

Research Roundup

February 2026

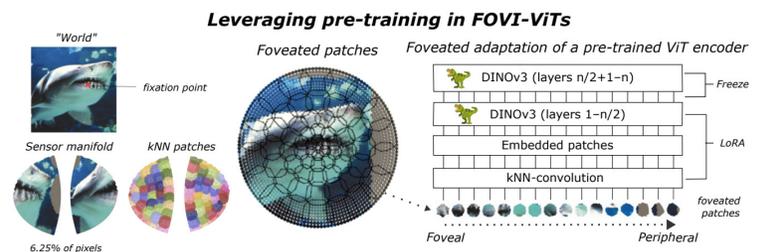
Discover our latest insights
in natural and artificial
intelligence research

⬡ A team led by Institute Investigator SueYeon Chung published a paper in *Nature Neuroscience* on **geometric properties of optimal neural coding for tasks with shared structure.**

⬡ Associate Faculty Patrick Slade and Harvard graduate student Haedo Cho introduce **OpenMetabolics**, "a biomechanically-informed activity monitor that employs a smartphone in a pants pocket which measures leg motion to estimate energy expenditure."

⬡ Research Fellow Gabriel Poesia and Harvard graduate student Simon Henniger introduce **The Token Games** (TTG), "an evaluation framework where models challenge each other by creating their own puzzles."

Featured Figure



Model architecture from the preprint "**FOVI: A biologically-inspired foveated interface for deep vision models.**"

Featured Formula

$$E_g = \frac{1}{\pi} \tan^{-1} \left(\sqrt{\frac{\pi}{2pc^2PR(\Psi)} + \frac{1}{f} + \frac{1}{s} - 1} \right)$$

Formula predicting generalization error from four geometric properties of neural representations, from the paper "**Neural population geometry and optimal coding of tasks with shared latent structure.**"

From the text: "we have written the generalization error as a strictly decreasing function of four geometric terms as well as the number of samples in the training dataset"

Featured Artifacts

Prescriptive Scaling Reveals the Evolution of Language Model Capabilities.

Hanlin Zhang, Jikai Jin, Vasilis Syrgkanis, and Sham Kakade.

[Get the Code](#) ▶

Learning Abstractions for Hierarchical Planning in Program-Synthesis Agents.

Zergham Ahmed, Kazuki Irie, Joshua B. Tenenbaum, Christopher J. Bates, and Samuel J. Gershman.

[Get the Code](#) ▶

A Critical Look at Targeted Instruction Selection: Disentangling What Matters (and What Doesn't).

Nihal V. Nayak, Paula Rodriguez-Diaz, Neha Hulkund, Sara Beery, and David Alvarez-Melis.

[Get the Code](#) ▶

AI innovation

LLMs Can Learn to Reason Via Off-Policy RL.

Daniel Ritter, Owen Oertell, Bradley Guo, Jonathan Chang, Kianté Brantley, and Wen Sun. arXiv:2602.19362v1 (2026).

Adaptive Time Series Reasoning via Segment Selection.

Shvat Messica, Jiawen Zhang, Kevin Li, Theodoros Tsiligkaridis, and Marinka Zitnik. arXiv:2602.18645v1 (2026).

A note on the dynamics of extended-context disordered kinetic spin models.

Jacob A. Zavatore-Veth and Cengiz Pehlevan. Journal of Physics A, Mathematical and Theoretical 59(4) :045001 (2026).

Graphon Mean-Field Subsampling for Cooperative Heterogeneous Multi-Agent Reinforcement Learning.

Emile Anand, Richard Hoffmann, Sarah Liaw, and Adam Wierman. arXiv:2602.16196v1 (2026).

Prescriptive Scaling Reveals the Evolution of Language Model Capabilities.

Hanlin Zhang, Jikai Jin, Vasilis Syrgkanis, and Sham Kakade. arXiv:2602.15327v1 (2026).

Boule or Baguette? A Study on Task Topology, Length Generalization, and the Benefit of Reasoning Traces.

William L. Tong, Ege Cakar, and Cengiz Pehlevan. arXiv:2602.14404v1 (2026).

Stop Training for the Worst: Progressive Unmasking Accelerates Masked Diffusion Training.

Jaeyeon Kim, Jonathan Geuter, David Alvarez-Melis, Sham Kakade, and Sitan Chen. arXiv:2602.10314v1 (2026).

Universal One-third Time Scaling in Learning Peaked Distributions.

Yizhou Liu, Ziming Liu, Cengiz Pehlevan, and Jeff Gore. arXiv:2602.03685v1 (2026).

Anytime Pretraining: Horizon-Free Learning-Rate Schedules with Weight Averaging.

Alexandru Meterez, Pranav Ajit Nair, Depen Morwani, Cengiz Pehlevan, and Sham Kakade. arXiv:2602.03702v1 (2026).

Learning Abstractions for Hierarchical Planning in Program-Synthesis Agents.

Zergham Ahmed, Kazuki Irie, Joshua B. Tenenbaum, Christopher J. Bates, and Samuel J. Gershman. arXiv:2602.00929v1 (2026).

Unsupervised Decomposition and Recombination with Discriminator-Driven Diffusion Models.

Archer Wang, Emile Anand, Yilun Du, and Marin Soljačić. arXiv:2601.22057v1 (2026).

Neuroscience & NeuroAI

Neural population geometry and optimal coding of tasks with shared latent structure.

Albert J. Wakhloo, Will Slatton, and SueYeon Chung. Nature Neuroscience 1-11 (2026).

FOVI: A biologically-inspired foveated interface for deep vision models.

Nicholas M. Blauch, George A. Alvarez, and Talia Konkle. arXiv:2602.03766v1 (2026)

How “Neural” is a Neural Foundation Model?

Johannes Bertram, Luciano Dyballa, T. Anderson Keller, Savik Kinger, and Steven W. Zucker. arXiv:2601.21508v1 (2026).

Sparse input representations explain odor discrimination in complex, concentration-varying mixtures.

Hannah McCalmon, George Cai, Constantine Tsibouris, Farhad Pashakhanloo, SueYeon Chung, Vikrant Kapoor, and Venkatesh N. Murthy. bioRxiv:2026.01.27.702074 (2026).

Applications of AI

Induction Meets Biology: Mechanisms of Repeat Detection in Protein Language Models.

Gal Kesten-Pomeranz, Yaniv Nikankin, Anja Reusch, Tomer Tsaban, Ora Schueler-Furman, and Yonatan Belinkov. arXiv:2602.23179v1 (2026).

Defining operational safety in clinical artificial intelligence systems.

Young-Tak Kim, Hyunji Kim, Manisha Bahl, Michael H. Lev, Ramon Gilberto González, Michael S. Gee, and Synho Do. NPJ Digital Medicine (2026).

OpenMetabolics: Estimating energy expenditure using a smartphone worn in a pocket.

Haedo Cho and Patrick Slade. Communications Engineering 5(1) :35 (2026).

STRAND: Sequence-Conditioned Transport for Single-Cell Perturbations.

Boyang Fu, George Dasoulas, Sameer Gabbita, Xiang Lin, Shanghua Gao, Xiaorui Su, Soumya Ghosh, and Marinka Zitnik. arXiv:2602.10156v1 (2026).

Scaling medical AI across clinical contexts.

Michelle M. Li, Ben Y. Reis, Adam Rodman, Tianxi Cai, Noa Dagan, Ran D. Balicer, Joseph Loscalzo, Isaac S. Kohane, and Marinka Zitnik. Nature Medicine 32(2) :439–48 (2026).



Interpretability and AI theory

Beyond Explainable AI (XAI): An Overdue Paradigm Shift and Post-XAI Research Directions.

Saleh Afroogh, Seyd Ishtiaque Ahmed, Petra Ahrweiler, David Alvarez-Melis, Mansur Maturidi Arief, Emilia Barakova, Falco J. Bargagli-Stoffi, et al. arXiv:2602.24176v1 (2026).

Emergent Manifold Separability during Reasoning in Large Language Models.

Alexandre Polo, Chanwoo Chun, and SueYeon Chung. arXiv:2602.20338v1 (2026).

The Token Games: Evaluating Language Model Reasoning with Puzzle Duels.

Simon Henniger and Gabriel Poesia. arXiv:2602.17831v1 (2026).

A Critical Look at Targeted Instruction Selection: Disentangling What Matters (and What Doesn't).

Nihal V. Nayak, Paula Rodriguez-Diaz, Neha Hulkund, Sara Beery, and David Alvarez-Melis. arXiv:2602.14696v1 (2026).

Cross-Modal Redundancy and the Geometry of Vision-Language Embeddings.

Grégoire Dhimoila, Thomas Fel, Victor Boutin, and Agustin Picard. arXiv:2602.06218v2 (2026).

Interpreting Physics in Video World Models.

Sonia Joseph, Quentin Garrido, Randall Balestrieri, Matthew Kowal, Thomas Fel, Shahab Bakhtiari, Blake Richards, and Mike Rabbat. arXiv:2602.07050v1 (2026).

Decomposing Query-Key Feature Interactions Using Contrastive Covariances.

Andrew Lee, Yonatan Belinkov, Fernanda Viégas, and Martin Wattenberg. arXiv:2602.04752v1 (2026).

A Random Matrix Theory Perspective on the Consistency of Diffusion Models.

Binxu Wang, Jacob Zavatore-Veth, and Cengiz Pehlevan. arXiv:2602.02908v1 (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/