ICLR 2026 Workshop on **Representational Alignment** (Re^4-Align)

🤖🧠💭

tl;dr: Representational alignment among artificial and biological neural systems continues to be a rapidly growing research area across machine learning, neuroscience, and cognitive science communities; we counted 688 papers submitted to ICLR 2026 on this set of interdisciplinary topics,[1] up from 443 papers submitted to ICLR 2025, and 303 to ICLR 2024, representing an average 51% yearly increase. The Re-Align Workshop at ICLR 2026 facilitates interdisciplinary discussion among these communities, highlights unexpected findings from last year's hackathon, and pushes beyond the foundational questions of alignment addressed in the previous workshops to focus on two novel and critical interdisciplinary applications of representational alignment: enabling neural control via representational alignment and evaluating the downstream behaviors that are enabled by representational alignment.

Contents

1 Summary

The first two editions of the Workshop on Representational Alignment (Re-Align) established a vibrant community bringing together researchers from **machine learning, neuroscience, and cognitive science** to tackle a foundational question: how can we meaningfully compare and align the internal representations of intelligent systems (Sucholutsky et al., 2023)? Building on this foundation, the third edition of the Re-Align Workshop (and fourth overall Re-Align event) pivots from asking *how* we measure alignment to *what we can conclude from observing alignment*. In particular, our next edition brings focus on what affordances alignment makes possible. In other words, **what can we do with alignment?** This focus on control and intervention directly engages with topics of broad interest to the ICLR community, including AI safety, interpretability, and the creation of more robust models. This is where a deep, bidirectional exchange of insights is invaluable: cognitive science and neuroscience provide theories of how the brain implements control, while AI offers powerful, manipulable models to test these theories. This year's workshop asks: Can we leverage representational alignment to intervene on and control both biological and artificial systems? Concretely, we are particularly interested in

1. **Neural control.** When does representational alignment allow us to meaningfully intervene on a system's behavior? In AI, this connects to the goals of mechanistic interpretability (Olah et al., 2020; Nanda et al., 2023) and the engineering challenge of building safer, steerable models. In neuroscience, it parallels the long-standing goal of understanding how local neural activity gives rise to global function. By exploring how to control representations of specific functions or concepts (Bau et al., 2017; Todd et al., 2023), we create a shared framework for moving from simply mapping circuits to actively understanding their causal role in both artificial and biological systems.
2. **Downstream behavior.** Representational alignment may also have consequences for deployment. Task vectors (Hendel et al., 2023) reveal the structure needed to adapt a model's behavior to a new task by adding, removing, or composing computations in representational space and are parallel to cognitive control

[1] via keywords: *neuroai|neuro|cognitive|cognitive science|cognitive sciences|cogai|behavior*

in biological intelligence, where the prefrontal cortex actively manipulates representations to guide flexible behavior (Miller & Cohen, 2001; Botvinick et al., 2001). Findings like task vectors invite a new set of research questions in representational alignment: When might aligned representations enable control on downstream behavior? And can we leverage alignment to build generalizable "control interfaces"—subspaces or other affordances where interventions generalize across tasks and models? Answering these questions demands evaluation frameworks that probe the relationship between representational alignment and flexible recontextualization (Studdiford et al., 2025) and that extend to complex domains such as collaboration and communication (Murthy et al., 2022; Fan et al., 2021) where this relationship may pay off.

We invite contributions within these two topic areas, as well as other topic areas related to representational alignment described in our prior workshop propposals in 2024 and 2025 and our position paper (Sucholutsky et al., 2023).

Finally, we invite participants to contribute through several avenues that bridge hands-on application and discussion. Building on our inaugural hackathon and the active debate around similarity metrics (Schaeffer et al., 2024; Lampinen et al., 2024), we will introduce a **challenge report track** (similar to the "task report" from Baby LM) for novel analyses stemming from our hackathon, and a **findings track**, which is similar to the workshop paper track from prior years, but also permits perspectives on the challenge.

1.1 New Component: Re -Align Challenge Leaderboard & Submission Track

Research in representational alignment converges on central questions but diverges in its answers. Studies report unexpected failures in validating alignment (Han et al., 2023), unclear driving factors (Conwell et al., 2024), and conflicting conclusions across analyses (Elmonzino & Bonner, 2024). At our previous edition, we addressed these inconsistencies through a **Representational Alignment Hackathon** with standardized stimulus sets and similarity measures. Participants formed two teams with opposing objectives: **the Blue Teams** investigated model universality by identifying heterogeneous populations that showed strong alignment, while **the Red Teams** examined model variability by revealing systematic differences among homogeneous populations that were expected to align. This standardized framework facilitated directly comparable insights and generated strong community engagement. Building on this momentum, participants identified clear next steps—consolidating findings across teams, enabling continuous comparison beyond the hackathon, and developing shared infrastructure to sustain collaboration.

To realize these goals, we propose **transforming the hackathon into a persistent shared-task challenge**—a community-driven platform promoting transparency, reproducibility, and collaboration in representational alignment research. This initiative will establish a common language for comparing and understanding alignment, enabling sustained progress through standardized models, shared datasets, and continuous evaluation. As part of this **Representational Alignment Challenge** our **new initiatives** are:

- The **Re -Align Leaderboard** will be a continuously updated, community-driven platform where researchers can submit models and datasets to engage in a friendly competition on representational alignment challenges over time. The leaderboard evaluates all submissions on shared benchmarks using pre-computed, group-level similarity metrics on withheld datasets, with results showcased on our official website. Inspired by successful initiatives such as BabyLM, this platform bridges the hackathon participation from last year with long-term, standardized benchmarking, promoting reproducibility, transparency, and collaboration across the community.

- The **Challenge Report Track** will be a dedicated paper track at the workshop, inviting participants to submit short reports describing their models, analyses, and findings from the challenge. Submissions will undergo light peer review, following the spirit of the recent ICLR Blogpost initiative, and will be compiled into a collective volume to ensure visibility and citability for emerging research. Beyond leaderboard results, the track also welcomes creative and critical contributions related to the representational alignment challenge—for instance, addressing questions like *"Why doesn't the next version of the Re-Align Challenge do cool idea X?"*—as well as introducing novel stimulus sets, evaluation metrics, or modeling approaches. To capture the broader impact, the organizers plan to author a synthesis paper highlighting *track winners* and *compelling negative results*, distilling the key advances and insights emerging across all submissions, after the workshop takes place, similar to the "post-competition analyses" called for as part of the NeurIPS Competition Track.

The Leaderboard will launch early 2026, offering an open platform where participants can access a standardized stimulus set, model repository, and similarity metrics to engage with the challenge. Submissions to the Challenge Report Track will be accepted until February 2026 and will undergo a light peer-review process, coordinated with the main research paper track to ensure aligned notification and acceptance timelines. The final synthesis and recognition of highlights, including track winners and outstanding reports, will be presented during the workshop event.

1.2 Anticipated audience

Re-Align was attended by more than 150 in-person participants at each of ICLR 2024 and ICLR 2025 (despite the community concerns about difficulties with travel to Singapore in 2025!); we expect a similar number of participants this year. We have a fairly even distribution of participants across ML, neuroscience, and cognitive science (as well as their intersections) with ML somewhat more represented than the other two areas as is expected from an event at ICLR. We had the following spread of papers across the previous two workshops: 45 machine learning, 22 neuroscience, and 27 cognitive science as identified by paper authors, which nicely reflected our interdisciplinarity. We aim for a similar distribution (approximately 2:1:1) this year. We highlight the following feedback from the first two years of the event on what people liked:

> 2025: *fantastic, highly interdisciplinary setting*
>
> 2025: *The poster sessions were really good.*
>
> 2025: *The gender balance among participants was better than any of the other workshops I attended.*
>
> 2024: *It was great throughout: the keynotes, meeting [sic] people from different fields; I really liked the size not too many people, and also not to [sic] few; getting a lot of input on a general level through the talks and on a more specific level in the poster session..etc.*
>
> 2024: *The amazing line-up of speakers, the quality of the posters, and the clear (email) communication beforehand.*
>
> 2024: *It was a great workshop. The talks and the posters were great. I particularly enjoyed the organized lunch.*
>
> 2024: *I really enjoyed seeing a lot of other work similar to mine and the possibility to talk to people working on the same topic.*

1.3 Prior context: Papers, workshops, debates

1.3.1 Papers

The alignment of representations between humans and machines remains a critical area of research at the intersection of cognitive science, neuroscience, and machine learning, focusing on how internal representations—whether biological or artificial—reflect structured information from the external world (Cao, 2022). Since the previous two editions of our workshop, the field has made substantial progress in understanding and comparing alignment between artificial and biological systems. Researchers have developed sophisticated metrics that differentiate models beyond simple representational comparisons (McNeal et al. 2024; Harvey, et al. 2024), investigated what alignment reveals about shared computational strategies (Ammar et al. 2025; Sartzetaki et al., 2025), developed methods to systematically modulate alignment between systems (Sundaram et al., 2024; Muttenthaler et al., 2024; Moussa et al., 2025), and clarified whether aligned representations reflect convergent computational principles or superficial similarities (Chen & Bonner, 2025; Hosseini et al., 2025).

Yet increasingly, the field is pivoting from measuring alignment to exploiting it—asking not just how much systems align, but **what interventions that alignment enables**. This shift is evident in recent work on **interventional approaches** that leverage representational alignment to **causally understand and control systems**. Recent work has applied mechanistic interpretability methods to identify task-specific representations and edit model behavior (Park et al., 2024; Casademunt et al. 2025). Other researchers have begun to causally intervene on alignment between biological and artificial systems and test causal hypotheses about underlying mechanisms such as those involving language (AlKhamissi et al., 2025; Moussa et al., 2025).

Beyond internal mechanisms, this interventional approach extends to understanding how aligned representations **support downstream task performance and flexible behavior**. For example, recent work has focused on leveraging representational alignment to facilitate adaptive, goal-directed behavior (Stolfo et al., 2025; Zhu et al., 2025), context-aware reasoning (Marjieh et al., 2025; Wang et al., 2025), efficient learning and teaching (Sucholutsky et al., 2023; Sucholutsky et al., 2025), and effective communication of values and preferences (Rane et al., 2024; Wynn et al., 2024). This convergence marks a fundamental transition in the field. Our third edition of the Re-Align ICLR workshop provides a timely platform for the interdisciplinary dialogue needed to advance this emerging research frontier.

1.3.2 Workshops

This proposal is for the third iteration of the Re-Align Workshop at ICLR. Our first two workshops successfully established a vibrant community at the intersection of neuroscience, cognitive science, and machine learning, united by the theme of representational alignment. Over the past summer, we also hosted a discussion event at CCN, the Re^3-Align Collaborative Hackathon, which discussed the hackathon we launched at the ICLR workshop last year, and provided guidance for our proposed challenge track at this year's ICLR Workshop. Alongside our efforts, the representational alignment landscape at major ML conferences has expanded considerably. At ICLR 2025, the Workshop on Bidirectional Human-AI Alignment examines mutual adaptation between humans and AI systems. At NeurIPS 2025, emerging workshops such as CogInterp and Data on the Brain & Mind join established venues like UniReps in exploring representations from diverse perspectives, spanning theory, methods, and datasets. However, **Re-Align occupies a unique position by centering representational alignment as a fundamental question across both artificial and biological neural systems**. While other venues treat alignment as one lens among many, we position it as a core lens for understanding intelligence. This year, we also extend beyond investigating how to align representations to critically exploring what such alignment enables—particularly for control, intervention, and steering of intelligent systems. This dual focus positions Re-Align as a vital bridge connecting foundational research in alignment with the practical challenges of building more interpretable, safe, and robust machine learning models.

1.3.3 Debates

Representational alignment has become a central topic in contemporary discussions about the nature of intelligence. Over the past two years, the field has engaged in lively debate around a core question: *How universal—or system-specific—are the representations that intelligent agents, biological or artificial, form about the world?* This question has captured sustained attention within the machine learning community and growing interest among neuroscientists and cognitive scientists (Muttenthaler et al., 2023; Huh et al., 2025; Chen & Bonner, 2025). The question has surfaced in well-attended community debates such as Universality and Idiosyncrasy of Perceptual Representations, in disagreements over appropriate metrics and methodologies, and in broader conversations about what "alignment" truly means. Yet despite the enthusiasm surrounding these discussions, definitive answers remain elusive. This year, we aim to confront these tensions directly through the Representational Alignment Challenge—a new initiative designed to provide a structured and rigorous framework for testing competing hypotheses.

A second, increasingly urgent question has risen to prominence but has yet to receive focused community attention: *If representations align, what does that alignment actually mean—and why should we care?* As the field reports increasingly higher scores of representational alignment, there is mounting pressure to clarify its interpretability and implications. Critics have argued that correlation-based analyses cannot establish causality (Chen et al., 2025), that learned representations may be systematically biased towards certain features (Lampinen et al., 2025), and that representational similarity can exist without functional equivalence (Braun et al. 2025). These concerns strike at the heart of what representational analyses can—and cannot—reveal about intelligence itself. We believe this tension can no longer remain siloed within individual papers or review exchanges. Our workshop this year will bring this debate to the forefront of community discourse by inviting discussion on the causal control and downstream applications of representational alignment.

2 Speakers & panelists

2.1 Invited speakers

All speakers below have confirmed their plans to **give an invited talk in person at the workshop.** We have invited these individuals as they are all researchers who have published high-impact and often interdisciplinary works in neuroscience, machine learning, and cognitive science. We give their biographies below, highlighting their cross-disciplinary expertise in machine learning (🤖), cognitive science (💭), and neuroscience (🧠).

David Bau (Northeastern University, USA) is an assistant professor in the Khoury College of Computer Sciences. He studies artificial intelligence, with a focus on machine learning interpretability, computer vision, and natural language processing. He completed his Ph.D. in Computer Science at MIT after a 20-year career as a software engineer at Google and Microsoft, where he helped lead the development of products like Google Image Search and Internet Explorer. His work has particularly impacted the field of AI interpretability, pioneering methods like Network Dissection and model editing to locate and alter factual knowledge within large language models. Dr. Bau's research has been recognized by a Sloan Research Fellowship and an NSF Graduate Research Fellowship. His teaching has been recognized with the Ruth and Joel Spira Award for Excellence in Teaching. His current research, which he likens to "the neuroscience of AI," focuses on understanding the internal mechanisms of large models to make them more transparent and controllable. He also directs the National Deep Inference Fabric, a $9M NSF-funded national research infrastructure for large-scale AI. 🤖🧠 David's work on interpretability (Todd et al., ICLR 2024) in large-scale machine learning models, especially his work on concept editing (Meng et al., ICLR 2023, Gandikota et al., ICCV 2023; Gandikota et al., WACV 2024), is highly relevant to this year's workshop theme on neural control.

Arturo Deza (Artificio, Peru) is an Assistant Professor in Computer Science at UTEC in Lima, Peru, and Co-Founder and CEO of Artificio, a Moonshot R&D company developing a benchmarking platform for the self-driving car industry. He completed his Ph.D. in Dynamical Neuroscience at UC Santa Barbara under Miguel Eckstein, focusing on biologically inspired vision models and the mechanisms of foveation in humans and machines. He previously held postdoctoral positions at MIT's McGovern Institute for Brain Research and Center for Brains, Minds and Machines with Tomaso Poggio, and at Harvard University with Talia Konkle. His work bridges vision science, computer vision, and machine learning, exploring how human visual processing can inform advanced computer and robot vision systems. 💭🤖 Arturo's work on human–machine alignment in autonomous driving is highly relevant to this year's workshop theme on the downstream applications of representational alignment (Cusipuma et al., CVPR 2025).

Alona Fyshe (University of Alberta, Canada) is an Associate Professor with joint appointments in Computing Science and Psychology at the University of Alberta, Canada. She is a Fellow of the Alberta Machine Intelligence Institute and holds a Canada CIFAR AI Chair. Alona's research lies at the intersection of computational linguistics, machine learning, and neuroscience. She completed her Ph.D. at Carnegie Mellon University under the supervision of Tom Mitchell and later held a position at the University of Victoria. Her work uses brain imaging techniques such as EEG and fMRI to compare human language processing with computational language models, bridging cognitive neuroscience and artificial intelligence. Currently, she focuses on developing more efficient language models inspired by how the human brain processes language. 🧠🤖 Alona's work on aligning representations between the brain and deep neural networks—highlighted by her recent study on decodable concepts in the brain (Efird et al., ICLR 2025)—is highly relevant to this year's workshop theme on neural control.

Phillip Isola (Massachusetts Institute of Technology, USA) is the Class of 1948 Career Development associate professor in EECS at MIT. He studies computer vision, machine learning, robotics, and AI. He completed his Ph.D. in Brain & Cognitive Sciences at MIT, and has since spent time at UC Berkeley, OpenAI, and Google Research. His work has particularly impacted generative AI and self-supervised representation learning. Dr. Isola's research has been recognized by a Google Faculty Research Award, a PAMI Young Researcher Award, a Samsung AI Researcher of the Year Award, a Packard Fellowship, and a Sloan Fellowship. His teaching has been recognized by the Ruth and Joel Spira Award for Distinguished Teaching. His current research focuses on trying to scientifically understand human-like intelligence. 🤖💭 Phil's work on universal representations across modalities in machine learning systems is highly topical for the workshop (Huh et al., ICML 2024; Sundaram et al., NeurIPS 2024).

Danielle Perszyk (Amazon, USA) is a cognitive scientist and Member of Technical Staff at Amazon's AGI SF Lab, where she leads the Human-Computer Interaction team in developing foundational capabilities for AI agents that operate seamlessly across digital and physical environments. Her pioneering approach shifts the industry's focus from asking "Is AGI within our reach?" to "What kind of general intelligence should we reach for?"—championing the development of useful general intelligence that augments human capabilities rather than replacing them. Drawing from her Ph.D. research at Northwestern University on language evolution and social cognition, she brings a unique perspective on how intelligence emerges and adapts. She previously applied these insights at Adept to build better human-AI interactions using language development principles, and at Google where she developed research programs for collecting human feedback to train AI models. At Amazon AGI, Dr. Perszyk is advancing AI agents that can think, act, and collaborate as naturally in the physical world as in the digital one. Her vision positions AI as a "collective subconscious" for repetitive tasks while ensuring humans remain the primary decision-makers. Dr. Perszyk has shared her insights at major industry venues including AI Engineer World's Fair, WebSummit, Columbia University's DAPLab workshop, and TechCrunch Sessions: AI, establishing herself as a leading voice in human-centered AI development. 💭🤖 Danielle's work on developing agents that understand both digital systems and the social-cognitive contexts driving human intelligence—combined with her framework for human-AI collaboration that keeps humans in control (Perszyk, AI Engineer World's Fair 2024; Perszyk, Amazon AGI technical reports 2024)—is highly topical for the workshop.

2.2 Invited panelists

We have sufficient speakers (5) for our panel, but will draw from our excellent shortlist of possible panellists if a speaker or two have a conflict; our shortlist comprises mid- and senior career researchers who have authored or spoken at one of the two prior Re-Align workshops.

3 Schedule

We give a tentative workshop schedule below.

start	dur.	event
8:45	0:15	opening remarks
9:00	0:30	invited talk: David 🤖🧠
9:30	0:30	invited talk: Arturo 💭🤖
10:00	0:15	contributed talk
10:15	0:15	contributed talk
10:30	0:20	discussion + coffee
10:50	1:40	poster session
12:30	1:45	lunch
14:15	0:30	invited talk: Alona 🧠🤖
14:45	0:30	invited talk: Phil 🤖💭
15:15	0:20	discussion + coffee
15:35	0:15	contributed talk
15:50	0:30	invited talk: Danielle 💭🤖
16:20	1:00	panel: David, Arturo, Alona, Phil, Danielle (and/or panellists)
17:20	0:10	closing remarks
17:30		**FIN.**

4 Diversity & inclusion

4.1 Diversity among organizers & invited participants

Representational alignment is an interdisciplinary research area that thrives on diverse perspectives. To that end, we have an organizing team and invited speaker roster with representing different fields (machine learning 🤖, cognitive science 💭, and neuroscience 🧠), a range of career stages (Ph.D. student to faculty on the organizing team; junior to senior faculty and researchers on the speaker roster), and distinct affiliations (12 affiliations for 13 individuals). We ensured that women take on both speaking (2 of 5) and organizing (3 of 8) roles at the workshop.

Our speakers and organizers represent distinct institutions across both industry and academia, and span geographic locations including the US, Canada, Europe, and South America. We have historically made a concerted effort to increase geographic diversity especially in the regions where ICLR is taking place (we recruited two APAC speakers last year) and this year are excited to welcome Arturo Deza, who is based in Peru 🇵🇪, as an invited speaker. Arturo's highly topical work is relevant to Re-Align, and we hope to leverage this connection to build stronger ties with the cognitive / neuroscience / AI community in South America.

4.2 Inclusive access to workshop components

Talks and posters. To ensure broad access to the workshop for those who are unable to attend in person, we are implementing several strategies to make the more standard components (invited talks, contributed talks and posters, panels) of our workshop accessible for remote participants. We are anticipating support from ICLR to have talks, panels, and key activities live-streamed via SlidesLive, with all participants including virtual attendees of ICLR able to join in the discussions. Essential materials, such as workshop papers, posters, and presentation slides, will also be published on our website to ensure ongoing access, like the past two year.

Contributing. Like last year, we will have two research tracks: a tiny / short paper track (3–5 pages) and a long paper track (up to 9 pages). Again like last year, reviewers will be instructed to evaluate papers within the track they were submitted to, so as not to penalize new contributors to the tiny / short paper track. This shorter length track provides an entry point for those who may be hesitant about submitting a full-length paper, in line with the ICLR 2026 Workshop guidance on a "Platform for Tiny or Short Papers." New this year, we will also have an inaugural challenge report track (3-5 pages), building on our inaugural hackathon last year; see Section 1.1. The Representational Alignment Challenge and associated challenge report track provide an additional entry point for those who want to get involved in representational alignment research, but are unsure where to start, by leveraging a common framework and dataset to provide participants with a structured and accessible way to contribute.

5 Workshop processes

5.1 Contributions: Submission & review process

We will accept contributed papers and have a formal peer review process facilitated by OpenReview. This year we will have three tracks. As in the previous two years we will have two research tracks: a **tiny / short paper track (3–5 pages)** and a **long paper track (up to 9 pages)**; but additionally we will now have a **challenge report track (3-5 pages)** for hackathon participants to report their approaches and findings. Again like the previous years, reviewers will be instructed to evaluate papers within the track they were submitted to, so as not to penalize new contributors and new ideas to the tiny / short paper track. All reviewers will be asked to list their conflicts of interest ahead of time and will be assigned papers accordingly to ensure a fair review process. Each paper will be reviewed by a minimum of 3 reviewers and the goal will be for each reviewer to be assigned no more than 3 papers (see Section 5.7 for details on our program committee). The diverse range of research areas represented on our organizing team will ensure that we can step in as emergency reviewers on any papers that have received fewer than 3 high-quality reviews by the reviewing deadline. The organizing committee will act as program chairs in making acceptance decisions given the reviewer evaluations; organizers with conflicts for a specific submission (due to collaboration, institutional affiliation, etc.) will recuse themselves from the decision process on that submission.

5.2 Outreach

We plan to advertise the workshop across several social media platforms, including Twitter/X, Mastodon, and Bluesky, as we did last year. with the goal of attracting a broad audience, including those who can't participate in person. We also plan to update the workshop website from the last two years (representational-alignment.github.io) after the proposal notification date, when planning for the workshop would be in full swing.

5.3 Sponsorship

At ICLR 2024, we secured sponsorship from a local Viennese company interested in machine learning (EY Vienna), which enabled us to host a sponsored community lunch that was a highlight for many of our participants (see "Anticipated Audience" above). This year, we plan to recruit sponsors for another community lunch as well as compute credits for participants in the challenge track. In our experience, we don't receive firm commitments from sponsors until the workshop has been confirmed to take place.

5.4 Dates & deadlines

We have established a submission, reviewing, and notification schedule for contributed papers in line with the ICLR 2026 Workshop Proposal guidelines, as follows:

Thursday, February 5th, 2026	submission deadline
Thursday, February 26th, 2026	internal reviewing deadline
Sunday, March 1st, 2026	notification date
Monday, April 20th, 2026	camera-ready copy deadline
April 27th or 28th, 2026	workshop date!

6 Committees

6.1 Organizing committee

🧰 denotes prior experience organizing workshops, challenges, and other community events and initiatives.
🔨 denotes the role that each organizer will take on at the ICLR 2026 Workshop on "Representational Alignment."

Badr AlKhamissi (EPFL, Switzerland) is a PhD candidate at EPFL, working with Antoine Bosselut (NLP Lab) and Martin Schrimpf (NeuroAI Lab). His research sits at the intersection of machine learning, neuroscience, and cognitive science, with a focus on developing language models that are better aligned with the human brain and behavior. Prior to EPFL, Badr was an AI Resident at Meta AI in Seattle, collaborating with Mona Diab and Asli Celikyilmaz, and spent a year as a Research Intern at Sony AI with Michael Spranger. He holds an MSc in Computational Cognitive Neuroscience (with distinction) from Goldsmiths, University of London, and a BSc in Computer Science (Summa Cum Laude) from the American University in Cairo. 🧰 Badr co-organized the ICML 2024 Workshop on LLMs and Cognition, served as a Social Chair for the ArabicNLP 2024 Conference, and as a Publicity Chair for the ArabicNLP 2025 Conference. He has also served on the program committees of several conferences and workshops (e.g., NeurIPS, ACL, EMNLP), including as an Area Chair for ACL Rolling Review. Badr developed and maintains the Egyptian in AI Research website, with the goal of showcasing Egyptian talent in AI and related fields. The initiative has inspired the creation of several spinoff communities, including Moroccans in AI Research and the Pakistanis in AI Research. Badr is also an active contributor to the Brain-Score repository which aims to improve the efficiency of communication between experimentalists and modelers by providing experimental data as accessible benchmarks, and providing unified computational models to experimentalists. 🔨 Badr will support on the **new submission track** initiative.

Brian Cheung (UC San Francisco, USA) is an incoming Assistant Professor in the Department of Bioengineering and Therapeutic Sciences at UCSF. Brian studies the convergence of representations across multiple levels: from the structural aspects of how intelligence is accomplished in biology and in-silicon, to the nature of how meaningful representations are generated from raw inputs, all the way to how these systems ultimately make decisions. Brian received his PhD from the Redwood Center for Theoretical Neuroscience at UC Berkeley while being advised by

Bruno Olshausen. Before starting at UCSF, Brian is currently a postdoc at MIT Brain and Cognitive Sciences and CSAIL working with Boris Katz, Tomaso Poggio and Phillip Isola. 🧰 Brian has co-organized the 2025 Brains, Minds, and Machines Summer School in Woods Hole. He has also co-organized the 2025 Re-Align workshops in ICLR 2025 and CCN 2025 and the Universality and Idiosyncrasy of Perceptual Representations Community event at CCN 2025. 🔨 Brian will advise on the 2026 workshop **challenge**.

Dota Dong (Max Planck Institute for Psycholinguistics, the Netherlands) is a Ph.D. student in Computational Cognitive Neuroscience at the Max Planck Institute (MPI) for Psycholinguistics and the Donders Center for Cognitive Neuroimaging. Dota studies how biological and artificial neural networks learn multimodal semantic representations from real-world experiences in both adults and infants. She uses computational methods to explore these questions, in conjunction with data and theories from neuroscience, linguistics, and psychology. 🧰 Dota organized a workshop at ICLR 2025 and a community event at CCN 2025, serves as the DEI Chair for CCN 2026, and was a member of the program committee for CCN 2025. She reviewed for multiple workshops at ICLR and NeurIPS, cognitive neuroscience conferences including CogSci and CCN, and served as an invited early-career researcher (ECR) reviewer for *Nature Communications*. 🔨 Dota will lead the **new submission track** initiative.

Erin Grant (University of Alberta, Canada) is an incoming Assistant Professor in the Departments of Psychology & Computing Science at the University of Alberta and a Fellow at the Alberta Machine Intelligence Institute (Amii). Erin studies prior knowledge and learning mechanisms in minds, brains, and machines using a combination of behavioral experiments, computational simulations, and analytical techniques, with the goal of grounding higher-level cognitive phenomena in a plausible neural implementation. Erin earned her Ph.D. in Computer Science from UC Berkeley, and during her Ph.D., spent time at OpenAI, Google Brain, and DeepMind. 🧰 Erin has co-organized 9 workshops at NeurIPS, ICML, and ICLR, in addition to events at other conferences: the hybrid (2018, 2020) and virtual (2021) NeurIPS Workshops on Meta-Learning; the hybrid ICLR 2019 Workshop on Structure & Priors in RL; the virtual NeurIPS 2020 Women in Machine Learning Affinity Workshop; the in-person ICLR 2024 and ICLR 2025 Re-Align Workshops, the in-person ICML 2024 Workshop on "In-Context Learning," and the in-person NeurIPS 2025 Workshop on "Data on the Brain & Mind." She has also served on the program committee for 29 workshops at ACL, ICML, ICLR, and NeurIPS. Erin has led diversity and inclusion at machine learning conferences as a Diversity, Inclusion / Next Generation & Accessibility Chair at NeurIPS 2022, 2023, and 2025, and a Diversity, Equity & Inclusion Chair at ICLR 2024, 2025, and 2026. 🔨 Erin will advise on the **new submission track** initiative.

Stephanie Fu (UC Berkeley, USA) is a PhD student at UC Berkeley, where she is advised by Trevor Darrell (BAIR). Her research is primarily in computer vision, with an emphasis on modeling/understanding human visual intelligence and using that knowledge to build better perceptual representations. Stephanie holds a BS in Computer Science and Engineering, BS in Music, and MEng in Computer Science from MIT. 🧰 Stephanie was the cofounder of TEDxMIT, launching the inaugural event in 2019 and continuing to organize until 2023. She also served on MIT's EECS Committee on Diversity, Equity, and Inclusion (CDEI), was an organizer for HackMIT, and served as Finance Chair for Battlecode. 🔨 Stephanie will support the **established paper track** at the 2026 workshop.

Kushin Mukherjee (Stanford University, USA) is a postdoctoral scholar in the Cognitive Tools Lab at Stanford University broadly interested in the human ability to use and understand **visualizations** (charts, graphs, drawings) in service of communication and discovery. His research focuses on developing computational cognitive models of visualization understanding to both (1) better characterize human cognition and (2) bridge the gap between modern AI systems and human-like understanding of visual concepts. Kushin completed his PhD in Psychology at UW-Madison while based in the Knowledge and Concepts Lab and the Schloss Visual Reasoning Lab and also interned at Apple AI/ML working with the Vis team on chart understanding in vision-language models. 🧰 Kushin has co-organized the 2024 COGGRAPH and the 2022 Images2Symbols workshops at the Cognitive Science Society's annual meetings, and served on the program committee for the 2025 VIS×AI workshop at VIS and the 2021 SVRHM workshop at NeurIPS . He has served as a reviewer for ML/HCI venues including ICLR, EMNLP, NeurIPS, ACL, CHI, and VIS, and for journals in the cognitive sciences including *Cognition*, *Open Mind, Communications Biology, Nature Reviews Psychology,* and *Nature*. 🔨 Kushin will support the 2026 workshop **challenge**.

Ilia Sucholutsky (New York University, USA) is a faculty fellow / assistant professor at the NYU Center for Data Science and an incoming assistant professor of Computer Science at Purdue University. Previously, he was a postdoctoral fellow in Computer Science with Tom Griffiths at Princeton University and a visiting scholar in Brain & Cognitive Sciences at MIT. Ilia works on enabling deep learning with small data, with a focus on efficient representation learning. His recent focus has been on using information theory to study representational alignment. 🧰 Ilia co-organized the CogSci 2023 Workshop on LLMs for Cognitive Science, the Neuromonster 2023 Representational Alignment Session, the CHAI 2023 Human Cognition Session, the ICLR 2024 Re-Align Workshop, the ICML 2024 Cognition and LLMs Workshop, the NeuroMonster 2024 AI Session, the NeurIPS 2024 BehaviorML Workshop, and the ICLR 2025 Re-Align Workshop, and has previously served on the program committees and session committees of other ML workshops, including several at ICML and NeurIPS. Ilia also served as an area chair for the ICLR 2023 and 2024 Tiny Papers track, and is serving as an area chair for ICLR 2025. 🔨 Ilia will advise on the **established paper track** at the 2026 workshop.

Siddharth Suresh (University of Wisconsin, Madison, USA) is a PhD student in Cognitive Science at the University of Wisconsin-Madison. His research focuses on uncovering the differences between human semantic representations and those in neural network models. During his PhD, he has spent time at Amazon AGI, Netflix, and is currently working as a consultant with Amazon AGI-SF Labs. 🧰 Sid co-organized the 2025 Re-Align workshops in ICLR 2025 and CCN 2025. Sid has served as a reviewer for ICLR, ACL, and EMNLP. He was also a reviewer at the first iteration of the Re-Align workshop at ICLR 2024 and led hackathon development at the workshop at ICLR 2025. 🔨 Sid will lead the 2026 workshop **challenge**.

6.2 Program committee

Drawing from a list of 705 reviewer candidates (accumulated from community interest and the previous workshops we have organized), last year we recruited a program committee of 186 reviewers across the disciplines of machine learning, neuroscience and cognitive science. Our program committee wrote 209 reviews for 62 submissions, with all submissions receiving at least 2 reviews, and 85% of submissions receiving at least 3 reviews. This was an increase from the previous year when we had 103 reviewers and 80% of submissions receiving at least 3 reviews. Both years, the reviews and ensuing discussion were noted as "constructive and helpful" by authors, which we attribute to the broad topical expertise and significant excitement of reviewers serving on our program committee. We plan to re-invite last year's program committee, alongside authors of papers accepted to Re-Align last year, with a goal of establishing a program committee of 200 reviewers, to sustain growth.