Advances in Programming Languages and Neurosymbolism (AIPLANS)

Jialu Bao[3], **Maddy Bowers**[2], **Breandan Considine**[1,9],
Younesse Kaddar[5,10], **Justine Gehring**[9,11], **Shawn Tan**[4,9]
David Chiang[6], Kolya Malkin[7], Xujie Si[8,12]
McGill[1], MIT[2], Cornell[3], Université de Montréal[4], Oxford[5], Notre Dame[6],
University of Toronto[7], University of Edinburgh[8], Mila[9], Cohere[10], Moderne[11], Vector[12]

Tagline: AIPLANS: Fusing machine learning with programming language theory to create neurosymbolic programming machines! `https://aiplans.github.io`

What was once a fringe topic in machine learning has catalyzed a broader movement, with summer schools [7, 4], workshops [3, 1, 6] and journals [2] across the globe. Since the first AIPLANS workshop at NeurIPS 2021, considerable progress has been made on reuniting the branching diciplines of machine learning and programming language theory. From the first wave of symbolic AI emerged type systems, model checkers, and automated theorem provers. And from early connectionist techniques came the current generation of neural networks, loosely inspired by biological systems. By fusing the insights from neural and symbolic AI, neurosymbolism hopes to align the tinkering strategy of computational biology with the explainability and rigour of the symbolic method.

Today's language models have grown more capable and spawned a cottage industry of applications. But doubts about their safety and reliability still linger. Machine learning is generally concerned with naturalness, generalization and scalability, while programming languages care much more deeply about soundness, completeness and interpretability. Can we have the best of both worlds? In a series of invited talks, contributed papers, and panel discussions, we plan to explore the intersection of machine learning and programming languages, with a particular focus on the following areas:

- Probabilistic model checking
- Language induction
- Proof search and program synthesis
- Rewriting systems
- Analytic combinatorics
- Computational group theory
- Vector symbolic architectures
- Automata and formal language theory

- Formal aspects of language modeling [5]
- Constrained sampling from LLMs
- Randomized complexity
- Algorithmic information theory
- Circuit lower bounds
- Reachability problems
- Proof assistance and automation
- Type theory and semantics

Proposed Workshop Logistics

AIPLANS will be a one-day in-person workshop welcoming an economically and geographically diverse audience to participate. Talks will be hosted in English, following the standard format of oral presentations and panel discussions, to be concluded with a poster session. Proceedings will be non-archival. Outside of standard videoconferencing and SlidesLive assistance, we anticipate no other technical requirements. If accepted, we expect to receive a hundred or so participants, including speakers and workshop submitters, based on attendance at similarly-themed workshops in prior years. A partial list of speakers who have confirmed their presence (pending workshop acceptance) follows:

Parisa Kordjamshidi

The workshop itself will run for approximately eight hours during daylight saving time on five continents, and feature up to four 45-minute keynote talks and up to six 20-minute contributed talks. In addition, AIPLANS will solicit four-page paper submissions in a CFP to be circulated pending workshop acceptance. To encourage submissions from the broader ML/PL community, accepted authors will be given an opportunity to showcase their work in a virtual poster session or lightning talk for outstanding contributions. We expect to receive 20-30 submissions in total, and pledge that each paper will receive at least two fair and independent reviews. To minimize potential conflicts of interest, AIPLANS will manage submissions via the OpenReview conference management system.

Those who traditionally publish in venues such as SIGPLAN, SIGSOFT and other ACM venues are encouraged to submit work that may be considered relevant to the machine learning and reasoning community, provided that effort is taken to ensure its accessibility. Special consideration will be given to didactic submissions of outstanding clarity. Further information, including evaluation criteria, examples of relevant literature, deadlines and workshop logistics will be provided for reference.
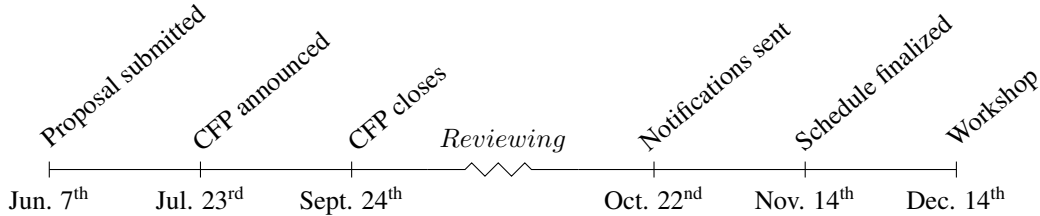

Figure 1: A tentative timeline for our proposed workshop at NeurIPS 2024.

If accepted, AIPLANS will announce its CFP and pursue contributions from the broader ML/PL community shortly thereafter. Two months later, the CFP will close on Sept. 17th. This deadline may be extended to no later than Sept. 24th, depending on the volume of submissions received, leaving sufficient time for referees and program chairs to leave feedback. Authors will be notified of acceptance no later than Oct. 22nd. We intend to finalize the schedule and coordinate presentation logistics between Nov. and Dec. 14th. Those who wish to prerecord talks will be given an opportunity to do so. The final workshop will consist of prerecorded and live talks with Q&A, followed by a moderated panel, and virtual poster session hosted on Gather.town. Further details about schedule and logistics will be made available, pending acceptance at: `https://aiplans.github.io`.

AIPLANS is an equal-opportunity workshop that celebrates cultural, linguistic, ethnic and intellectual diversity in all forms. Not only are we committed to nondiscrimination on the basis of, e.g., race, creed, age, gender, orientation, physical or mental handicap, but also aim to encourage individuals from other disadvantaged and underrepresented socioeconomic backgrounds to participate. Should our workshop be accepted, scholarships covering the cost of registration will be extended for those who wish to attend but would otherwise be unable to do so due to financial hardship. If needed, AIPLANS may pursue industry sponsorship for this initiative to enable a wider audience to attend. Further details about registration and funding availability will be provided in a timely manner.

Confirmed Workshop Organizers
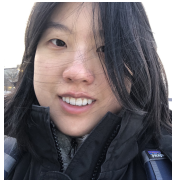
Jialu Bao is a Ph.D. student at Cornell advised by Prof. Justin Hsu, who is working on the verification of randomized algorithms. Before moving to Cornell with her advisor, she spent two years at University of Wisconsin – Madison as a Ph.D. student, and prior to that, did her undergrad at Cornell majoring in Mathematics and Computer Science.

⌂ https://baojia.lu 🐦 @howowhy

Maddy Bowers is a Ph.D. student at MIT, co-advised by Armando Solar-Lezama in EECS and Josh Tenenbaum in BCS, whose research combines methods from programming languages (PL) research with machine learning to tackle problems in artificial intelligence.

⌂ https://mlb2251.github.io/ 🐦 @mattlbowers

Breandan Considine is a Ph.D. student at McGill University co-supervised by Jin Guo and Xujie Si. His research studies how to reason about the behavior of real-world programs and build more intelligent programming tools for developers. Previously, he organized the first AIPLANS workshop at NeurIPS and co-organized the ICLR workshop, Rethinking ML Papers.

⌂ https://breandan.github.io 🐦 @breandan

Younesse Kaddar is a Ph.D. student in theoretical computer science at the University of Oxford, working on programming language semantics, Bayesian probabilistic programming, and category theory.

⌂ https://younesse.net/ 🐦 @you_kad

Justine Gehring is a research engineer at Moderne working in the field of Machine Learning (ML) for code and Graph Neural Networks (GNNs). Her focus lies in generating code under challenging circumstances, specifically in scenarios such as sparse data where library-specific code is required, as well as managing a substantial amount of code at a time. She completed her Master's Degree in Computer Science at Mila & McGill.

⌂ https://justine-gehring.github.io/ 🐦 @GehringJustine
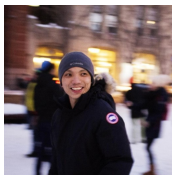
Shawn Tan is a Ph.D. candidate at Mila, Université de Montréal. He is interested in differentiable methods for structured prediction, specifically in the domain of natural language. He co-authored the Ordered Neurons paper which won best paper at ICLR 2019.

⌂ http://blog.wtf.sg 🐦 @tanshawn

Confirmed Program Committee

David Chiang is a Professor at Notre Dame University. His research is in natural language processing, the subfield of computer science that aims to enable computers to understand and produce human language. He focuses mainly on language translation, and has interests in syntactic parsing and other areas as well.

⌂ https://www3.nd.edu/~dchiang/ 🐦 @davidweichiang
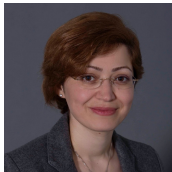
Parisa Kordjamshidi is an assistant professor at Michigan State University and director of the Heterogeneous Learning and Reasoning (HLR) lab. Her main research interests are artificial intelligence, machine learning, natural language processing, and declarative learning based programming (DeLBP). She works on the extraction of formal semantics and structured representations from natural language, with a specific focus on spatial semantics representation and structured output learning models.

⌂ https://www.cse.msu.edu/~kordjams/

Kolya Malkin is an Associate Professor at the University of Edinburgh working on algorithms for deep-learning-based reasoning and their applications, in particular induction of compositional structure in generative models, modeling of posteriors over high-dimensional explanatory variables, uncertainty-aware explanations for observed data, human-like symbolic, formal, and mathematical reasoning and tracking land use patterns over time and monitoring the effects of climate change.

⌂ https://malkin1729.github.io/
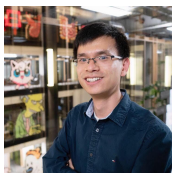
Xujie Si is an Assistant Professor and Canada CIFAR AI Chair in the School of Computer Science at the University of Toronto and Vector Institute. He finished his Ph.D. in Computer and Information Science at the University of Pennsylvania in 2020, advised by Prof. Mayur Naik. Xujie received his M.S. in computer science from Vanderbilt University in 2014, before which he obtained his B.E. (with Honors) from Nankai Unversity in 2011. He spent the summer of 2019 as a research scientist intern at DeepMind working with Pushmeet Kohli in the Robust AI team.

⌂ https://www.cs.mcgill.ca/~xsi 🐦 @XujieSi

References

[1] Vaishak Belle, Martin Grohe, and Efi Tsamoura. Samsung ai neurosymbolic workshop. 2022.

[2] Tarek Besold, Artur d'Avila Garcez, and Pascal Hitzler. Neurosymbolic artificial intelligence. 2023.

[3] Tarek Besold et al. Neurosymbolic ai: From hybrid models to hybrid intelligence. 2023.

[4] Omar Costilla Reyes et al. Summer school on neurosymbolic programming. 2024.

[5] Ryan Cotterell, Anej Svete, Clara Meister, Tianyu Liu, and Li Du. Formal aspects of language modeling. *arXiv preprint arXiv:2311.04329*, 2023.

[6] Filip Llievski, Jacopo de Berardinis, Jongmo Kim, and Nitisha Jain. First international workshop on generative neuro-symbolic ai. 2024.

[7] Asim Munawar, Alexander Gray, et al. Neuro-symbolic ai summer school. 2023.