



Research Roundup

July 2026

Discover our latest insights
in natural and artificial
intelligence research

⬡ Kempner visiting scholar Yonatan Belinkov and his collaborators present BetaDescribe, an AI system that translates a protein's amino acid sequence into a detailed natural-language description. Introduced in the journal ***PNAS***, the system describes properties including a protein's function, catalytic activity, involvement in metabolic pathways, location within a cell, and structural domains.

⬡ A team including Kempner Institute Investigator Yilun Du introduces ***DriftWorld***, a world model that helps robots plan by rapidly predicting the visual outcomes of possible actions.

⬡ A team including Kempner Graduate Fellow Mary Letey and Kempner associate faculty member Cengiz Pehlevan ***introduces*** an "effective context length" for LLMs, finding that a prompt with closely related examples can behave like a shorter prompt containing independent examples.

Featured Figure

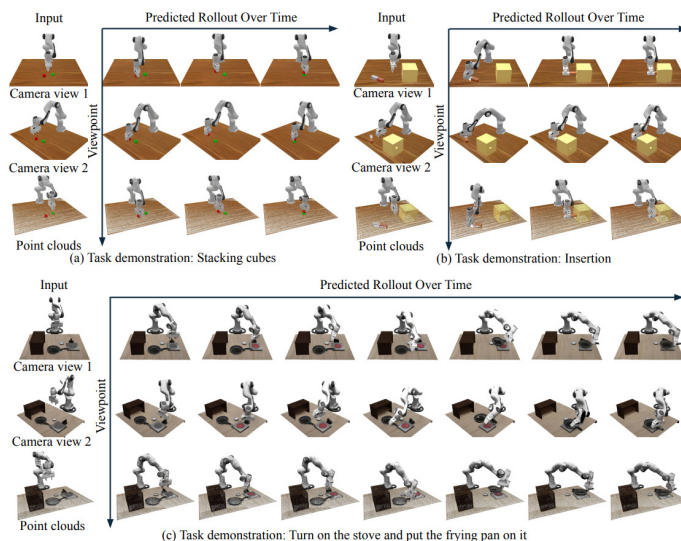


Diagram from the preprint ***Structured 4D Latent Predictive Model for Robot Planning***. Given input observations in the first column, the model "unrolls the 4D latent dynamics to generate future 3D structures over time."

Featured Artifacts

Formal Disco: Scalable Open-Ended Generation of Formally Verified Programs.

Gabriel Poesia, Simon Henniger, Tzu-Han Hsu, Yilun Du, and Nada Amin. (2026).

[Get the Code](#) ➤

DriftWorld: Fast World Modeling through Drifting.

Susie Lu, Haonan Chen, Weirui Ye, and Yilun Du. (2026).

[Get the Code](#) ➤

Featured Formula

$$s^*(H, B) \gtrsim 1 - \begin{cases} B d^{-\frac{a^2}{2a-1}}, & B \lesssim 1, \\ B^{\frac{a}{2a-1}} d^{-\frac{a^2}{2a-1}}, & 1 \lesssim B \lesssim d^{1/2} \\ d^{-a/2}, & B \gtrsim d^{1/2} \end{cases}$$

Inequality characterizing the optimal spectral radius of accelerated SGD, from the preprint ***Compute Efficiency and Serial Runtime Tradeoffs for Stochastic Momentum Methods.***



AI Innovation

DriftWorld: Fast World Modeling through Drifting.

Susie Lu, Haonan Chen, Weirui Ye, and Yilun Du. arXiv:2607.15065 (2026).

Formal Disco: Scalable Open-Ended Generation of Formally Verified Programs.

Gabriel Poesia, Simon Henniger, Tzu-Han Hsu, Yilun Du, and Nada Amin. arXiv:2607.04631 (2026).

OmniPilot: An Uncertainty-Aware LLM Inference Advisor for Heterogeneous GPU Clusters.

D. Balamurugan and Thomas W. Bush. arXiv:2607.01579 (2026).

Structured 4D Latent Predictive Model for Robot Planning.

Zhiyi Li, Peilin Wu, Xiaoshen Han, Ruojin Cai, and Yilun Du. arXiv:2607.01166 (2026).

MemoBench: Benchmarking World Modeling in Dynamically Changing Environments.

Haoyu Chen, Kaichen Zhou, Hang Hua, Kaile Zhang, Jingwen Qian, Wufei Ma, Haonan Chen, Chunjiang Liu, Yizhou Zhao, Xiaoyuan Wang, Weiyue Li, Alan Yuille, Paul Pu Liang, and Yilun Du. arXiv:2606.27537v6 (2026).

GB-LSR: A Fast Local Spectral Image Representation with a Single Global Bandwidth for Continuous Reconstruction and Super-Resolution.

Max Shad and Naeem Khoshnevis. arXiv:2606.19617 (2026).

Interpretability and AI Theory

The Log Log Jam in Gaussian State Tomography.

Sitan Chen, Weiyuan Gong, Qi Ye, and Zhihan Zhang. arXiv:2607.12983 (2026).

Sequential Correlations Change In-Context Learning: Effective Context Length and Architectural Mismatch.

Mary Letey, Yue M. Lu, Cengiz Pehlevan, and Jacob Zavatone-Veth. arXiv:2607.03660 (2026).

The State-Prediction Separation Hypothesis.

Giovanni Monea, Nathan Godey, Kianté Brantley, and Yoav Artzi. arXiv:2607.01218 (2026).

Compute Efficiency and Serial Runtime Tradeoffs for Stochastic Momentum Methods.

Depen Morwani, Alexandru Meterez, Pranav Nair, and Sham Kakade. arXiv:2606.19179v1 (2026).

Applications of AI

BetaDescribe: Providing Rich Descriptions from Protein Sequences.

Edo Dotan, Iris Lyubman, Marcelo Ehrlich, Eran Bacharach, Tal Pupko, and Yonatan Belinkov. Proceedings of the National Academy of Sciences 123(27): e2537345123 (2026).

Generalizable AI Predicts Immunotherapy Outcomes Across Cancers and Treatments.

Wanxiang Shen, Intae Moon, Thinh H. Nguyen, Michelle M. Li, Yepeng Huang, Nitya Nair, Daniel Marbach, and Marinka Zitnik. Nature Medicine (2026).

Using Natural Language Processing of Clinical Notes to Supplement Structured Electronic Health Record Data for Phenotyping Smoking and Obesity in a Healthcare System.

Jie Yang, Bowen Gu, Haritha Pillai, Joyce Lii, David Cronkite, Keith A. Marsolo, and Rishi J. Desai. Pharmacoepidemiology and Drug Safety 35(7): e70415 (2026).

BRIDGE: Benchmarking Large Language Models for Understanding Real-World Clinical Practice Texts.

Jiageng Wu, Bowen Gu, Ren Zhou, Kevin Xie, Doug Snyder, Yixing Jiang, Valentina Carducci, Richard Wyss, Rishi J. Desai, Emily Alsentzer, Leo Anthony Celi, Adam Rodman, Sebastian Schneeweiss, Jonathan H. Chen, Santiago Romero-Brufau, Kueiyu Joshua Lin and Jie Yang. Nature Biomedical Engineering (2026).

Neuroscience & NeuroAI

People Make Graded Judgments About the Inconceivable.

Jennifer Hu, Felix Sosa, and Tomer Ullman. Cognition 275: 106627 (2026).

The Extended Language Network: Language-Responsive Brain Areas Whose Contributions to Language Remain to Be Discovered.

Agata Wolna, Aaron Wright, Colton Casto, Samuel Hutchinson, Benjamin Lipkin, and Evelina Fedorenko. The Journal of Neuroscience (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 Roundups:
kempnerinstitute.harvard.edu/kempner-community/research-roundup/