

Circuits of LLM Classifiers.
Khalaf, Hadi, Sajani Vithana, Kunal Talreja, Haewon Jeong, Viveck R. Cadambe, and Flavio Calmon. arXiv:2610.30076 (2026).

Is Sycophancy a Form of In-Context Updating?
Shapira, Itai, Myra Cheng, and Hadas Orgad. arXiv:2610.28325 (2026).

Transformers Learn Two Competing Sorting Algorithms, and Only One Length-Generalizes.
Gatmiry, Khashayar, Nathan W. Henry, Chloe H. Su, Emmanuel Abbe, Samy Bengio, Melanie Weber, Jonathan A. Kelner, and Peter L. Bartlett. arXiv:2610.26974 (2026).

In-context classification: Calibration and noise in linear attention with softmax readout.
Rysmakhanov, Arman, Mary Letey, Yue M. Lu, Jacob Zavatone-Veth, and Cengiz Pehlevan. arXiv:2610.17061 (2026).

What can linear attention learn from nonlinear teachers in-context?
Letey, Mary, Arman Rysmakhanov, Yue M. Lu, Cengiz Pehlevan, and Jacob Zavatone-Veth. arXiv:2610.10761 (2026).

A Solvable Model of Adaptive Learning Rate Rescaling: Acceleration, Stability & Scaling.
Lavie, Itay, Clarissa Lauditi, and Cengiz Pehlevan. arXiv:2610.06701 (2026).

When Do Biological Reasoning Models Use Their Biological Inputs?
Fang, Ada, Nikitha Thoduguli, Lukas Fesser, Hanlin Zhang, Sham M. Kakade, and Marinka Zitnik. arXiv:2610.00898 (2026).

Decoding Alignment without Encoding Alignment: A critique of similarity analysis in neuroscience.
Bertram, Johannes, Luciano Dyballa, T. Anderson Keller, Savik Kinger, and Steven W. Zucker. arXiv:2605.05907 (2026).

Applications of AI

Safe Multi-Robot Collaboration With Uniform Exclusion Using Trust.
Kaminsky, Thomas, Áron Vékássy, Hammad Izhar, Angelia Nedić, and Stephanie Gil. arXiv:2610.28251 (2026).

AfriGuard: Safety and Robust Refusal For African Languages.
Azime, Israel Abebe, Jesujoba O. Alabi, Iffat Maab, Alexander D. Cai, Tadesse Destaw Belay, Atnafu Lambebo Tonja, Jonas Kgomo et al. arXiv:2610.26474 (2026).

Dual-graph deconvolution of enzyme fitness: an interpretable framework to identify and explain catalytic modulators.
Garcia Busto, Sergio, Ethan Eschbach, Thomas A. Hopf, Chris Sander, Sergio Bacallado, and Debora S. Marks. bioRxiv (2026).

Generative modeling of spatially resolved perturbation responses with CONCERT.
Lin, Xiang, Zhenglun Kong, Boyang Fu, Soumya Ghosh, Manolis Kellis, and Marinka Zitnik. bioRxiv (2026).

Neuroscience & NeuroAI

Temporal smoothness constraints on efficient neurobiological codes imply temporal specialization.
Vastola, John J., and Samuel J. Gershman. arXiv:2610.36272 (2026).

A theory of noisy linear recurrent neural network learning dynamics.
Vastola, John J., and Kanaka Rajan. arXiv:2610.35785 (2026).

Solvable models of learning to pursue a moving target.
Vastola, John J., and Kanaka Rajan. arXiv:2610.30721 (2026).

Learning Perceptual Geometry of Natural Odor Mixtures from LLM-Derived Descriptors.
Pashakhanloo, Farhad, and Venkatesh N. Murthy. arXiv:2610.17164 (2026).

Efficiency and robustness partition the solution space for memory in recurrent networks.
Qian, William, and Cengiz Pehlevan. arXiv:2610.04697 (2026).

A network model of path control.
Osman, Mohammed Abdal Monium, Wenyi Zhang, Cengiz Pehlevan, and Rachel I. Wilson. bioRxiv (2026).

Spontaneous behavior predicts colony membership, individual identity, and rank in naked mole-rat societies.
Haba, Yuki, Ryan Schwark, Caleb Weinreb, Simon Ogundare, William Foster, Yu-Young Wesley Tsai, Jerry Lyu et al. bioRxiv (2026).

Cooperation recruits corticolimbic projections and induces synaptic change in rats.
Kietzman, Henry W., R. Allie Cauchon, Amelia Johnson, David R. Backer Peral, Robin Bonomi, Yiyun Huang, Hayde Sanchez, Alya Bagdas, Andy Chen, Shreya Saxena, S. William Li, and Jane R. Taylor. bioRxiv (2026).

Symmetries and continuous attractors in disordered neural circuits.
Clark, David G., L. F. Abbott, and Haim Sompolinsky. Nature Neuroscience (2026).

Associative synaptic plasticity creates dynamic persistent activity.
Wakhloo, Albert J., David G. Clark, and L. F. Abbott. Proceedings of the National Academy of Sciences (2026).

Theory-driven computational psychiatry and its challenges: An optimistic outlook.
Fisher, Elizabeth L., Noor Sajid, Laura Convertino, and Jakob Hohwy. Neuroscience & Biobehavioral Reviews (2026).

Note: *This is a partial list of articles and preprints published by Kempner-affiliated researchers in the last month. Papers are listed by topic and publication/upload date, with the most recent first.*

View this report and other Kempner Research-Round-ups:
kempnerinstitute.harvard.edu/kempner-community/research-roundup/