In-context learning in natural and artificial intelligence

Akshay K. Jagadish (akshay.jagadish@helmholtz-munich.de)
Helmholtz Center for Computational Health, Munich, Germany

Ishita Dasgupta (dasgupta.ishita@gmail.com)
Google Deepmind

Jacques Pesnot Lerousseau (jacques.pesnot-lerousseau@univ-amu.fr)
University of Oxford, United Kingdom
Institute for Language, Communication, and the Brain, Aix-Marseille Université, Marseille, France

Marcel Binz (marcel.binz@helmholtz-munich.de)
Helmholtz Center for Computational Health, Munich, Germany

Keywords: in-context learning; neural networks; large language models; meta-learning

Introduction

In-context learning refers to the ability of a neural network to learn from information presented in its context (Brown et al., 2020). While traditional learning in neural networks requires adjusting network weights for every new task, in-context learning operates purely by updating internal activations without needing any updates to network weights. The emergence of this ability in large language models has led to a paradigm shift in machine learning and has forced researchers to reconceptualize how they think about learning in neural networks. Looking beyond language models, we can find in-context learning in many computational models relevant to cognitive science, including those that emerge from meta-learning (Binz et al., 2023).

The present workshop aims to delineate and discuss the implications of this phenomenon for the cognitive sciences. In order to accomplish this goal, we have invited experts who will present recent advances on the topic of *in-context learning in natural and artificial intelligence*. The selected speakers cover a broad spectrum of topics, including (a) understanding in-context learning from a computational perspective (Chan et al., 2022), (b) using in-context learning to model human behavior (Binz, Gershman, Schulz, & Endres, 2022), and (c) applications in neuroscience and linguistics (Whittington, Dorrell, Behrens, Ganguli, & El-Gaby, 2023).

There is growing evidence for the presence of in-context learning-like systems in humans (Binz et al., 2023). In contrast to traditional learning schemes for neural networks, in-context learning is fast and sample-efficient, and thereby able to capture the human ability to learn from just a few observations. It accomplishes this by having learned (in its weights) the inductive biases relevant to the environment it operates in. In-context learning is therefore able to exploit environmental structures such as learning curricula (Flesch, Balaguer, Dekker, Nili, & Summerfield, 2018) or compositional representations (Lake & Baroni, 2023) to learn rapidly from a few examples. Taken together, these features offer a new perspective on many aspects of human cognition — some of which we will cover in our workshop.

Goals and scope

The goals of this workshop are two-fold. First, for the first time ever, bring together researchers working on different aspects of in-context learning (including psychology, neuroscience, linguistics, and computer science). For this, we have invited a diverse group of researchers to map out the following questions:

- How well can human learning be modeled using in-context learning?
- Which neural architectures support in-context learning?
- When and why do natural and artificial systems rely on in-context versus in-weights learning?
- How does in-context learning relate to classical concepts from cognitive science?

In addition, we aim to make the concept of in-context learning more accessible to a wider audience. To accomplish this goal, we have included an introductory lecture on the topic as well as a podium discussion relating in-context learning to classical cognitive concepts.

Target audience

We aim to target an interdisciplinary group almost as wide as the conference itself, including psychologists, neuroscientists, linguists, and computer scientists. The *dynamics of cognition* theme of this year's conference provides the perfect background for our workshop, as in-context learning algorithms are —by definition— systems that dynamically adapt to new situations.

Organizers and presenters

Akshay K. Jagadish (Organizer) is a PhD student at the Max Planck Institute for Biological Cybernetics, Tübingen. His current research is dedicated towards understanding the ingredients essential for explaining human adaptive behavior across multiple task domains.

In L. K. Samuelson, S. L. Frank, M. Toneva, A. Mackey, & E. Hazeltine (Eds.), *Proceedings of the 46th Annual Conference of the Cognitive Science Society.*

Ishita Dasgupta (Organizer) is a research scientist at Google Deepmind. She uses advances in machine learning to build models of human reasoning, applies cognitive science approaches toward understanding black-box AI systems, and combines these insights to build better, more human-like artificial intelligence.

Jacques Pesnot Lerousseau (Organizer) is a postdoc at the Institute for Language, Communication, and the Brain, Marseille. His current research addresses the question of in-context learning in human brains and artificial neural networks, aiming to uncover the mechanisms behind rule generalization in the brain and algorithms.

Marcel Binz (Organizer) is a research scientist at Helmholtz Munich. He works on modeling human behavior using ideas from meta-learning, resource rationality, and language models.

Roma Patel is a fourth-year PhD student at Brown University. Her research uses language to structure reinforcement learning, aiming towards building more intelligent and interpretable agents that can learn to use language to communicate and coordinate with each other.

Stephanie Chan is a senior research scientist at Google Deepmind. Having a background in both cognitive and computer science, she studies how data distributional properties drive emergent in-context learning.

James Whittington is a Sir Henry Wellcome postdoctoral fellow at Stanford University & the University of Oxford. He works on building models and theories for understanding structured neural representations in brains and machines.

Tom McCoy is an assistant professor in the Department of Linguistics at Yale University. He studies the computational principles that underlie human language using techniques from cognitive science, machine learning, and natural language processing.

Greta Tucktuke is a PhD candidate at the Department of Brain and Cognitive Sciences at MIT. She studies how language is processed in the biological brain, and how the representations and processes in artificial neural networks models compare to those in humans.

The panel will be moderated by **Christopher Summerfield** and includes **Morgan Barense**, **Alison Gopnick**, **Tom Griffiths**, **Micha Heilbron** and **Brenden Lake** as panelists.

Workshop structure

We propose a full-day workshop consisting of four parts. The first three are sessions that involve a set of three talks (20 minutes plus 5 minutes for discussions). They are centered around in-context learning in computer science, psychology, and neuroscience/linguistics respectively. The workshop concludes with a panel that discusses the broader implications and connections of in-context learning in cognitive science. The topics of the talks are as follows:

Presenter	Topic
Binz	Introduction to in-context learning
Chan	What do you need for in-context learning in transformers?
Patel	Towards understanding the (conceptual) structure of language models
Dasgupta	Concepts and categories within context
Jagadish	Ecologically rational meta-learned inference explains human category learning
Pesnot Lerousseau	Training data distribution drives in-context learning in humans and transformers
Whittington	Different algorithms for in-context learning in prefrontal cortex and the hippocampal formation
Tucktuke	Modeling human language processing using large language model
McCoy	Understanding and controlling neural networks through the problem they are trained to solve
Summerfield, Barense, Gopnick, Griffiths, Heilbron, Lake	Podium discussion

References

Binz, M., Dasgupta, I., Jagadish, A. K., Botvinick, M., Wang, J. X., & Schulz, E. (2023). Meta-learned models of cognition. *Behavioral and Brain Sciences*, 1–38.

Binz, M., Gershman, S. J., Schulz, E., & Endres, D. (2022). Heuristics from bounded meta-learned inference. *Psychological review*.

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., ... others (2020). Language models are few-shot learners. *Advances in neural information processing systems*, *33*, 1877–1901.

Chan, S., Santoro, A., Lampinen, A., Wang, J., Singh, A., Richemond, P., ... Hill, F. (2022). Data distributional properties drive emergent in-context learning in transformers. *Advances in Neural Information Processing Systems*, *35*, 18878–18891.

Flesch, T., Balaguer, J., Dekker, R., Nili, H., & Summerfield, C. (2018). Comparing continual task learning in minds and machines. *Proceedings of the National Academy of Sciences*, *115*(44), E10313–E10322.

Lake, B. M., & Baroni, M. (2023). Human-like systematic generalization through a meta-learning neural network. *Nature*, 1–7.

Whittington, J. C., Dorrell, W., Behrens, T. E., Ganguli, S., & El-Gaby, M. (2023). On prefrontal working memory and hippocampal episodic memory: Unifying memories stored in weights and activation slots. *bioRxiv*, 2023–11.