Physics Supernova: AI Agent Matches Elite Gold Medalists at IPhO 2025

Jiahao Qiu*[1], Jingzhe Shi*[2], Xinzhe Juan[3,4], Zelin Zhao[5],
Jiayi Geng[1], Shilong Liu[1], Hongru Wang[1], Sanfeng Wu[1], Mengdi Wang[1]
[1]Princeton University [2]Tsinghua University [3]Shanghai Jiao Tong University
[4]University of Michigan [5]King's College London
* equal contribution

Abstract

Physics provides fundamental laws that describe and predict the natural world. AI systems aspiring toward more general, real-world intelligence must therefore demonstrate strong physics problem solving abilities: to formulate and apply physical laws for explaining and predicting physical processes. The International Physics Olympiad (IPhO)–the world's most prestigious physics competition–offers a rigorous benchmark for this purpose. We introduce Physics Supernova, an AI agent system with superior physics problem-solving abilities that match elite IPhO gold medalists. On the IPhO 2025 theory problems, Physics Supernova attains 23.5/30 points, ranking 14$^{\text{th}}$ of 406 contestants and surpassing the median performance of human gold medalists. We extensively analyzed Physics Supernova's capabilities and flexibility across diverse physics tasks. These results show that principled tool integration within agent systems can deliver competitive improvements in solving challenging science problems.[1]

1 Introduction

The supreme task of the physicist is to arrive at those universal elementary laws from which the cosmos can be built up by pure deduction.

Albert Einstein

Physics aims to compactly formulate the fundamental laws that govern the behavior of the universe [1, 2]. The mastery of physics entails constructing rigorous abstractions [3, 4]. As AI systems increasingly integrate into the physical world and advance toward Artificial General Intelligence (AGI) and potentially Artificial Super intelligence (ASI), a deep grounding in physics becomes a critical foundation for developing competent and reliable intelligence [5, 6, 7]. In this work, we investigate how to enhance AI systems' physics problem-solving capabilities with agent-based architectures. To assess progress in this challenging domain, we benchmark our models on the theory problems of the 2025 International Physics Olympiad (IPhO [8]) , a globally recognized competition that emphasizes deep conceptual understanding, abstraction, and advanced problem-solving in physics.

The International Physics Olympiad, or IPhO [8], is widely viewed as a prestigious physics competition. The 2025 International Physics Olympiad (IPhO) was held in July in France [9]. Unlike earlier benchmarks evaluating physics problem-solving abilities [10, 11, 12] that might risk data contamination, use coarse evaluation (e.g., final-answer-only scoring), offer limited novelty, and lack assessments of precise figure reading/measurement, etc., IPhO 2025 Theory Problems exhibit

[1]The codes are available at https://github.com/CharlesQ9/Physics-Supernova.

39th Conference on Neural Information Processing Systems (NeurIPS 2025) Workshop: Evaluating the Evolving LLM Lifecycle: Benchmarks, Emergent Abilities, and Scaling.

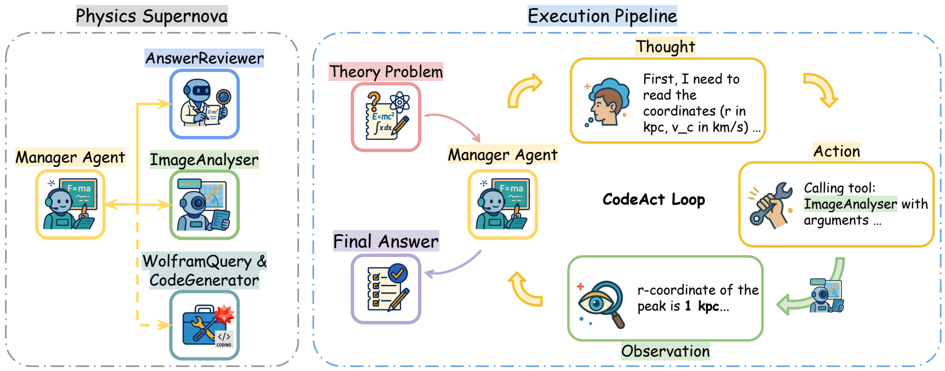


Figure 1: Our proposed agent system: Physics Supernova, for solving theory problems in physics.

several distinctive features: (1) new release date in July 2025 [9]; (2) fine-grained, part-level scoring, enabling detailed assessment; (3) uncommon physics models that challenge standard approaches; and (4) explicit figure-based measurement requirements (see Table 1). In this work, we utilize **IPhO 2025 Theory Problems** as the benchmark for evaluating AI's capability in physics. IPhO contestants are provided with calculators and fundamental constant tables; moreover, expert physicists routinely work with extra tools: these facts highlight the importance of external tool use for physicists in general. However, previous benchmarks mainly evaluate the performance of base LLMs, with limited integration of tools and limited evaluation of LLM-based agent systems [12, 13, 14].

LLM-based agent systems demonstrate significant advantages over LLM alone in planning, complex reasoning and tool using [15, 16]. ReAct [17] introduces the idea of Reasoning-Acting loop for agents. More recently, self-evolving agents have been proposed, with even fewer designs based on humans [18, 19, 20]. However, previous work mainly examines general-purpose agents with a focus on possible virtual-assisting style tasks (e.g., GAIA [21]) or other domains like math [22], history [23], etc., not covering the domain of complex physics problems.

In this work, with a focus on physics reasoning and problem-solving ability, we introduce **Physics Supernova**, an agent system equipped with physics-oriented tools targeting physics problem-solving. We equip LLMs with tools to improve physical reasoning ability: the Image Analyzer Tool to accurately extract data, and the Answer Reviewer Tool to self-review and refine. With these tools, Physics Supernova achieves **Gold Medal** in IPhO 2025 Theory Problems: it ranks top 10% on all three problems tested, and ranks 14^{th} among 406 human contestants in IPhO 2025 Theory Problems.

To conclude, we summarize our main contribution as follows: **(1)** We develop Physics Supernova, an agent system combining Large Language Models with physics-specific tools, exhibiting strong capabilities across a wide range of physics tasks. **(2)** We show that Physics Supernova achieves gold-medalist-level performance on IPhO 2025 Theory Problems, ranking 14^{th} among all 406 human contestants worldwide, exceeding the official gold medalist median score. **(3)** We conduct further analysis and experiments to show Physics Supernova's capability and flexibility for physics problem solving tasks.

2 Method

2.1 Agent Architecture

We introduce **Physics Supernova**, an AI agent system designed to solve complex physics theory problems. It follows the `CodeAgent` architecture from the `smolagents` framework [24]. As illustrated in Figure 1, the system consists of a **Manager Agent** $\mathcal{M}$ and a set of domain-specific tools $\mathcal{D} = \{d_i\}_{i=1}^{n}$, where each d_i is physics-problem-oriented tool.

Unlike prior work in mathematical problem-solving that often relies on fixed, manually hard-coded workflows [25, 26], our approach emphasizes **flexible self-planning**. The Manager Agent is granted access to the tool set $\mathcal{D}$ but is not provided with a hard-coded predefined execution graph or action script. The agent is able to call different Tools according to its progress made in its process of solving

Problem	Subpart	Subpart Content {# Scoring Points†}	Model Novelty	Computation Difficulty	Image Ability	Top 10% Score
Theory Problem 1 (Modern Physics)	Part A	Bohr Model {18}	Common	Low	No Image	9.0/10.0
	Part B	Rotation of a Galaxy {18}	Rare	Med	Accurate Meas	
	Part C	Mass Distribution of a Galaxy {24}	Rare	Med	Accurate Meas	
	Part D	Tully-Fisher Relation & MOND {23}	Novel	Med	Rough Meas	
Theory Problem 2 (Thermodynamics & Dynamics)	Part A	Pulling on a Submerged Tube {9}	Rare	Med	Understand	5.0/10.0
	Part B	Two-Part Barometric Tube {10}	Rare	High	Understand	
	Part C	Cox's Timepiece {38}	Novel	High	Understand	
Theory Problem 3 (Thermodynamics & Dynamics)	Part A	Nucleation & Growth of Bubbles {26}	Rare	Med	Rough Meas	8.0/10.0
	Part B	Acoustic Emission of Bubbles {20}	Rare	Med	Understand	
	Part C	Popping Champagne {19}	Rare	Med	Understand	

Table 1: Theory problems of IPhO 2025: Subpart contents (and number of scoring points† for each subpart, obtained from official solutions at `https://ipho.olimpicos.net/`), physics model novelty (among Physics Olympaids), computation difficulty levels, required image-reading skills, and top 10% human scores. For physics model novelty: Common, Rare and Novel represent how often such physical model appears in previous Physics Olympiads. For computation Difficulty: Low , Med and High show the requirement for tedious computation and careful attention for applying formulas. For Image reading requirements: No Image and Understand do not require measurements, while Rough Meas and Accurate Meas requires rough/accurate reading from the image. Examples of these problems are shown in Appendix A.

problems. Overall, the agent operates in an iterative **Reason–Act** loop. In each round, it generates a natural language-based reasoning step and justify the selection of tools $d \subseteq \mathcal{D}$ (*Reason*); it then produces code for calling tools, which are then called and produce intermediate observations (*Act*). These observations are then coorporated into next reasoning step, enabling refinement of agent's understanding.The loop continues until final answers are produced for all questions.

2.2 Physics-Problem-Oriented Toolset

We develop agent tools based on observations and experience on physics problems [9, 10, 12] and LLM agents [19, 23, 27]. Here, we discuss the physics-problems-oriented tools we equip the manager agent with: the *ImageAnalyzer* and the *AnswerReviewer*.[2]

ImageAnalyzer: Reading experimental results and extracting critical information from figures is important for expert physicists. To enhance image handling, ***ImageAnalyzer*** routes a high-res image to a dedicated Vision Language Model, addressing precise tasks such as reading numeric values and making measurements.

AnswerReviewer: Physicists routinely evaluate whether their theoretical results are physically meaningful. To enhance AI's ability of rethinking, ***AnswerReviewer*** is provided: it classifies likely error types and locates erroneous expressions through the process. We provide ablations studies where the reviewing tool improves performance in Section C.1.

With only ***ImageAnalyzer*** and ***AnswerReviewer***, Physics Supernova empowered by state-of-the-art LLM (Gemini 2.5 Pro [28]) can achieve medium gold-level performance in IPhO 2025 Theory Problems (Section 3). Moreover, this system supports the integration of additional advanced physics-related tools, such as the WolframAlpha question-answering engine (Section C.3) that can assist with computationally intensive physics tasks.

3 Experiment: Physics Supernova Excels in IPhO 2025 Theory Problems

3.1 Experiment Setup

Benchmarking Dataset IPhO 2025 has 3 theory problems and 2 experimental problems in which each problem counts for 10.0 points, adding up to 50.0 points in total. Among 406 contestants from more than 90 countries and regions, 37 ones with the highest total scores are awarded the Gold Medal. Theory score is the sum of all three theory problems. The minimum and median theory scores for gold medalists are 19.4 & 22.8. Table 1 shows the detailed contents of each theory problem.

[2]The tools and prompts are not specially tuned on IPhO 2025 problems to avoid data leakage. We run experiments on IPhO 2025 theory problems and report scores in Section 3 after creating these tools.

Problem	Part A	Part B	Part C	Part D	SUM
Theory 1 total score	2.2	2.5	3.0	2.3	10.0
LLM Only	2.20 ± 0.00	2.28 ± 0.04	2.06 ± 0.13	1.92 ± 0.11	8.46 ± 0.16
Physics Supernova	2.20 ± 0.00	2.20 ± 0.00	2.46 ± 0.05	2.16 ± 0.05	$\mathbf{9.02 \pm 0.11}$
Human Top 10%	/	/	/	/	~ 9.0
Theory 2 total score	1.3	2.0	6.7	/	10.0
LLM Only	1.16 ± 0.31	1.08 ± 0.19	3.04 ± 1.02	/	5.30 ± 0.99
Physics Supernova	1.30 ± 0.00	1.16 ± 0.22	3.62 ± 0.79	/	$\mathbf{6.08 \pm 0.77}$
Human Top 10%	/	/	/	/	~ 5.0
Theory 3 total score	4.3	3.3	2.4	/	10.0
LLM Only	3.70 ± 0.07	2.74 ± 0.36	1.22 ± 0.18	/	7.66 ± 0.51
Physics Supernova	4.02 ± 0.22	3.26 ± 0.09	1.12 ± 0.11	/	$\mathbf{8.40 \pm 0.27}$
Human Top 10%	/	/	/	/	~ 8.0
Theory Part total score	/	/	/	/	30.0
LLM Only	/	/	/	/	21.4 ± 1.1
Physics Supernova	/	/	/	/	$\mathbf{23.5 \pm 0.8}$
Gold Medalists (mediam)	/	/	/	/	22.8

Table 2: Experiment results for Physics Supernova (mean$\pm$std) across multiple problems and parts (with Gemini 2.5 Pro), for 5 rounds. Our agent results rank top 10% among human contestants on all three Theory problems.

Experiment Details We test Physics Supernova on each of the three theory problems, with chosen pre-defined tools. As smolagents [24] lack a built-in summarization memory, we implemented a lightweight summarizing tool to summarize existing progress, which is provided beside the ***Image-Analyzer*** and ***AnswerReviewer*** tools. [3]

3.2 Main Experimental Results

The results of our main experiments are shown in Table 2. The mean theory score of Gemini 2.5 Pro alone ranks 30$^{\text{th}}$ among 406 contestants, while the mean theory score of **Physics Supernova ranked 14$^{\text{th}}$ among all 406 contestants, surpassing the medium theory score of the gold medalists**. Noticeable, Physics Supernova ranks top 10% on all three Theory problems tested. Comparing Table 1 and Table 2, we make the following observations. **(1)** Physics Supernova successfully improves LLM performance on all three problems. **(2)** Larger advantages are achieved on more difficult problems (i.e., larger improvement is achieved on Problems 2 and 3). **(3)** Harder problems show larger variance (i.e. larger variance on Problem 2). We provide more analysis and results of additional tools targeting more general physics problems in Appendix C.

4 Conclusion

In this work, we introduce Physics Supernova, a flexible agent system for solving Olympiad-level physics problems and beyond. By providing task-specific tools including ImageAnalyzer, AnswerReviewer, etc., we extend the capability of state-of-the-art LLMs on physics problems. This aspect was previously viewed mostly as benchmarks of base, simple LLMs [10, 11, 12]. On the newly released IPhO 2025 theory problems, our agent system achieves gold-medal-level performance. Specifically, our method ranks top 10% among human contestants in all three theory problems. It achieves 23.5 points on the theory problems, ranks #14 among all 406 human contestants, and surpasses the median theory score of the gold medalists.

Overall, our proposed Physics Supernova successfully improves the physics problem solving ability of LLMs through agent paradigms. This shows the potential of agent systems to improve the capability of LLMs for scientific reasoning and physics-related tasks, and further implies their potential for developing super intelligence that embeds into the real world.

[3]Throughout the experiments, we use Gemini 2.5 Pro [28] for the agent system and all the LLM-requiring tools. For each problem, human experts score LLM answers in detail based on the official IPhO 2025 scoring criteria (we provide examples of scoring criteria in Appendix A)[4]. For each problem, the experiment is carried out 5 times, where mean and standard deviation are reported in Table2.

References

[1] Richard P. Feynman. *The Character of Physical Law*. MIT Press, 1965.

[2] P. J. E. Peebles. *Principles of Physical Cosmology*. Princeton University Press, 1993.

[3] Emmy Noether. Invariante variationsprobleme. *Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse*, pages 235–257, 1918.

[4] Karl J. Åström and Richard M. Murray. *Feedback Systems: An Introduction for Scientists and Engineers*. Princeton University Press, 2008.

[5] Edward A. Lee. Cyber physical systems: Design challenges. In *Proc. IEEE Int'l Symposium on Object/Component/Service-Oriented Real-Time Distributed Computing (ISORC)*, pages 363–369, 2008.

[6] Nick Bostrom. *Superintelligence: Paths, Dangers, Strategies*. Oxford University Press, 2014.

[7] Yann LeCun. A path towards autonomous machine intelligence. OpenReview position paper, 2022.

[8] International Physics Olympiad (IPhO). Official Website: `https://www.ipho-new.org/`, ongoing. Accessed: 2025-07-29.

[9] International Physics Olympiad (IPhO) 2025. Official Website: `https://www.ipho2025.fr/`, ongoing. Accessed: 2025-07-29.

[10] Kun Xiang, Heng Li, Terry Jingchen Zhang, Yinya Huang, Zirong Liu, Peixin Qu, Jixi He, Jiaqi Chen, Yu-Jie Yuan, Jianhua Han, Hang Xu, Hanhui Li, Mrinmaya Sachan, and Xiaodan Liang. Seephys: Does seeing help thinking? – benchmarking vision-based physics reasoning, 2025.

[11] Chaoqun He, Renjie Luo, Yuzhuo Bai, Shengding Hu, Zhen Leng Thai, Junhao Shen, Jinyi Hu, Xu Han, Yujie Huang, Yuxiang Zhang, Jie Liu, Lei Qi, Zhiyuan Liu, and Maosong Sun. Olympiadbench: A challenging benchmark for promoting agi with olympiad-level bilingual multimodal scientific problems, 2024.

[12] Shi Qiu, Shaoyang Guo, Zhuo-Yang Song, Yunbo Sun, Zeyu Cai, Jiashen Wei, Tianyu Luo, Yixuan Yin, Haoxu Zhang, Yi Hu, Chenyang Wang, Chencheng Tang, Haoling Chang, Qi Liu, Ziheng Zhou, Tianyu Zhang, Jingtian Zhang, Zhangyi Liu, Minghao Li, Yuku Zhang, Boxuan Jing, Xianqi Yin, Yutong Ren, Zizhuo Fu, Jiaming Ji, Weike Wang, Xudong Tian, Anqi Lv, Laifu Man, Jianxiang Li, Feiyu Tao, Qihua Sun, Zhou Liang, Yushu Mu, Zhongxuan Li, Jing-Jun Zhang, Shutao Zhang, Xiaotian Li, Xingqi Xia, Jiawei Lin, Zheyu Shen, Jiahang Chen, Qiuhao Xiong, Binran Wang, Fengyuan Wang, Ziyang Ni, Bohan Zhang, Fan Cui, Changkun Shao, Qing-Hong Cao, Ming xing Luo, Yaodong Yang, Muhan Zhang, and Hua Xing Zhu. Phybench: Holistic evaluation of physical perception and reasoning in large language models, 2025.

[13] Charlie Snell, Jaehoon Lee, Kelvin Xu, and Aviral Kumar. Scaling llm test-time compute optimally can be more effective than scaling model parameters, 2024.

[14] Yangzhen Wu, Zhiqing Sun, Shanda Li, Sean Welleck, and Yiming Yang. Scaling inference computation: Compute-optimal inference for problem-solving with language models. In *The 4th Workshop on Mathematical Reasoning and AI at NeurIPS'24*, 2024.

[15] Hongru Wang, Lingzhi Wang, Yiming Du, Liang Chen, Jingyan Zhou, Yufei Wang, and Kam-Fai Wong. A survey of the evolution of language model-based dialogue systems: Data, task and models, 2025.

[16] Takeshi Kojima, Shixiang Shane Gu, Machel Reid, Yutaka Matsuo, and Yusuke Iwasawa. Large language models are zero-shot reasoners, 2023.

[17] Shunyu Yao, Jeffrey Zhao, Dian Yu, Nan Du, Izhak Shafran, Karthik R Narasimhan, and Yuan Cao. React: Synergizing reasoning and acting in language models. In *The Eleventh International Conference on Learning Representations*, 2023.

[18] Huan ang Gao, Jiayi Geng, Wenyue Hua, Mengkang Hu, Xinzhe Juan, Hongzhang Liu, Shilong Liu, Jiahao Qiu, Xuan Qi, Yiran Wu, Hongru Wang, Han Xiao, Yuhang Zhou, Shaokun Zhang, Jiayi Zhang, Jinyu Xiang, Yixiong Fang, Qiwen Zhao, Dongrui Liu, Qihan Ren, Cheng Qian, Zhenhailong Wang, Minda Hu, Huazheng Wang, Qingyun Wu, Heng Ji, and Mengdi Wang. A survey of self-evolving agents: On path to artificial super intelligence, 2025.

[19] Jiahao Qiu, Xuan Qi, Tongcheng Zhang, Xinzhe Juan, Jiacheng Guo, Yifu Lu, Yimin Wang, Zixin Yao, Qihan Ren, Xun Jiang, Xing Zhou, Dongrui Liu, Ling Yang, Yue Wu, Kaixuan Huang, Shilong Liu, Hongru Wang, and Mengdi Wang. Alita: Generalist agent enabling scalable agentic reasoning with minimal predefinition and maximal self-evolution, 2025.

[20] Lang Feng, Zhenghai Xue, Tingcong Liu, and Bo An. Group-in-group policy optimization for llm agent training, 2025.

[21] Grégoire Mialon, Clémentine Fourrier, Thomas Wolf, Yann LeCun, and Thomas Scialom. Gaia: a benchmark for general ai assistants. In *The Twelfth International Conference on Learning Representations*, 2023.

[22] Thang Luong and Edward Lockhart. Advanced version of gemini with deep think officially achieves gold-medal standard at the international mathematical olympiad. Google DeepMind Blog, July 2025.

[23] Jiahao Qiu, Fulian Xiao, Yimin Wang, Yuchen Mao, Yijia Chen, Xinzhe Juan, Shu Zhang, Siran Wang, Xuan Qi, Tongcheng Zhang, Zixin Yao, Jiacheng Guo, Yifu Lu, Charles Argon, Jundi Cui, Daixin Chen, Junran Zhou, Shuyao Zhou, Zhanpeng Zhou, Ling Yang, Shilong Liu, Hongru Wang, Kaixuan Huang, Xun Jiang, Yuming Cao, Yue Chen, Yunfei Chen, Zhengyi Chen, Ruowei Dai, Mengqiu Deng, Jiye Fu, Yunting Gu, Zijie Guan, Zirui Huang, Xiaoyan Ji, Yumeng Jiang, Delong Kong, Haolong Li, Jiaqi Li, Ruipeng Li, Tianze Li, Zhuoran Li, Haixia Lian, Mengyue Lin, Xudong Liu, Jiayi Lu, Jinghan Lu, Wanyu Luo, Ziyue Luo, Zihao Pu, Zhi Qiao, Ruihuan Ren, Liang Wan, Ruixiang Wang, Tianhui Wang, Yang Wang, Zeyu Wang, Zihua Wang, Yujia Wu, Zhaoyi Wu, Hao Xin, Weiao Xing, Ruojun Xiong, Weijie Xu, Yao Shu, Yao Xiao, Xiaorui Yang, Yuchen Yang, Nan Yi, Jiadong Yu, Yangyuxuan Yu, Huiting Zeng, Danni Zhang, Yunjie Zhang, Zhaoyu Zhang, Zhiheng Zhang, Xiaofeng Zheng, Peirong Zhou, Linyan Zhong, Xiaoyin Zong, Ying Zhao, Zhenxin Chen, Lin Ding, Xiaoyu Gao, Bingbing Gong, Yichao Li, Yang Liao, Guang Ma, Tianyuan Ma, Xinrui Sun, Tianyi Wang, Han Xia, Ruobing Xian, Gen Ye, Tengfei Yu, Wentao Zhang, Yuxi Wang, Xi Gao, and Mengdi Wang. On path to multimodal historical reasoning: Histbench and histagent, 2025.

[24] Aymeric Roucher, Albert Villanova del Moral, Thomas Wolf, Leandro von Werra, and Erik Kaunismäki. 'smolagents': a smol library to build great agentic systems. `https://github.com/huggingface/smolagents`, 2025.

[25] Yichen Huang and Lin F. Yang. Gemini 2.5 pro capable of winning gold at imo 2025, 2025.

[26] Haohan Lin, Zhiqing Sun, Sean Welleck, and Yiming Yang. Lean-star: Learning to interleave thinking and proving, 2025.

[27] Matthew Renze and Erhan Guven. The benefits of a concise chain of thought on problem-solving in large language models. In *2024 2nd International Conference on Foundation and Large Language Models (FLLM)*, page 476483. IEEE, November 2024.

[28] Gheorghe Comanici, Eric Bieber, Mike Schaekermann, Ice Pasupat, Noveen Sachdeva, Inderjit Dhillon, Marcel Blistein, Ori Ram, Dan Zhang, Evan Rosen, Luke Marris, Sam Petulla, Colin Gaffney, Asaf Aharoni, Nathan Lintz, Tiago Cardal Pais, Henrik Jacobsson, Idan Szpektor, Nan-Jiang Jiang, Krishna Haridasan, Ahmed Omran, Nikunj Saunshi, Dara Bahri, Gaurav Mishra, Eric Chu, Toby Boyd, Brad Hekman, Aaron Parisi, Chaoyi Zhang, Kornraphop Kawintiranon, Tania Bedrax-Weiss, Oliver Wang, Ya Xu, Ollie Purkiss, Uri Mendlovic, Ilaï Deutel, Nam Nguyen, Adam Langley, Flip Korn, Lucia Rossazza, Alexandre Ramé, Sagar Waghmare, Helen Miller, Nathan Byrd, Ashrith Sheshan, Raia Hadsell Sangnie Bhardwaj, Pawel Janus, Tero Rissa, Dan Horgan, Sharon Silver, Ayzaan Wahid, Sergey Brin, Yves Raimond, Klemen Kloboves, Cindy Wang, Nitesh Bharadwaj Gundavarapu, Ilia Shumailov, Bo Wang, Mantas Pajarskas, Joe Heyward, Martin Nikoltchev, Maciej Kula, Hao Zhou, Zachary Garrett, Sushant Kafle, Sercan Arik, Ankita Goel, Mingyao Yang, Jiho Park, Koji Kojima, Parsa Mahmoudieh, Koray Kavukcuoglu, Grace Chen, Doug Fritz, Anton Bulyenov, Sudeshna Roy, Dimitris Paparas, Hadar Shemtov, Bo-Juen Chen, Robin Strudel, David Reitter, Aurko Roy, Andrey Vlasov, Changwan Ryu, Chas Leichner, Haichuan Yang, Zelda Mariet, Denis Vnukov, Tim Sohn, Amy Stuart, Wei Liang, Minmin Chen, Praynaa Rawlani, Christy Koh, JD Co-Reyes, Guangda Lai, Praseem Banzal, Dimitrios Vytiniotis, Jieru Mei, Mu Cai, Mohammed Badawi, Corey Fry, Ale Hartman, Daniel Zheng, Eric Jia, James Keeling, Annie Louis, Ying Chen, Efren Robles, Wei-Chih Hung, Howard Zhou, Nikita Saxena, Sonam Goenka, Olivia Ma, Zach Fisher, Mor Hazan Taege, Emily Graves, David Steiner, Yujia Li, Sarah Nguyen, Rahul Sukthankar, Joe

Stanton, Ali Eslami, Gloria Shen, Berkin Akin, Alexey Guseynov, Yiqian Zhou, Jean-Baptiste Alayrac, Armand Joulin, Efrat Farkash, Ashish Thapliyal, Stephen Roller, Noam Shazeer, Todor Davchev, Terry Koo, Hannah Forbes-Pollard, Kartik Audhkhasi, Greg Farquhar, Adi Mayrav Gilady, Maggie Song, John Aslanides, Piermaria Mendolicchio, Alicia Parrish, John Blitzer, Pramod Gupta, Xiaoen Ju, Xiaochen Yang, Puranjay Datta, Andrea Tacchetti, Sanket Vaibhav Mehta, Gregory Dibb, Shubham Gupta, Federico Piccinini, Raia Hadsell, Sujee Rajayogam, Jiepu Jiang, Patrick Griffin, Patrik Sundberg, Jamie Hayes, Alexey Frolov, Tian Xie, Adam Zhang, Kingshuk Dasgupta, Uday Kalra, Lior Shani, Klaus Macherey, Tzu-Kuo Huang, Liam MacDermed, Karthik Duddu, Paulo Zacchello, Zi Yang, Jessica Lo, Kai Hui, Matej Kastelic, Derek Gasaway, Qijun Tan, Summer Yue, Pablo Barrio, John Wieting, Weel Yang, Andrew Nystrom, Solomon Demmessie, Anselm Levskaya, Fabio Viola, Chetan Tekur, Greg Billock, George Necula, Mandar Joshi, Rylan Schaeffer, Swachhand Lokhande, Christina Sorokin, Pradeep Shenoy, Mia Chen, Mark Collier, Hongji Li, Taylor Bos, Nevan Wichers, Sun Jae Lee, Angéline Pouget, Santhosh Thangaraj, Kyriakos Axiotis, Phil Crone, Rachel Sterneck, Nikolai Chinaev, Victoria Krakovna, Oleksandr Ferludin, Ian Gemp, Stephanie Winkler, Dan Goldberg, Ivan Korotkov, Kefan Xiao, Malika Mehrotra, Sandeep Mariserla, Vihari Piratla, Terry Thurk, Khiem Pham, Hongxu Ma, Alexandre Senges, Ravi Kumar, Clemens Meyer, Ellie Talius, Nuo Wang Pierse, Ballie Sandhu, Horia Toma, Kuo Lin, Swaroop Nath, Tom Stone, Dorsa Sadigh, Nikita Gupta, Arthur Guez, Avi Singh, Matt Thomas, Tom Duerig, Yuan Gong, Richard Tanburn, Lydia Lihui Zhang, Phuong Dao, Mohamed Hammad, Sirui Xie, Shruti Rijhwani, Ben Murdoch, Duhyeon Kim, Will Thompson, Heng-Tze Cheng, Daniel Sohn, Pablo Sprechmann, Qiantong Xu, Srinivas Tadepalli, Peter Young, Ye Zhang, Hansa Srinivasan, Miranda Aperghis, Aditya Ayyar, Hen Fitoussi, Ryan Burnell, David Madras, Mike Dusenberry, Xi Xiong, Tayo Oguntebi, Ben Albrecht, Jörg Bornschein, Jovana Mitrovi, Mason Dimarco, Bhargav Kanagal Shamanna, Premal Shah, Eren Sezener, Shyam Upadhyay, Dave Lacey, Craig Schiff, Sebastien Baur, Sanjay Ganapathy, Eva Schnider, Mateo Wirth, Connor Schenck, Andrey Simanovsky, Yi-Xuan Tan, Philipp Fränken, Dennis Duan, Bharath Mankalale, Nikhil Dhawan, Kevin Sequeira, Zichuan Wei, Shivanker Goel, Caglar Unlu, Yukun Zhu, Haitian Sun, Ananth Balashankar, Kurt Shuster, Megh Umekar, Mahmoud Alnahlawi, Aäron van den Oord, Kelly Chen, Yuexiang Zhai, Zihang Dai, Kuang-Huei Lee, Eric Doi, Lukas Zilka, Rohith Vallu, Disha Shrivastava, Jason Lee, Hisham Husain, Honglei Zhuang, Vincent Cohen-Addad, Jarred Barber, James Atwood, Adam Sadovsky, Quentin Wellens, Steven Hand, Arunkumar Rajendran, Aybuke Turker, CJ Carey, Yuanzhong Xu, Hagen Soltau, Zefei Li, Xinying Song, Conglong Li, Iurii Kemaev, Sasha Brown, Andrea Burns, Viorica Patraucean, Piotr Stanczyk, Renga Aravamudhan, Mathieu Blondel, Hila Noga, Lorenzo Blanco, Will Song, Michael Isard, Mandar Sharma, Reid Hayes, Dalia El Badawy, Avery Lamp, Itay Laish, Olga Kozlova, Kelvin Chan, Sahil Singla, Srinivas Sunkara, Mayank Upadhyay, Chang Liu, Aijun Bai, Jarek Wilkiewicz, Martin Zlocha, Jeremiah Liu, Zhuowan Li, Haiguang Li, Omer Barak, Ganna Raboshchuk, Jiho Choi, Fangyu Liu, Erik Jue, Mohit Sharma, Andreea Marzoca, Robert Busa-Fekete, Anna Korsun, Andre Elisseeff, Zhe Shen, Sara Mc Carthy, Kay Lamerigts, Anahita Hosseini, Hanzhao Lin, Charlie Chen, Fan Yang, Kushal Chauhan, Mark Omernick, Dawei Jia, Karina Zainullina, Demis Hassabis, Danny Vainstein, Ehsan Amid, Xiang Zhou, Ronny Votel, Eszter Vértes, Xinjian Li, Zongwei Zhou, Angeliki Lazaridou, Brendan McMahan, Arjun Narayanan, Hubert Soyer, Sujoy Basu, Kayi Lee, Bryan Perozzi, Qin Cao, Leonard Berrada, Rahul Arya, Ke Chen, Katrina, Xu, Matthias Lochbrunner, Alex Hofer, Sahand Sharifzadeh, Renjie Wu, Sally Goldman, Pranjal Awasthi, Xuezhi Wang, Yan Wu, Claire Sha, Biao Zhang, Maciej Mikua, Filippo Graziano, Siobhan Mcloughlin, Irene Giannoumis, Youhei Namiki, Chase Malik, Carey Radebaugh, Jamie Hall, Ramiro Leal-Cavazos, Jianmin Chen, Vikas Sindhwani, David Kao, David Greene, Jordan Griffith, Chris Welty, Ceslee Montgomery, Toshihiro Yoshino, Liangzhe Yuan, Noah Goodman, Assaf Hurwitz Michaely, Kevin Lee, KP Sawhney, Wei Chen, Zheng Zheng, Megan Shum, Nikolay Savinov, Etienne Pot, Alex Pak, Morteza Zadimoghaddam, Sijal Bhatnagar, Yoad Lewenberg, Blair Kutzman, Ji Liu, Lesley Katzen, Jeremy Selier, Josip Djolonga, Dmitry Lepikhin, Kelvin Xu, Jacky Liang, Jiewen Tan, Benoit Schillings, Muge Ersoy, Pete Blois, Bernd Bandemer, Abhimanyu Singh, Sergei Lebedev, Pankaj Joshi, Adam R. Brown, Evan Palmer, Shreya Pathak, Komal Jalan, Fedir Zubach, Shuba Lall, Randall Parker, Alok Gunjan, Sergey Rogulenko, Sumit Sanghai, Zhaoqi Leng, Zoltan Egyed, Shixin Li, Maria Ivanova, Kostas Andriopoulos, Jin Xie, Elan Rosenfeld, Auriel Wright, Ankur Sharma, Xinyang Geng, Yicheng Wang, Sam Kwei, Renke Pan, Yujing Zhang, Gabby Wang, Xi Liu, Chak Yeung, Elizabeth Cole, Aviv Rosenberg, Zhen Yang, Phil Chen, George Polovets, Pranav Nair, Rohun Saxena, Josh

Smith, Shuo yiin Chang, Aroma Mahendru, Svetlana Grant, Anand Iyer, Irene Cai, Jed McGiffin, Jiaming Shen, Alanna Walton, Antonious Girgis, Oliver Woodman, Rosemary Ke, Mike Kwong, Louis Rouillard, Jinmeng Rao, Zhihao Li, Yuntao Xu, Flavien Prost, Chi Zou, Ziwei Ji, Alberto Magni, Tyler Liechty, Dan A. Calian, Deepak Ramachandran, Igor Krivokon, Hui Huang, Terry Chen, Anja Hauth, Anastasija Ili, Weijuan Xi, Hyeontaek Lim, Vlad-Doru Ion, Pooya Moradi, Metin Toksoz-Exley, Kalesha Bullard, Miltos Allamanis, Xiaomeng Yang, Sophie Wang, Zhi Hong, Anita Gergely, Cheng Li, Bhavishya Mittal, Vitaly Kovalev, Victor Ungureanu, Jane Labanowski, Jan Wassenberg, Nicolas Lacasse, Geoffrey Cideron, Petar Devi, Annie Marsden, Lynn Nguyen, Michael Fink, Yin Zhong, Tatsuya Kiyono, Desi Ivanov, Sally Ma, Max Bain, Kiran Yalasangi, Jennifer She, Anastasia Petrushkina, Mayank Lunayach, Carla Bromberg, Sarah Hodkinson, Vilobh Meshram, Daniel Vlasic, Austin Kyker, Steve Xu, Jeff Stanway, Zuguang Yang, Kai Zhao, Matthew Tung, Seth Odoom, Yasuhisa Fujii, Justin Gilmer, Eunyoung Kim, Felix Halim, Quoc Le, Bernd Bohnet, Seliem El-Sayed, Behnam Neyshabur, Malcolm Reynolds, Dean Reich, Yang Xu, Erica Moreira, Anuj Sharma, Zeyu Liu, Mohammad Javad Hosseini, Naina Raisinghani, Yi Su, Ni Lao, Daniel Formoso, Marco Gelmi, Almog Gueta, Tapomay Dey, Elena Gribovskaya, Domagoj evid, Sidharth Mudgal, Garrett Bingham, Jianling Wang, Anurag Kumar, Alex Cullum, Feng Han, Konstantinos Bousmalis, Diego Cedillo, Grace Chu, Vladimir Magay, Paul Michel, Ester Hlavnova, Daniele Calandriello, Setareh Ariafar, Kaisheng Yao, Vikash Sehwag, Arpi Vezer, Agustin Dal Lago, Zhenkai Zhu, Paul Kishan Rubenstein, Allen Porter, Anirudh Baddepudi, Oriana Riva, Mihai Dorin Istin, Chih-Kuan Yeh, Zhi Li, Andrew Howard, Nilpa Jha, Jeremy Chen, Raoul de Liedekerke, Zafarali Ahmed, Mikel Rodriguez, Tanuj Bhatia, Bangju Wang, Ali Elqursh, David Klinghoffer, Peter Chen, Pushmeet Kohli, Te I, Weiyang Zhang, Zack Nado, Jilin Chen, Maxwell Chen, George Zhang, Aayush Singh, Adam Hillier, Federico Lebron, Yiqing Tao, Ting Liu, Gabriel Dulac-Arnold, Jingwei Zhang, Shashi Narayan, Buhuang Liu, Orhan Firat, Abhishek Bhowmick, Bingyuan Liu, Hao Zhang, Zizhao Zhang, Georges Rotival, Nathan Howard, Anu Sinha, Alexander Grushetsky, Benjamin Beyret, Keerthana Gopalakrishnan, James Zhao, Kyle He, Szabolcs Payrits, Zaid Nabulsi, Zhaoyi Zhang, Weijie Chen, Edward Lee, Nova Fallen, Sreenivas Gollapudi, Aurick Zhou, Filip Paveti, Thomas Köppe, Shiyu Huang, Rama Pasumarthi, Nick Fernando, Felix Fischer, Daria urko, Yang Gao, James Svensson, Austin Stone, Haroon Qureshi, Abhishek Sinha, Apoorv Kulshreshtha, Martin Matysiak, Jieming Mao, Carl Saroufim, Aleksandra Faust, Qingnan Duan, Gil Fidel, Kaan Katircioglu, Raphaël Lopez Kaufman, Dhruv Shah, Weize Kong, Abhishek Bapna, Gellért Weisz, Emma Dunleavy, Praneet Dutta, Tianqi Liu, Rahma Chaabouni, Carolina Parada, Marcus Wu, Alexandra Belias, Alessandro Bissacco, Stanislav Fort, Li Xiao, Fantine Huot, Chris Knutsen, Yochai Blau, Gang Li, Jennifer Prendki, Juliette Love, Yinlam Chow, Pichi Charoenpanit, Hidetoshi Shimokawa, Vincent Coriou, Karol Gregor, Tomas Izo, Arjun Akula, Mario Pinto, Chris Hahn, Dominik Paulus, Jiaxian Guo, Neha Sharma, Cho-Jui Hsieh, Adaeze Chukwuka, Kazuma Hashimoto, Nathalie Rauschmayr, Ling Wu, Christof Angermueller, Yulong Wang, Sebastian Gerlach, Michael Pliskin, Daniil Mirylenka, Min Ma, Lexi Baugher, Bryan Gale, Shaan Bijwadia, Nemanja Rakievi, David Wood, Jane Park, Chung-Ching Chang, Babi Seal, Chris Tar, Kacper Krasowiak, Yiwen Song, Georgi Stephanov, Gary Wang, Marcello Maggioni, Stein Xudong Lin, Felix Wu, Shachi Paul, Zixuan Jiang, Shubham Agrawal, Bilal Piot, Alex Feng, Cheolmin Kim, Tulsee Doshi, Jonathan Lai, Chuqiao, Xu, Sharad Vikram, Ciprian Chelba, Sebastian Krause, Vincent Zhuang, Jack Rae, Timo Denk, Adrian Collister, Lotte Weerts, Xianghong Luo, Yifeng Lu, Håvard Garnes, Nitish Gupta, Terry Spitz, Avinatan Hassidim, Lihao Liang, Izhak Shafran, Peter Humphreys, Kenny Vassigh, Phil Wallis, Virat Shejwalkar, Nicolas Perez-Nieves, Rachel Hornung, Melissa Tan, Beka Westberg, Andy Ly, Richard Zhang, Brian Farris, Jongbin Park, Alec Kosik, Zeynep Cankara, Andrii Maksai, Yunhan Xu, Albin Cassirer, Sergi Caelles, Abbas Abdolmaleki, Mencher Chiang, Alex Fabrikant, Shravya Shetty, Luheng He, Mai Giménez, Hadi Hashemi, Sheena Panthaplackel, Yana Kulizhskaya, Salil Deshmukh, Daniele Pighin, Robin Alazard, Disha Jindal, Seb Noury, Pradeep Kumar S, Siyang Qin, Xerxes Dotiwalla, Stephen Spencer, Mohammad Babaeizadeh, Blake JianHang Chen, Vaibhav Mehta, Jennie Lees, Andrew Leach, Penporn Koanantakool, Ilia Akolzin, Ramona Comanescu, Junwhan Ahn, Alexey Svyatkovskiy, Basil Mustafa, David D'Ambrosio, Shiva Mohan Reddy Garlapati, Pascal Lamblin, Alekh Agarwal, Shuang Song, Pier Giuseppe Sessa, Pauline Coquinot, John Maggs, Hussain Masoom, Divya Pitta, Yaqing Wang, Patrick Morris-Suzuki, Billy Porter, Johnson Jia, Jeffrey Dudek, Raghavender R, Cosmin Paduraru, Alan Ansell, Tolga Bolukbasi, Tony Lu, Ramya Ganeshan, Zi Wang, Henry Griffiths, Rodrigo Benenson, Yifan He, James Swirhun, George Papamakarios, Aditya Chawla, Kuntal

Sengupta, Yan Wang, Vedrana Milutinovic, Igor Mordatch, Zhipeng Jia, Jamie Smith, Will Ng, Shitij Nigam, Matt Young, Eugen Vuak, Blake Hechtman, Sheela Goenka, Avital Zipori, Kareem Ayoub, Ashok Popat, Trilok Acharya, Luo Yu, Dawn Bloxwich, Hugo Song, Paul Roit, Haiqiong Li, Aviel Boag, Nigamaa Nayakanti, Bilva Chandra, Tianli Ding, Aahil Mehta, Cath Hope, Jiageng Zhang, Idan Heimlich Shtacher, Kartikeya Badola, Ryo Nakashima, Andrei Sozanschi, Iulia Coma, Ante uul, Emily Caveness, Julian Odell, Matthew Watson, Dario de Cesare, Phillip Lippe, Derek Lockhart, Siddharth Verma, Huizhong Chen, Sean Sun, Lin Zhuo, Aditya Shah, Prakhar Gupta, Alex Muzio, Ning Niu, Amir Zait, Abhinav Singh, Meenu Gaba, Fan Ye, Prajit Ramachandran, Mohammad Saleh, Raluca Ada Popa, Ayush Dubey, Frederick Liu, Sara Javanmardi, Mark Epstein, Ross Hemsley, Richard Green, Nishant Ranka, Eden Cohen, Chuyuan Kelly Fu, Sanjay Ghemawat, Jed Borovik, James Martens, Anthony Chen, Pranav Shyam, André Susano Pinto, Ming-Hsuan Yang, Alexandru ifrea, David Du, Boqing Gong, Ayushi Agarwal, Seungyeon Kim, Christian Frank, Saloni Shah, Xiaodan Song, Zhiwei Deng, Ales Mikhalap, Kleopatra Chatziprimou, Timothy Chung, Toni Creswell, Susan Zhang, Yennie Jun, Carl Lebsack, Will Truong, Slavica Andai, Itay Yona, Marco Fornoni, Rong Rong, Serge Toropov, Afzal Shama Soudagar, Andrew Audibert, Salah Zaiem, Zaheer Abbas, Andrei Rusu, Sahitya Potluri, Shitao Weng, Anastasios Kementsietsidis, Anton Tsitsulin, Daiyi Peng, Natalie Ha, Sanil Jain, Tejasi Latkar, Simeon Ivanov, Cory McLean, Anirudh GP, Rajesh Venkataraman, Canoee Liu, Dilip Krishnan, Joel D'sa, Roey Yogev, Paul Collins, Benjamin Lee, Lewis Ho, Carl Doersch, Gal Yona, Shawn Gao, Felipe Tiengo Ferreira, Adnan Ozturel, Hannah Muckenhirn, Ce Zheng, Gargi Balasubramaniam, Mudit Bansal, George van den Driessche, Sivan Eiger, Salem Haykal, Vedant Misra, Abhimanyu Goyal, Danilo Martins, Gary Leung, Jonas Valfridsson, Four Flynn, Will Bishop, Chenxi Pang, Yoni Halpern, Honglin Yu, Lawrence Moore, Yuvein, Zhu, Sridhar Thiagarajan, Yoel Drori, Zhisheng Xiao, Lucio Dery, Rolf Jagerman, Jing Lu, Eric Ge, Vaibhav Aggarwal, Arjun Khare, Vinh Tran, Oded Elyada, Ferran Alet, James Rubin, Ian Chou, David Tian, Libin Bai, Lawrence Chan, Lukasz Lew, Karolis Misiunas, Taylan Bilal, Aniket Ray, Sindhu Raghuram, Alex Castro-Ros, Viral Carpenter, CJ Zheng, Michael Kilgore, Josef Broder, Emily Xue, Praveen Kallakuri, Dheeru Dua, Nancy Yuen, Steve Chien, John Schultz, Saurabh Agrawal, Reut Tsarfaty, Jingcao Hu, Ajay Kannan, Dror Marcus, Nisarg Kothari, Baochen Sun, Ben Horn, Matko Bonjak, Ferjad Naeem, Dean Hirsch, Lewis Chiang, Boya Fang, Jie Han, Qifei Wang, Ben Hora, Antoine He, Mario Lui, Beer Changpinyo, Anshuman Tripathi, John Youssef, Chester Kwak, Philippe Schlattner, Cat Graves, Rémi Leblond, Wenjun Zeng, Anders Andreassen, Gabriel Rasskin, Yue Song, Eddie Cao, Junhyuk Oh, Matt Hoffman, Wojtek Skut, Yichi Zhang, Jon Stritar, Xingyu Cai, Saarthak Khanna, Kathie Wang, Shriya Sharma, Christian Reisswig, Younghoon Jun, Aman Prasad, Tatiana Sholokhova, Preeti Singh, Adi Gerzi Rosenthal, Anian Ruoss, Françoise Beaufays, Sean Kirmani, Dongkai Chen, Johan Schalkwyk, Jonathan Herzig, Been Kim, Josh Jacob, Damien Vincent, Adrian N Reyes, Ivana Balazevic, Léonard Hussenot, Jon Schneider, Parker Barnes, Luis Castro, Spandana Raj Babbula, Simon Green, Serkan Cabi, Nico Duduta, Danny Driess, Rich Galt, Noam Velan, Junjie Wang, Hongyang Jiao, Matthew Mauger, Du Phan, Miteyan Patel, Vlado Gali, Jerry Chang, Eyal Marcus, Matt Harvey, Julian Salazar, Elahe Dabir, Suraj Satishkumar Sheth, Amol Mandhane, Hanie Sedghi, Jeremiah Willcock, Amir Zandieh, Shruthi Prabhakara, Aida Amini, Antoine Miech, Victor Stone, Massimo Nicosia, Paul Niemczyk, Ying Xiao, Lucy Kim, Sawek Kwasiborski, Vikas Verma, Ada Maksutaj Oflazer, Christoph Hirnschall, Peter Sung, Lu Liu, Richard Everett, Michiel Bakker, Ágoston Weisz, Yufei Wang, Vivek Sampathkumar, Uri Shaham, Bibo Xu, Yasemin Altun, Mingqiu Wang, Takaaki Saeki, Guanjie Chen, Emanuel Taropa, Shanthal Vasanth, Sophia Austin, Lu Huang, Goran Petrovic, Qingyun Dou, Daniel Golovin, Grigory Rozhdestvenskiy, Allie Culp, Will Wu, Motoki Sano, Divya Jain, Julia Proskurnia, Sébastien Cevey, Alejandro Cruzado Ruiz, Piyush Patil, Mahdi Mirzazadeh, Eric Ni, Javier Snaider, Lijie Fan, Alexandre Fréchette, AJ Pierigiovanni, Shariq Iqbal, Kenton Lee, Claudio Fantacci, Jinwei Xing, Lisa Wang, Alex Irpan, David Raposo, Yi Luan, Zhuoyuan Chen, Harish Ganapathy, Kevin Hui, Jiazhong Nie, Isabelle Guyon, Heming Ge, Roopali Vij, Hui Zheng, Dayeong Lee, Alfonso Castaño, Khuslen Baatarsukh, Gabriel Ibagon, Alexandra Chronopoulou, Nicholas FitzGerald, Shashank Viswanadha, Safeen Huda, Rivka Moroshko, Georgi Stoyanov, Prateek Kolhar, Alain Vaucher, Ishaan Watts, Adhi Kuncoro, Henryk Michalewski, Satish Kambala, Bat-Orgil Batsaikhan, Alek Andreev, Irina Jurenka, Maigo Le, Qihang Chen, Wael Al Jishi, Sarah Chakera, Zhe Chen, Aditya Kini, Vikas Yadav, Aditya Siddhant, Ilia Labzovsky, Balaji Lakshminarayanan, Carrie Grimes Bostock, Pankil Botadra, Ankesh Anand, Colton Bishop, Sam

Conway-Rahman, Mohit Agarwal, Yani Donchev, Achintya Singhal, Félix de Chaumont Quitry, Natalia Ponomareva, Nishant Agrawal, Bin Ni, Kalpesh Krishna, Masha Samsikova, John Karro, Yilun Du, Tamara von Glehn, Caden Lu, Christopher A. Choquette-Choo, Zhen Qin, Tingnan Zhang, Sicheng Li, Divya Tyam, Swaroop Mishra, Wing Lowe, Colin Ji, Weiyi Wang, Manaal Faruqui, Ambrose Slone, Valentin Dalibard, Arunachalam Narayanaswamy, John Lambert, Pierre-Antoine Manzagol, Dan Karliner, Andrew Bolt, Ivan Lobov, Aditya Kusupati, Chang Ye, Xuan Yang, Heiga Zen, Nelson George, Mukul Bhutani, Olivier Lacombe, Robert Riachi, Gagan Bansal, Rachel Soh, Yue Gao, Yang Yu, Adams Yu, Emily Nottage, Tania Rojas-Esponda, James Noraky, Manish Gupta, Ragha Kotikalapudi, Jichuan Chang, Sanja Deur, Dan Graur, Alex Mossin, Erin Farnese, Ricardo Figueira, Alexandre Moufarek, Austin Huang, Patrik Zochbauer, Ben Ingram, Tongzhou Chen, Zelin Wu, Adrià Puigdomènech, Leland Rechis, Da Yu, Sri Gayatri Sundara Padmanabhan, Rui Zhu, Chu ling Ko, Andrea Banino, Samira Daruki, Aarush Selvan, Dhruva Bhaswar, Daniel Hernandez Diaz, Chen Su, Salvatore Scellato, Jennifer Brennan, Woohyun Han, Grace Chung, Priyanka Agrawal, Urvashi Khandelwal, Khe Chai Sim, Morgane Lustman, Sam Ritter, Kelvin Guu, Jiawei Xia, Prateek Jain, Emma Wang, Tyrone Hill, Mirko Rossini, Marija Kostelac, Tautvydas Misiunas, Amit Sabne, Kyuyeun Kim, Ahmet Iscen, Congchao Wang, José Leal, Ashwin Sreevatsa, Utku Evci, Manfred Warmuth, Saket Joshi, Daniel Suo, James Lottes, Garrett Honke, Brendan Jou, Stefani Karp, Jieru Hu, Himanshu Sahni, Adrien Ali Taïga, William Kong, Samrat Ghosh, Renshen Wang, Jay Pavagadhi, Natalie Axelsson, Nikolai Grigorev, Patrick Siegler, Rebecca Lin, Guohui Wang, Emilio Parisotto, Sharath Maddineni, Krishan Subudhi, Eyal Ben-David, Elena Pochernina, Orgad Keller, Thi Avrahami, Zhe Yuan, Pulkit Mehta, Jialu Liu, Sherry Yang, Wendy Kan, Katherine Lee, Tom Funkhouser, Derek Cheng, Hongzhi Shi, Archit Sharma, Joe Kelley, Matan Eyal, Yury Malkov, Corentin Tallec, Yuval Bahat, Shen Yan, Xintian, Wu, David Lindner, Chengda Wu, Avi Caciularu, Xiyang Luo, Rodolphe Jenatton, Tim Zaman, Yingying Bi, Ilya Kornakov, Ganesh Mallya, Daisuke Ikeda, Itay Karo, Anima Singh, Colin Evans, Praneeth Netrapalli, Vincent Nallatamby, Isaac Tian, Yannis Assael, Vikas Raunak, Victor Carbune, Ioana Bica, Lior Madmoni, Dee Cattle, Snchit Grover, Krishna Somandepalli, Sid Lall, Amelio Vázquez-Reina, Riccardo Patana, Jiaqi Mu, Pranav Talluri, Maggie Tran, Rajeev Aggarwal, RJ Skerry-Ryan, Jun Xu, Mike Burrows, Xiaoyue Pan, Edouard Yvinec, Di Lu, Zhiying Zhang, Duc Dung Nguyen, Hairong Mu, Gabriel Barcik, Helen Ran, Lauren Beltrone, Krzysztof Choromanski, Dia Kharrat, Samuel Albanie, Sean Purser-haskell, David Bieber, Carrie Zhang, Jing Wang, Tom Hudson, Zhiyuan Zhang, Han Fu, Johannes Mauerer, Mohammad Hossein Bateni, AJ Maschinot, Bing Wang, Muye Zhu, Arjun Pillai, Tobias Weyand, Shuang Liu, Oscar Akerlund, Fred Bertsch, Vittal Premachandran, Alicia Jin, Vincent Roulet, Peter de Boursac, Shubham Mittal, Ndaba Ndebele, Georgi Karadzhov, Sahra Ghalebikesabi, Ricky Liang, Allen Wu, Yale Cong, Nimesh Ghelani, Sumeet Singh, Bahar Fatemi, Warren, Chen, Charles Kwong, Alexey Kolganov, Steve Li, Richard Song, Chenkai Kuang, Sobhan Miryoosefi, Dale Webster, James Wendt, Arkadiusz Socala, Guolong Su, Artur Mendonça, Abhinav Gupta, Xiaowei Li, Tomy Tsai, Qiong, Hu, Kai Kang, Angie Chen, Sertan Girgin, Yongqin Xian, Andrew Lee, Nolan Ramsden, Leslie Baker, Madeleine Clare Elish, Varvara Krayvanova, Rishabh Joshi, Jiri Simsa, Yao-Yuan Yang, Piotr Ambroszczyk, Dipankar Ghosh, Arjun Kar, Yuan Shangguan, Yumeya Yamamori, Yaroslav Akulov, Andy Brock, Haotian Tang, Siddharth Vashishtha, Rich Munoz, Andreas Steiner, Kalyan Andra, Daniel Eppens, Qixuan Feng, Hayato Kobayashi, Sasha Goldshtein, Mona El Mahdy, Xin Wang, Jilei, Wang, Richard Killam, Tom Kwiatkowski, Kavya Kopparapu, Serena Zhan, Chao Jia, Alexei Bendebury, Sheryl Luo, Adrià Recasens, Timothy Knight, Jing Chen, Mohak Patel, YaGuang Li, Ben Withbroe, Dean Weesner, Kush Bhatia, Jie Ren, Danielle Eisenbud, Ebrahim Songhori, Yanhua Sun, Travis Choma, Tasos Kementsietsidis, Lucas Manning, Brian Roark, Wael Farhan, Jie Feng, Susheel Tatineni, James Cobon-Kerr, Yunjie Li, Lisa Anne Hendricks, Isaac Noble, Chris Breaux, Nate Kushman, Liqian Peng, Fuzhao Xue, Taylor Tobin, Jamie Rogers, Josh Lipschultz, Chris Alberti, Alexey Vlaskin, Mostafa Dehghani, Roshan Sharma, Tris Warkentin, Chen-Yu Lee, Benigno Uria, Da-Cheng Juan, Angad Chandorkar, Hila Sheftel, Ruibo Liu, Elnaz Davoodi, Borja De Balle Pigem, Kedar Dhamdhere, David Ross, Jonathan Hoech, Mahdis Mahdieh, Li Liu, Qiujia Li, Liam McCafferty, Chenxi Liu, Markus Mircea, Yunting Song, Omkar Savant, Alaa Saade, Colin Cherry, Vincent Hellendoorn, Siddharth Goyal, Paul Pucciarelli, David Vilar Torres, Zohar Yahav, Hyo Lee, Lars Lowe Sjoesund, Christo Kirov, Bo Chang, Deepanway Ghoshal, Lu Li, Gilles Baechler, Sébastien Pereira, Tara Sainath, Anudhyan Boral, Dominik Grewe, Afief Halumi, Nguyet Minh Phu, Tianxiao Shen, Marco Tulio Ribeiro, Dhriti Varma, Alex Kaskasoli, Vlad Feinberg,

Navneet Potti, Jarrod Kahn, Matheus Wisniewski, Shakir Mohamed, Arnar Mar Hrafnkelsson, Bobak Shahriari, Jean-Baptiste Lespiau, Lisa Patel, Legg Yeung, Tom Paine, Lantao Mei, Alex Ramirez, Rakesh Shivanna, Li Zhong, Josh Woodward, Guilherme Tubone, Samira Khan, Heng Chen, Elizabeth Nielsen, Catalin Ionescu, Utsav Prabhu, Mingcen Gao, Qingze Wang, Sean Augenstein, Neesha Subramaniam, Jason Chang, Fotis Iliopoulos, Jiaming Luo, Myriam Khan, Weicheng Kuo, Denis Teplyashin, Florence Perot, Logan Kilpatrick, Amir Globerson, Hongkun Yu, Anfal Siddiqui, Nick Sukhanov, Arun Kandoor, Umang Gupta, Marco Andreetto, Moran Ambar, Donnie Kim, Pawe Wesoowski, Sarah Perrin, Ben Limonchik, Wei Fan, Jim Stephan, Ian Stewart-Binks, Ryan Kappedal, Tong He, Sarah Cogan, Romina Datta, Tong Zhou, Jiayu Ye, Leandro Kieliger, Ana Ramalho, Kyle Kastner, Fabian Mentzer, Wei-Jen Ko, Arun Suggala, Tianhao Zhou, Shiraz Butt, Hana Strejek, Lior Belenki, Subhashini Venugopalan, Mingyang Ling, Evgenii Eltyshev, Yunxiao Deng, Geza Kovacs, Mukund Raghavachari, Hanjun Dai, Tal Schuster, Steven Schwarcz, Richard Nguyen, Arthur Nguyen, Gavin Buttimore, Shrestha Basu Mallick, Sudeep Gandhe, Seth Benjamin, Michal Jastrzebski, Le Yan, Sugato Basu, Chris Apps, Isabel Edkins, James Allingham, Immanuel Odisho, Tomas Kocisky, Jewel Zhao, Linting Xue, Apoorv Reddy, Chrysovalantis Anastasiou, Aviel Atias, Sam Redmond, Kieran Milan, Nicolas Heess, Herman Schmit, Allan Dafoe, Daniel Andor, Tynan Gangwani, Anca Dragan, Sheng Zhang, Ashyana Kachra, Gang Wu, Siyang Xue, Kevin Aydin, Siqi Liu, Yuxiang Zhou, Mahan Malihi, Austin Wu, Siddharth Gopal, Candice Schumann, Peter Stys, Alek Wang, Mirek Olák, Dangyi Liu, Christian Schallhart, Yiran Mao, Demetra Brady, Hao Xu, Tomas Mery, Chawin Sitawarin, Siva Velusamy, Tom Cobley, Alex Zhai, Christian Walder, Nitzan Katz, Ganesh Jawahar, Chinmay Kulkarni, Antoine Yang, Adam Paszke, Yinan Wang, Bogdan Damoc, Zalán Borsos, Ray Smith, Jinning Li, Mansi Gupta, Andrei Kapishnikov, Sushant Prakash, Florian Luisier, Rishabh Agarwal, Will Grathwohl, Kuangyuan Chen, Kehang Han, Nikhil Mehta, Andrew Over, Shekoofeh Azizi, Lei Meng, Niccolò Dal Santo, Kelvin Zheng, Jane Shapiro, Igor Petrovski, Jeffrey Hui, Amin Ghafouri, Jasper Snoek, James Qin, Mandy Jordan, Caitlin Sikora, Jonathan Malmaud, Yuheng Kuang, Aga wietlik, Ruoxin Sang, Chongyang Shi, Leon Li, Andrew Rosenberg, Shubin Zhao, Andy Crawford, Jan-Thorsten Peter, Yun Lei, Xavier Garcia, Long Le, Todd Wang, Julien Amelot, Dave Orr, Praneeth Kacham, Dana Alon, Gladys Tyen, Abhinav Arora, James Lyon, Alex Kurakin, Mimi Ly, Theo Guidroz, Zhipeng Yan, Rina Panigrahy, Pingmei Xu, Thais Kagohara, Yong Cheng, Eric Noland, Jinhyuk Lee, Jonathan Lee, Cathy Yip, Maria Wang, Efrat Nehoran, Alexander Bykovsky, Zhihao Shan, Ankit Bhagatwala, Chaochao Yan, Jie Tan, Guillermo Garrido, Dan Ethier, Nate Hurley, Grace Vesom, Xu Chen, Siyuan Qiao, Abhishek Nayyar, Julian Walker, Paramjit Sandhu, Mihaela Rosca, Danny Swisher, Mikhail Dektiarev, Josh Dillon, George-Cristian Muraru, Manuel Tragut, Artiom Myaskovsky, David Reid, Marko Velic, Owen Xiao, Jasmine George, Mark Brand, Jing Li, Wenhao Yu, Shane Gu, Xiang Deng, François-Xavier Aubet, Soheil Hassas Yeganeh, Fred Alcober, Celine Smith, Trevor Cohn, Kay McKinney, Michael Tschannen, Ramesh Sampath, Gowoon Cheon, Liangchen Luo, Luyang Liu, Jordi Orbay, Hui Peng, Gabriela Botea, Xiaofan Zhang, Charles Yoon, Cesar Magalhaes, Pawe Stradomski, Ian Mackinnon, Steven Hemingray, Kumaran Venkatesan, Rhys May, Jaeyoun Kim, Alex Druinsky, Jingchen Ye, Zheng Xu, Terry Huang, Jad Al Abdallah, Adil Dostmohamed, Rachana Fellinger, Tsendsuren Munkhdalai, Akanksha Maurya, Peter Garst, Yin Zhang, Maxim Krikun, Simon Bucher, Aditya Srikanth Veerubhotla, Yaxin Liu, Sheng Li, Nishesh Gupta, Jakub Adamek, Hanwen Chen, Bernett Orlando, Aleksandr Zaks, Joost van Amersfoort, Josh Camp, Hui Wan, HyunJeong Choe, Zhichun Wu, Kate Olszewska, Weiren Yu, Archita Vadali, Martin Scholz, Daniel De Freitas, Jason Lin, Amy Hua, Xin Liu, Frank Ding, Yichao Zhou, Boone Severson, Katerina Tsihlas, Samuel Yang, Tammo Spalink, Varun Yerram, Helena Pankov, Rory Blevins, Ben Vargas, Sarthak Jauhari, Matt Miecnikowski, Ming Zhang, Sandeep Kumar, Clement Farabet, Charline Le Lan, Sebastian Flennerhag, Yonatan Bitton, Ada Ma, Arthur Brainskas, Eli Collins, Niharika Ahuja, Sneha Kudugunta, Anna Bortsova, Minh Giang, Wanzheng Zhu, Ed Chi, Scott Lundberg, Alexey Stern, Subha Puttagunta, Jing Xiong, Xiao Wu, Yash Pande, Amit Jhindal, Daniel Murphy, Jon Clark, Marc Brockschmidt, Maxine Deines, Kevin R. McKee, Dan Bahir, Jiajun Shen, Minh Truong, Daniel McDuff, Andrea Gesmundo, Edouard Rosseel, Bowen Liang, Ken Caluwaerts, Jessica Hamrick, Joseph Kready, Mary Cassin, Rishikesh Ingale, Li Lao, Scott Pollom, Yifan Ding, Wei He, Lizzetth Bellot, Joana Iljazi, Ramya Sree Boppana, Shan Han, Tara Thompson, Amr Khalifa, Anna Bulanova, Blagoj Mitrevski, Bo Pang, Emma Cooney, Tian Shi, Rey Coaguila, Tamar Yakar, Marc'aurelio Ranzato, Nikola Momchev, Chris Rawles, Zachary Charles, Young Maeng, Yuan Zhang, Rishabh Bansal, Xiaokai Zhao,

Brian Albert, Yuan Yuan, Sudheendra Vijayanarasimhan, Roy Hirsch, Vinay Ramasesh, Kiran Vodrahalli, Xingyu Wang, Arushi Gupta, DJ Strouse, Jianmo Ni, Roma Patel, Gabe Taubman, Zhouyuan Huo, Dero Gharibian, Marianne Monteiro, Hoi Lam, Shobha Vasudevan, Aditi Chaudhary, Isabela Albuquerque, Kilol Gupta, Sebastian Riedel, Chaitra Hegde, Avraham Ruderman, András György, Marcus Wainwright, Ashwin Chaugule, Burcu Karagol Ayan, Tomer Levinboim, Sam Shleifer, Yogesh Kalley, Vahab Mirrokni, Abhishek Rao, Prabakar Radhakrishnan, Jay Hartford, Jialin Wu, Zhenhai Zhu, Francesco Bertolini, Hao Xiong, Nicolas Serrano, Hamish Tomlinson, Myle Ott, Yifan Chang, Mark Graham, Jian Li, Marco Liang, Xiangzhu Long, Sebastian Borgeaud, Yanif Ahmad, Alex Grills, Diana Mincu, Martin Izzard, Yuan Liu, Jinyu Xie, Louis O'Bryan, Sameera Ponda, Simon Tong, Michelle Liu, Dan Malkin, Khalid Salama, Yuankai Chen, Rohan Anil, Anand Rao, Rigel Swavely, Misha Bilenko, Nina Anderson, Tat Tan, Jing Xie, Xing Wu, Lijun Yu, Oriol Vinyals, Andrey Ryabtsev, Rumen Dangovski, Kate Baumli, Daniel Keysers, Christian Wright, Zoe Ashwood, Betty Chan, Artem Shtefan, Yaohui Guo, Ankur Bapna, Radu Soricut, Steven Pecht, Sabela Ramos, Rui Wang, Jiahao Cai, Trieu Trinh, Paul Barham, Linda Friso, Eli Stickgold, Xiangzhuo Ding, Siamak Shakeri, Diego Ardila, Eleftheria Briakou, Phil Culliton, Adam Raveret, Jingyu Cui, David Saxton, Subhrajit Roy, Javad Azizi, Pengcheng Yin, Lucia Loher, Andrew Bunner, Min Choi, Faruk Ahmed, Eric Li, Yin Li, Shengyang Dai, Michael Elabd, Sriram Ganapathy, Shivani Agrawal, Yiqing Hua, Paige Kunkle, Sujeevan Rajayogam, Arun Ahuja, Arthur Conmy, Alex Vasiloff, Parker Beak, Christopher Yew, Jayaram Mudigonda, Bartek Wydrowski, Jon Blanton, Zhengdong Wang, Yann Dauphin, Zhuo Xu, Martin Polacek, Xi Chen, Hexiang Hu, Pauline Sho, Markus Kunesch, Mehdi Hafezi Manshadi, Eliza Rutherford, Bo Li, Sissie Hsiao, Iain Barr, Alex Tudor, Matija Kecman, Arsha Nagrani, Vladimir Pchelin, Martin Sundermeyer, Aishwarya P S, Abhijit Karmarkar, Yi Gao, Grishma Chole, Olivier Bachem, Isabel Gao, Arturo BC, Matt Dibb, Mauro Verzetti, Felix Hernandez-Campos, Yana Lunts, Matthew Johnson, Julia Di Trapani, Raphael Koster, Idan Brusilovsky, Binbin Xiong, Megha Mohabey, Han Ke, Joe Zou, Tea Saboli, Víctor Campos, John Palowitch, Alex Morris, Linhai Qiu, Pranavaraj Ponnuramu, Fangtao Li, Vivek Sharma, Kiranbir Sodhia, Kaan Tekelioglu, Aleksandr Chuklin, Madhavi Yenugula, Erika Gemzer, Theofilos Strinopoulos, Sam El-Husseini, Huiyu Wang, Yan Zhong, Edouard Leurent, Paul Natsev, Weijun Wang, Dre Mahaarachchi, Tao Zhu, Songyou Peng, Sami Alabed, Cheng-Chun Lee, Anthony Brohan, Arthur Szlam, GS Oh, Anton Kovsharov, Jenny Lee, Renee Wong, Megan Barnes, Gregory Thornton, Felix Gimeno, Omer Levy, Martin Sevenich, Melvin Johnson, Jonathan Mallinson, Robert Dadashi, Ziyue Wang, Qingchun Ren, Preethi Lahoti, Arka Dhar, Josh Feldman, Dan Zheng, Thatcher Ulrich, Liviu Panait, Michiel Blokzijl, Cip Baetu, Josip Matak, Jitendra Harlalka, Maulik Shah, Tal Marian, Daniel von Dincklage, Cosmo Du, Ruy Ley-Wild, Bethanie Brownfield, Max Schumacher, Yury Stuken, Shadi Noghabi, Sonal Gupta, Xiaoqi Ren, Eric Malmi, Felix Weissenberger, Blanca Huergo, Maria Bauza, Thomas Lampe, Arthur Douillard, Mojtaba Seyedhosseini, Roy Frostig, Zoubin Ghahramani, Kelvin Nguyen, Kashyap Krishnakumar, Chengxi Ye, Rahul Gupta, Alireza Nazari, Robert Geirhos, Pete Shaw, Ahmed Eleryan, Dima Damen, Jennimaria Palomaki, Ted Xiao, Qiyin Wu, Quan Yuan, Phoenix Meadowlark, Matthew Bilotti, Raymond Lin, Mukund Sridhar, Yannick Schroecker, Da-Woon Chung, Jincheng Luo, Trevor Strohman, Tianlin Liu, Anne Zheng, Jesse Emond, Wei Wang, Andrew Lampinen, Toshiyuki Fukuzawa, Folawiyo Campbell-Ajala, Monica Roy, James Lee-Thorp, Lily Wang, Iftekhar Naim, Tony, Nguy ên, Guy Bensky, Aditya Gupta, Dominika Rogoziska, Justin Fu, Thanumalayan Sankaranarayana Pillai, Petar Velikovi, Shahar Drath, Philipp Neubeck, Vaibhav Tulsyan, Arseniy Klimovskiy, Don Metzler, Sage Stevens, Angel Yeh, Junwei Yuan, Tianhe Yu, Kelvin Zhang, Alec Go, Vincent Tsang, Ying Xu, Andy Wan, Isaac Galatzer-Levy, Sam Sobell, Abodunrinwa Toki, Elizabeth Salesky, Wenlei Zhou, Diego Antognini, Sholto Douglas, Shimu Wu, Adam Lelkes, Frank Kim, Paul Cavallaro, Ana Salazar, Yuchi Liu, James Besley, Tiziana Refice, Yiling Jia, Zhang Li, Michal Sokolik, Arvind Kannan, Jon Simon, Jo Chick, Avia Aharon, Meet Gandhi, Mayank Daswani, Keyvan Amiri, Vighnesh Birodkar, Abe Ittycheriah, Peter Grabowski, Oscar Chang, Charles Sutton, Zhixin, Lai, Umesh Telang, Susie Sargsyan, Tao Jiang, Raphael Hoffmann, Nicole Brichtova, Matteo Hessel, Jonathan Halcrow, Sammy Jerome, Geoff Brown, Alex Tomala, Elena Buchatskaya, Dian Yu, Sachit Menon, Pol Moreno, Yuguo Liao, Vicky Zayats, Luming Tang, SQ Mah, Ashish Shenoy, Alex Siegman, Majid Hadian, Okwan Kwon, Tao Tu, Nima Khajehnouri, Ryan Foley, Parisa Haghani, Zhongru Wu, Vaishakh Keshava, Khyatti Gupta, Tony Bruguier, Rui Yao, Danny Karmon, Luisa Zintgraf, Zhicheng Wang, Enrique Piqueras, Junehyuk Jung, Jenny Brennan, Diego Machado, Marissa Giustina, MH Tessler, Kamyu Lee, Qiao Zhang, Joss Moore, Kaspar Daugaard, Alexander Frömmgen, Jennifer Beattie, Fred Zhang, Daniel Kasenberg, Ty Geri, Danfeng Qin, Gaurav Singh Tomar, Tom Ouyang, Tianli Yu, Luowei Zhou, Rajiv Mathews, Andy Davis, Yaoyiran Li, Jai Gupta, Damion Yates, Linda Deng, Elizabeth Kemp, Ga-Young Joung, Sergei Vassilvitskii, Mandy Guo, Pallavi LV, Dave Dopson, Sami Lachgar, Lara McConnaughey, Himadri Choudhury, Dragos Dena, Aaron Cohen, Joshua Ainslie, Sergey Levi, Parthasarathy Gopavarapu, Polina Zablotskaia, Hugo Vallet, Sanaz Bahargam, Xiaodan Tang, Nenad Tomasev, Ethan Dyer, Daniel Balle, Hongrae Lee, William Bono, Jorge Gonzalez Mendez, Vadim Zubov, Shentao Yang, Ivor Rendulic, Yanyan Zheng, Andrew Hogue, Golan Pundak, Ralph Leith, Avishkar Bhoopchand, Michael Han, Mislav ani, Tom Schaul, Manolis Delakis, Tejas Iyer, Guanyu Wang, Harman Singh, Abdelrahman Abdelhamed, Tara Thomas, Siddhartha Brahma, Hilal Dib, Naveen Kumar, Wenxuan Zhou, Liang Bai, Pushkar Mishra, Jiao Sun, Valentin Anklin, Roykrong Sukkerd, Lauren Agubuzu, Anton Briukhov, Anmol Gulati, Maximilian Sieb, Fabio Pardo, Sara Nasso, Junquan Chen, Kexin Zhu, Tiberiu Sosea, Alex Goldin, Keith Rush, Spurthi Amba Hombaiah, Andreas Noever, Allan Zhou, Sam Haves, Mary Phuong, Jake Ades, Yi ting Chen, Lin Yang, Joseph Pagadora, Stan Bileschi, Victor Cotruta, Rachel Saputro, Arijit Pramanik, Sean Ammirati, Dan Garrette, Kevin Villela, Tim Blyth, Canfer Akbulut, Neha Jha, Alban Rrustemi, Arissa Wongpanich, Chirag Nagpal, Yonghui Wu, Morgane Rivière, Sergey Kishchenko, Pranesh Srinivasan, Alice Chen, Animesh Sinha, Trang Pham, Bill Jia, Tom Hennigan, Anton Bakalov, Nithya Attaluri, Drew Garmon, Daniel Rodriguez, Dawid Wegner, Wenhao Jia, Evan Senter, Noah Fiedel, Denis Petek, Yuchuan Liu, Cassidy Hardin, Harshal Tushar Lehri, Joao Carreira, Sara Smoot, Marcel Prasetya, Nami Akazawa, Anca Stefanoiu, Chia-Hua Ho, Anelia Angelova, Kate Lin, Min Kim, Charles Chen, Marcin Sieniek, Alice Li, Tongfei Guo, Sorin Baltateanu, Pouya Tafti, Michael Wunder, Nadav Olmert, Divyansh Shukla, Jingwei Shen, Neel Kovelamudi, Balaji Venkatraman, Seth Neel, Romal Thoppilan, Jerome Connor, Frederik Benzing, Axel Stjerngren, Golnaz Ghiasi, Alex Polozov, Joshua Howland, Theophane Weber, Justin Chiu, Ganesh Poomal Girirajan, Andreas Terzis, Pidong Wang, Fangda Li, Yoav Ben Shalom, Dinesh Tewari, Matthew Denton, Roee Aharoni, Norbert Kalb, Heri Zhao, Junlin Zhang, Angelos Filos, Matthew Rahtz, Lalit Jain, Connie Fan, Vitor Rodrigues, Ruth Wang, Richard Shin, Jacob Austin, Roman Ring, Mariella Sanchez-Vargas, Mehadi Hassen, Ido Kessler, Uri Alon, Gufeng Zhang, Wenhu Chen, Yenai Ma, Xiance Si, Le Hou, Azalia Mirhoseini, Marc Wilson, Geoff Bacon, Becca Roelofs, Lei Shu, Gautam Vasudevan, Jonas Adler, Artur Dwornik, Tayfun Terzi, Matt Lawlor, Harry Askham, Mike Bernico, Xuanyi Dong, Chris Hidey, Kevin Kilgour, Gaël Liu, Surya Bhupatiraju, Luke Leonhard, Siqi Zuo, Partha Talukdar, Qing Wei, Aliaksei Severyn, Vít Listík, Jong Lee, Aditya Tripathi, SK Park, Yossi Matias, Hao Liu, Alex Ruiz, Rajesh Jayaram, Jackson Tolins, Pierre Marcenac, Yiming Wang, Bryan Seybold, Henry Prior, Deepak Sharma, Jack Weber, Mikhail Sirotenko, Yunhsuan Sung, Dayou Du, Ellie Pavlick, Stefan Zinke, Markus Freitag, Max Dylla, Montse Gonzalez Arenas, Natan Potikha, Omer Goldman, Connie Tao, Rachita Chhaparia, Maria Voitovich, Pawan Dogra, Andrija Ranatovi, Zak Tsai, Chong You, Oleaser Johnson, George Tucker, Chenjie Gu, Jae Yoo, Maryam Majzoubi, Valentin Gabeur, Bahram Raad, Rocky Rhodes, Kashyap Kolipaka, Heidi Howard, Geta Sampemane, Benny Li, Chulayuth Asawaroengchai, Duy Nguyen, Chiyuan Zhang, Timothee Cour, Xinxin Yu, Zhao Fu, Joe Jiang, Po-Sen Huang, Gabriela Surita, Iñaki Iturrate, Yael Karov, Michael Collins, Martin Baeuml, Fabian Fuchs, Shilpa Shetty, Swaroop Ramaswamy, Sayna Ebrahimi, Qiuchen Guo, Jeremy Shar, Gabe Barth-Maron, Sravanti Addepalli, Bryan Richter, Chin-Yi Cheng, Eugénie Rives, Fei Zheng, Johannes Griesser, Nishanth Dikkala, Yoel Zeldes, Ilkin Safarli, Dipanjan Das, Himanshu Srivastava, Sadh MNM Khan, Xin Li, Aditya Pandey, Larisa Markeeva, Dan Belov, Qiqi Yan, Mikoaj Rybiski, Tao Chen, Megha Nawhal, Michael Quinn, Vineetha Govindaraj, Sarah York, Reed Roberts, Roopal Garg, Namrata Godbole, Jake Abernethy, Anil Das, Lam Nguyen Thiet, Jonathan Tompson, John Nham, Neera Vats, Ben Caine, Wesley Helmholz, Francesco Pongetti, Yeongil Ko, James An, Clara Huiyi Hu, Yu-Cheng Ling, Julia Pawar, Robert Leland, Keisuke Kinoshita, Waleed Khawaja, Marco Selvi, Eugene Ie, Danila Sinopalnikov, Lev Proleev, Nilesh Tripuraneni, Michele Bevilacqua, Seungji Lee, Clayton Sanford, Dan Suh, Dustin Tran, Jeff Dean, Simon Baumgartner, Jens Heitkaemper, Sagar Gubbi, Kristina Toutanova, Yichong Xu, Chandu Thekkath, Keran Rong, Palak Jain, Annie Xie, Yan Virin, Yang Li, Lubo Litchev, Richard Powell, Tarun Bharti, Adam Kraft, Nan Hua, Marissa Ikonomidis, Ayal Hitron, Sanjiv Kumar, Loic Matthey, Sophie Bridgers, Lauren Lax, Ishaan Malhi, Ondrej Skopek, Ashish Gupta, Jiawei Cao, Mitchelle Rasquinha, Siim Põder, Wojciech Stokowiec, Nicholas Roth, Guowang Li, Michaël Sander, Joshua Kessinger, Vihan Jain, Edward

Loper, Wonpyo Park, Michal Yarom, Liqun Cheng, Guru Guruganesh, Kanishka Rao, Yan Li, Catarina Barros, Mikhail Sushkov, Chun-Sung Ferng, Rohin Shah, Ophir Aharoni, Ravin Kumar, Tim McConnell, Peiran Li, Chen Wang, Fernando Pereira, Craig Swanson, Fayaz Jamil, Yan Xiong, Anitha Vijayakumar, Prakash Shroff, Kedar Soparkar, Jindong Gu, Livio Baldini Soares, Eric Wang, Kushal Majmundar, Aurora Wei, Kai Bailey, Nora Kassner, Chizu Kawamoto, Goran ui, Victor Gomes, Abhirut Gupta, Michael Guzman, Ishita Dasgupta, Xinyi Bai, Zhufeng Pan, Francesco Piccinno, Hadas Natalie Vogel, Octavio Ponce, Adrian Hutter, Paul Chang, Pan-Pan Jiang, Ionel Gog, Vlad Ionescu, James Manyika, Fabian Pedregosa, Harry Ragan, Zach Behrman, Ryan Mullins, Coline Devin, Aroonalok Pyne, Swapnil Gawde, Martin Chadwick, Yiming Gu, Sasan Tavakkol, Andy Twigg, Naman Goyal, Ndidi Elue, Anna Goldie, Srinivasan Venkatachary, Hongliang Fei, Ziqiang Feng, Marvin Ritter, Isabel Leal, Sudeep Dasari, Pei Sun, Alif Raditya Rochman, Brendan O'Donoghue, Yuchen Liu, Jim Sproch, Kai Chen, Natalie Clay, Slav Petrov, Sailesh Sidhwani, Ioana Mihailescu, Alex Panagopoulos, AJ Piergiovanni, Yunfei Bai, George Powell, Deep Karkhanis, Trevor Yacovone, Petr Mitrichev, Joe Kovac, Dave Uthus, Amir Yazdanbakhsh, David Amos, Steven Zheng, Bing Zhang, Jin Miao, Bhuvana Ramabhadran, Soroush Radpour, Shantanu Thakoor, Josh Newlan, Oran Lang, Orion Jankowski, Shikhar Bharadwaj, Jean-Michel Sarr, Shereen Ashraf, Sneha Mondal, Jun Yan, Ankit Singh Rawat, Sarmishta Velury, Greg Kochanski, Tom Eccles, Franz Och, Abhanshu Sharma, Ethan Mahintorabi, Alex Gurney, Carrie Muir, Vered Cohen, Saksham Thakur, Adam Bloniarz, Asier Mujika, Alexander Pritzel, Paul Caron, Altaf Rahman, Fiona Lang, Yasumasa Onoe, Petar Sirkovic, Jay Hoover, Ying Jian, Pablo Duque, Arun Narayanan, David Soergel, Alex Haig, Loren Maggiore, Shyamal Buch, Josef Dean, Ilya Figotin, Igor Karpov, Shaleen Gupta, Denny Zhou, Muhuan Huang, Ashwin Vaswani, Christopher Semturs, Kaushik Shivakumar, Yu Watanabe, Vinodh Kumar Rajendran, Eva Lu, Yanhan Hou, Wenting Ye, Shikhar Vashishth, Nana Nti, Vytenis Sakenas, Darren Ni, Doug DeCarlo, Michael Bendersky, Sumit Bagri, Nacho Cano, Elijah Peake, Simon Tokumine, Varun Godbole, Carlos Guía, Tanya Lando, Vittorio Selo, Seher Ellis, Danny Tarlow, Daniel Gillick, Alessandro Epasto, Siddhartha Reddy Jonnalagadda, Meng Wei, Meiyan Xie, Ankur Taly, Michela Paganini, Mukund Sundararajan, Daniel Toyama, Ting Yu, Dessie Petrova, Aneesh Pappu, Rohan Agrawal, Senaka Buthpitiya, Justin Frye, Thomas Buschmann, Remi Crocker, Marco Tagliasacchi, Mengchao Wang, Da Huang, Sagi Perel, Brian Wieder, Hideto Kazawa, Weiyue Wang, Jeremy Cole, Himanshu Gupta, Ben Golan, Seojin Bang, Nitish Kulkarni, Ken Franko, Casper Liu, Doug Reid, Sid Dalmia, Jay Whang, Kevin Cen, Prasha Sundaram, Johan Ferret, Berivan Isik, Lucian Ionita, Guan Sun, Anna Shekhawat, Muqthar Mohammad, Philip Pham, Ronny Huang, Karthik Raman, Xingyi Zhou, Ross Mcilroy, Austin Myers, Sheng Peng, Jacob Scott, Paul Covington, Sofia Erell, Pratik Joshi, João Gabriel Oliveira, Natasha Noy, Tajwar Nasir, Jake Walker, Vera Axelrod, Tim Dozat, Pu Han, Chun-Te Chu, Eugene Weinstein, Anand Shukla, Shreyas Chandrakaladharan, Petra Poklukar, Bonnie Li, Ye Jin, Prem Eruvbetine, Steven Hansen, Avigail Dabush, Alon Jacovi, Samrat Phatale, Chen Zhu, Steven Baker, Mo Shomrat, Yang Xiao, Jean Pouget-Abadie, Mingyang Zhang, Fanny Wei, Yang Song, Helen King, Yiling Huang, Yun Zhu, Ruoxi Sun, Juliana Vicente Franco, Chu-Cheng Lin, Sho Arora, Hui, Li, Vivian Xia, Luke Vilnis, Mariano Schain, Kaiz Alarakyia, Laurel Prince, Aaron Phillips, Caleb Habtegebriel, Luyao Xu, Huan Gui, Santiago Ontanon, Lora Aroyo, Karan Gill, Peggy Lu, Yash Katariya, Dhruv Madeka, Shankar Krishnan, Shubha Srinivas Raghvendra, James Freedman, Yi Tay, Gaurav Menghani, Peter Choy, Nishita Shetty, Dan Abolafia, Doron Kukliansky, Edward Chou, Jared Lichtarge, Ken Burke, Ben Coleman, Dee Guo, Larry Jin, Indro Bhattacharya, Victoria Langston, Yiming Li, Suyog Kotecha, Alex Yakubovich, Xinyun Chen, Petre Petrov, Tolly Powell, Yanzhang He, Corbin Quick, Kanav Garg, Dawsen Hwang, Yang Lu, Srinadh Bhojanapalli, Kristian Kjems, Ramin Mehran, Aaron Archer, Hado van Hasselt, Ashwin Balakrishna, JK Kearns, Meiqi Guo, Jason Riesa, Mikita Sazanovich, Xu Gao, Chris Sauer, Chengrun Yang, XiangHai Sheng, Thomas Jimma, Wouter Van Gansbeke, Vitaly Nikolaev, Wei Wei, Katie Millican, Ruizhe Zhao, Justin Snyder, Levent Bolelli, Maura O'Brien, Shawn Xu, Fei Xia, Wentao Yuan, Arvind Neelakantan, David Barker, Sachin Yadav, Hannah Kirkwood, Farooq Ahmad, Joel Wee, Jordan Grimstad, Boyu Wang, Matthew Wiethoff, Shane Settle, Miaosen Wang, Charles Blundell, Jingjing Chen, Chris Duvarney, Grace Hu, Olaf Ronneberger, Alex Lee, Yuanzhen Li, Abhishek Chakladar, Alena Butryna, Georgios Evangelopoulos, Guillaume Desjardins, Jonni Kanerva, Henry Wang, Averi Nowak, Nick Li, Alyssa Loo, Art Khurshudov, Laurent El Shafey, Nagabhushan Baddi, Karel Lenc, Yasaman Razeghi, Tom Lieber, Amer Sinha, Xiao Ma, Yao Su, James Huang, Asahi Ushio, Hanna Klimczak-Pluciska, Kareem Mohamed, JD Chen,

Simon Osindero, Stav Ginzburg, Lampros Lamprou, Vasilisa Bashlovkina, Duc-Hieu Tran, Ali Khodaei, Ankit Anand, Yixian Di, Ramy Eskander, Manish Reddy Vuyyuru, Jasmine Liu, Aishwarya Kamath, Roman Goldenberg, Mathias Bellaiche, Juliette Pluto, Bill Rosgen, Hassan Mansoor, William Wong, Suhas Ganesh, Eric Bailey, Scott Baird, Dan Deutsch, Jinoo Baek, Xuhui Jia, Chansoo Lee, Abe Friesen, Nathaniel Braun, Kate Lee, Amayika Panda, Steven M. Hernandez, Duncan Williams, Jianqiao Liu, Ethan Liang, Arnaud Autef, Emily Pitler, Deepali Jain, Phoebe Kirk, Oskar Bunyan, Jaume Sanchez Elias, Tongxin Yin, Machel Reid, Aedan Pope, Nikita Putikhin, Bidisha Samanta, Sergio Guadarrama, Dahun Kim, Simon Rowe, Marcella Valentine, Geng Yan, Alex Salcianu, David Silver, Gan Song, Richa Singh, Shuai Ye, Hannah DeBalsi, Majd Al Merey, Eran Ofek, Albert Webson, Shibl Mourad, Ashwin Kakarla, Silvio Lattanzi, Nick Roy, Evgeny Sluzhaev, Christina Butterfield, Alessio Tonioni, Nathan Waters, Sudhindra Kopalle, Jason Chase, James Cohan, Girish Ramchandra Rao, Robert Berry, Michael Voznesensky, Shuguang Hu, Kristen Chiafullo, Sharat Chikkerur, George Scrivener, Ivy Zheng, Jeremy Wiesner, Wolfgang Macherey, Timothy Lillicrap, Fei Liu, Brian Walker, David Welling, Elinor Davies, Yangsibo Huang, Lijie Ren, Nir Shabat, Alessandro Agostini, Mariko Iinuma, Dustin Zelle, Rohit Sathyanarayana, Andrea D'olimpio, Morgan Redshaw, Matt Ginsberg, Ashwin Murthy, Mark Geller, Tatiana Matejovicova, Ayan Chakrabarti, Ryan Julian, Christine Chan, Qiong Hu, Daniel Jarrett, Manu Agarwal, Jeshwanth Challagundla, Tao Li, Sandeep Tata, Wen Ding, Maya Meng, Zhuyun Dai, Giulia Vezzani, Shefali Garg, Jannis Bulian, Mary Jasarevic, Honglong Cai, Harish Rajamani, Adam Santoro, Florian Hartmann, Chen Liang, Bartek Perz, Apoorv Jindal, Fan Bu, Sungyong Seo, Ryan Poplin, Adrian Goedeckemeyer, Badih Ghazi, Nikhil Khadke, Leon Liu, Kevin Mather, Mingda Zhang, Ali Shah, Alex Chen, Jinliang Wei, Keshav Shivam, Yuan Cao, Donghyun Cho, Angelo Scorza Scarpati, Michael Moffitt, Clara Barbu, Ivan Jurin, Ming-Wei Chang, Hongbin Liu, Hao Zheng, Shachi Dave, Christine Kaeser-Chen, Xiaobin Yu, Alvin Abdagic, Lucas Gonzalez, Yanping Huang, Peilin Zhong, Cordelia Schmid, Bryce Petrini, Alex Wertheim, Jifan Zhu, Hoang Nguyen, Kaiyang Ji, Yanqi Zhou, Tao Zhou, Fangxiaoyu Feng, Regev Cohen, David Rim, Shubham Milind Phal, Petko Georgiev, Ariel Brand, Yue Ma, Wei Li, Somit Gupta, Chao Wang, Pavel Dubov, Jean Tarbouriech, Kingshuk Majumder, Huijian Li, Norman Rink, Apurv Suman, Yang Guo, Yinghao Sun, Arun Nair, Xiaowei Xu, Mohamed Elhawaty, Rodrigo Cabrera, Guangxing Han, Julian Eisenschlos, Junwen Bai, Yuqi Li, Yamini Bansal, Thibault Sellam, Mina Khan, Hung Nguyen, Justin Mao-Jones, Nikos Parotsidis, Jake Marcus, Cindy Fan, Roland Zimmermann, Yony Kochinski, Laura Graesser, Feryal Behbahani, Alvaro Caceres, Michael Riley, Patrick Kane, Sandra Lefdal, Rob Willoughby, Paul Vicol, Lun Wang, Shujian Zhang, Ashleah Gill, Yu Liang, Gautam Prasad, Soroosh Mariooryad, Mehran Kazemi, Zifeng Wang, Kritika Muralidharan, Paul Voigtlaender, Jeffrey Zhao, Huanjie Zhou, Nina D'Souza, Aditi Mavalankar, Séb Arnold, Nick Young, Obaid Sarvana, Chace Lee, Milad Nasr, Tingting Zou, Seokhwan Kim, Lukas Haas, Kaushal Patel, Neslihan Bulut, David Parkinson, Courtney Biles, Dmitry Kalashnikov, Chi Ming To, Aviral Kumar, Jessica Austin, Alex Greve, Lei Zhang, Megha Goel, Yeqing Li, Sergey Yaroshenko, Max Chang, Abhishek Jindal, Geoff Clark, Hagai Taitelbaum, Dale Johnson, Ofir Roval, Jeongwoo Ko, Anhad Mohananey, Christian Schuler, Shenil Dodhia, Ruichao Li, Kazuki Osawa, Claire Cui, Peng Xu, Rushin Shah, Tao Huang, Ela Gruzewska, Nathan Clement, Mudit Verma, Olcan Sercinoglu, Hai Qian, Viral Shah, Masa Yamaguchi, Abhinit Modi, Takahiro Kosakai, Thomas Strohmann, Junhao Zeng, Beliz Gunel, Jun Qian, Austin Tarango, Krzysztof Jastrzbski, Robert David, Jyn Shan, Parker Schuh, Kunal Lad, Willi Gierke, Mukundan Madhavan, Xinyi Chen, Mark Kurzeja, Rebeca Santamaria-Fernandez, Dawn Chen, Alexandra Cordell, Yuri Chervonyi, Frankie Garcia, Nithish Kannen, Vincent Perot, Nan Ding, Shlomi Cohen-Ganor, Victor Lavrenko, Junru Wu, Georgie Evans, Cicero Nogueira dos Santos, Madhavi Sewak, Ashley Brown, Andrew Hard, Joan Puigcerver, Zeyu Zheng, Yizhong Liang, Evgeny Gladchenko, Reeve Ingle, Uri First, Pierre Sermanet, Charlotte Magister, Mihajlo Velimirovi, Sashank Reddi, Susanna Ricco, Eirikur Agustsson, Hartwig Adam, Nir Levine, David Gaddy, Dan Holtmann-Rice, Xuanhui Wang, Ashutosh Sathe, Abhijit Guha Roy, Bla Bratani, Alen Carin, Harsh Mehta, Silvano Bonacina, Nicola De Cao, Mara Finkelstein, Verena Rieser, Xinyi Wu, Florent Altché, Dylan Scandinaro, Li Li, Nino Vieillard, Nikhil Sethi, Garrett Tanzer, Zhi Xing, Shibo Wang, Parul Bhatia, Gui Citovsky, Thomas Anthony, Sharon Lin, Tianze Shi, Shoshana Jakobovits, Gena Gibson, Raj Apte, Lisa Lee, Mingqing Chen, Arunkumar Byravan, Petros Maniatis, Kellie Webster, Andrew Dai, Pu-Chin Chen, Jiaqi Pan, Asya Fadeeva, Zach Gleicher, Thang Luong, and Niket Kumar Bhumihar. Gemini 2.5: Pushing the frontier with advanced reasoning, multimodality, long context, and next generation agentic capabilities, 2025.

[29] Wolfram Alpha LLC. Wolfram|alpha. `https://www.wolframalpha.com/`, 2009. Accessed: 2025-08-12.

[30] Alec Radford, Jeff Wu, Rewon Child, David Luan, Dario Amodei, and Ilya Sutskever. Language models are unsupervised multitask learners. 2019.

[31] Hongru Wang, Cheng Qian, Manling Li, Jiahao Qiu, Boyang Xue, Mengdi Wang, Heng Ji, and Kam-Fai Wong. Toward a theory of agents as tool-use decision-makers, 2025.

[32] Qingyun Wu, Gagan Bansal, Jieyu Zhang, Yiran Wu, Beibin Li, Erkang Zhu, Li Jiang, Xiaoyun Zhang, Shaokun Zhang, Jiale Liu, Ahmed Hassan Awadallah, Ryen W White, Doug Burger, and Chi Wang. Autogen: Enabling next-gen llm applications via multi-agent conversation, 2023.

[33] OpenAI. Introducing deep research.

[34] Manus Team. Manus, 2024.

[35] Hongru Wang, Cheng Qian, Wanjun Zhong, Xiusi Chen, Jiahao Qiu, Shijue Huang, Bowen Jin, Mengdi Wang, Kam-Fai Wong, and Heng Ji. Acting less is reasoning more! teaching model to act efficiently, 2025.

[36] Guanting Dong, Hangyu Mao, Kai Ma, Licheng Bao, Yifei Chen, Zhongyuan Wang, Zhongxia Chen, Jiazhen Du, Huiyang Wang, Fuzheng Zhang, Guorui Zhou, Yutao Zhu, Ji-Rong Wen, and Zhicheng Dou. Agentic reinforced policy optimization, 2025.

[37] Xufang Luo, Yuge Zhang, Zhiyuan He, Zilong Wang, Siyun Zhao, Dongsheng Li, Luna K. Qiu, and Yuqing Yang. Agent lightning: Train any ai agents with reinforcement learning, 2025.

[38] AlphaProof and AlphaGeometry teams. Ai achieves silver-medal standard solving international mathematical olympiad problems. *Google DeepMind Blog*, July 2024. Published July 25, 2024.

[39] Jiahao Qiu, Yinghui He, Xinzhe Juan, Yimin Wang, Yuhan Liu, Zixin Yao, Yue Wu, Xun Jiang, Ling Yang, and Mengdi Wang. Emoagent: Assessing and safeguarding human-ai interaction for mental health safety, 2025.

[40] Luoxin Chen, Jinming Gu, Liankai Huang, Wenhao Huang, Zhicheng Jiang, Allan Jie, Xiaoran Jin, Xing Jin, Chenggang Li, Kaijing Ma, Cheng Ren, Jiawei Shen, Wenlei Shi, Tong Sun, He Sun, Jiahui Wang, Siran Wang, Zhihong Wang, Chenrui Wei, Shufa Wei, Yonghui Wu, Yuchen Wu, Yihang Xia, Huajian Xin, Fan Yang, Huaiyuan Ying, Hongyi Yuan, Zheng Yuan, Tianyang Zhan, Chi Zhang, Yue Zhang, Ge Zhang, Tianyun Zhao, Jianqiu Zhao, Yichi Zhou, and Thomas Hanwen Zhu. Seed-prover: Deep and broad reasoning for automated theorem proving, 2025.

[41] Kaiyue Feng, Yilun Zhao, Yixin Liu, Tianyu Yang, Chen Zhao, John Sous, and Arman Cohan. Physics: Benchmarking foundation models on university-level physics problem solving, 2025.

[42] Song Dai, Yibo Yan, Jiamin Su, Dongfang Zihao, Yubo Gao, Yonghua Hei, Jungang Li, Junyan Zhang, Sicheng Tao, Zhuoran Gao, and Xuming Hu. Physicsarena: The first multimodal physics reasoning benchmark exploring variable, process, and solution dimensions, 2025.

[43] Xinyu Zhang, Yuxuan Dong, Yanrui Wu, Jiaxing Huang, Chengyou Jia, Basura Fernando, Mike Zheng Shou, Lingling Zhang, and Jun Liu. Physreason: A comprehensive benchmark towards physics-based reasoning, 2025.

[44] Isaac Newton. *Philosophiæ Naturalis Principia Mathematica*. Jussu Societatis Regiæ ac Typis Josephi Streater, 1687.

[45] Shiekh Zia Uddin, Sachin Vaidya, Shrish Choudhary, Zhuo Chen, Raafat K. Salib, Luke Huang, Dirk R. Englund, and Marin Soljai. Ai-driven robotics for free-space optics, 2025.

[46] Mario Carneiro. Lean4lean: Towards a verified typechecker for lean, in lean, 2024.

[47] Peiyang Song, Kaiyu Yang, and Anima Anandkumar. Lean copilot: Large language models as copilots for theorem proving in lean, 2025.

[48] Maxwell P. Bobbin, Samiha Sharlin, Parivash Feyzishendi, An Hong Dang, Catherine M. Wraback, and Tyler R. Josephson. Formalizing chemical physics using the lean theorem prover, 2023.

[49] P. Smolensky. Connectionist ai, symbolic ai, and the brain. *Artificial Intelligence Review*, 1(3):95–109, 1987.

[50] Leonardo de Moura and Sebastian Ullrich. The lean 4 theorem prover and programming language (system description). In *Automated Deduction CADE 28*, pages 625–635. Springer, Cham, 2021.

[51] Z. Z. Ren, Zhihong Shao, Junxiao Song, Huajian Xin, Haocheng Wang, Wanjia Zhao, Liyue Zhang, Zhe Fu, Qihao Zhu, Dejian Yang, Z. F. Wu, Zhibin Gou, Shirong Ma, Hongxuan Tang, Yuxuan Liu, Wenjun Gao, Daya Guo, and Chong Ruan. Deepseek-prover-v2: Advancing formal mathematical reasoning via reinforcement learning for subgoal decomposition, 2025.

[52] Numina & Kimi Team. Kimina-prover preview: Towards large formal reasoning models with reinforcement learning, 2025.

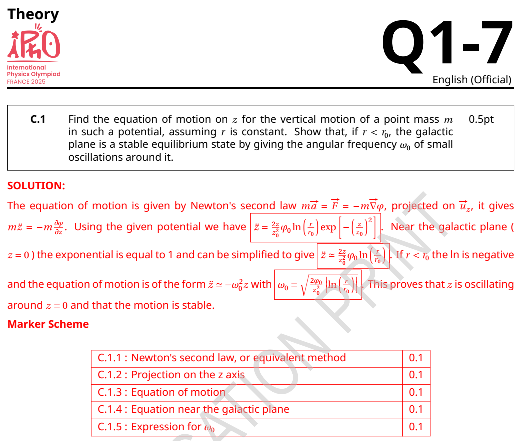
[53] Yichi Zhou, Jianqiu Zhao, Yongxin Zhang, Bohan Wang, Siran Wang, Luoxin Chen, Jiahui Wang, Haowei Chen, Allan Jie, Xinbo Zhang, Haocheng Wang, Trung Luong, Rong Ye, Phan Nhat Hoang, Huishuai Zhang, Peng Sun, and Hang Li. Solving formal math problems by decomposition and iterative reflection, 2025.

[54] Kaito Baba, Chaoran Liu, Shuhei Kurita, and Akiyoshi Sannai. Prover agent: An agent-based framework for formal mathematical proofs, 2025.

[55] Azim Ospanov, Farzan Farnia, and Roozbeh Yousefzadeh. Apollo: Automated llm and lean collaboration for advanced formal reasoning, 2025.

A Examples of IPhO 2025 Problems Scoring Criteria

We provide two example of the IPhO 2025 scoring criteria (corresponding to Theory Problem 1 Part C.1 and Theory Problem 3 Part C.2, correspondingly) in Figure 2, obtained from `https://ipho.olimpicos.net/`.



Theory

International Physics Olympiad FRANCE 2025

Q1-7

English (Official)

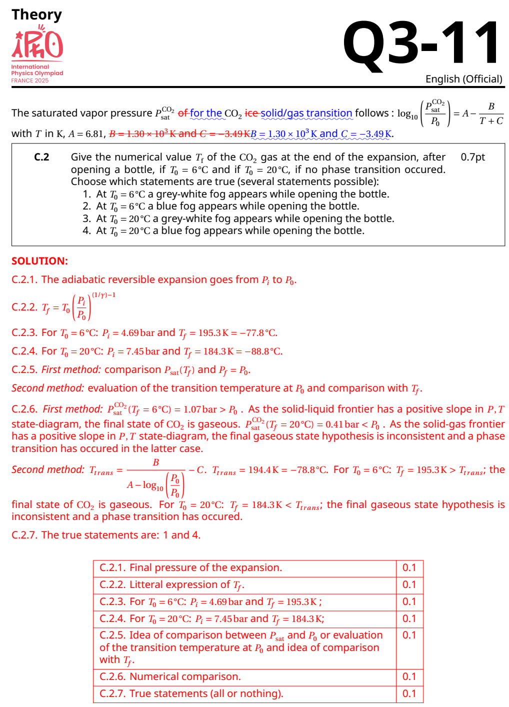
C.1	Find the equation of motion on z for the vertical motion of a point mass m in such a potential, assuming r is constant. Show that, if $r < r_0$, the galactic plane is a stable equilibrium state by giving the angular frequency ω_0 of small oscillations around it.	0.5pt

SOLUTION:

The equation of motion is given by Newton's second law $m\vec{a} = \vec{F} = -m\vec{\nabla}\varphi$, projected on $\vec{u}_z$, it gives $m\ddot{z} = -m\frac{\partial\varphi}{\partial z}$. Using the given potential we have $\boxed{\ddot{z} = \frac{2z}{z_0^2}\varphi_0 \ln\left(\frac{r}{r_0}\right)\exp\left[-\left(\frac{z}{z_0}\right)^2\right]}$. Near the galactic plane ($z = 0$) the exponential is equal to 1 and can be simplified to give $\boxed{\ddot{z} \simeq \frac{2z}{z_0^2}\varphi_0 \ln\left(\frac{r}{r_0}\right)}$. If $r < r_0$ the ln is negative and the equation of motion is of the form $\ddot{z} \simeq -\omega_0^2 z$ with $\boxed{\omega_0 = \sqrt{\frac{2\varphi_0}{z_0^2}\left|\ln\left(\frac{r}{r_0}\right)\right|}}$. This proves that z is oscillating around $z = 0$ and that the motion is stable.

Marker Scheme

C.1.1 : Newton's second law, or equivalent method	0.1
C.1.2 : Projection on the z axis	0.1
C.1.3 : Equation of motion	0.1
C.1.4 : Equation near the galactic plane	0.1
C.1.5 : Expression for ω_0	0.1



Theory

International Physics Olympiad FRANCE 2025

Q3-11

English (Official)

The saturated vapor pressure $P_{sat}^{CO_2}$ ~~of~~ for the CO_2 ~~ice~~ solid/gas transition follows : $\log_{10}\left(\frac{P_{sat}^{CO_2}}{P_0}\right) = A - \frac{B}{T+C}$ with T in K, $A = 6.81$, ~~$B = 1.30 \times 10^3$ K and $C = -3.49$ K~~ $B = 1.30 \times 10^3$ K and $C = -3.49$ K.

C.2	Give the numerical value T_f of the CO_2 gas at the end of the expansion, after opening a bottle, if $T_0 = 6$°C and if $T_0 = 20$°C, if no phase transition occured. Choose which statements are true (several statements possible): 1. At $T_0 = 6$°C a grey-white fog appears while opening the bottle. 2. At $T_0 = 6$°C a blue fog appears while opening the bottle. 3. At $T_0 = 20$°C a grey-white fog appears while opening the bottle. 4. At $T_0 = 20$°C a blue fog appears while opening the bottle.	0.7pt

SOLUTION:

C.2.1. The adiabatic reversible expansion goes from P_i to P_0.

C.2.2. $T_f = T_0\left(\frac{P_i}{P_0}\right)^{(1/\gamma)-1}$

C.2.3. For $T_0 = 6$°C: $P_i = 4.69$ bar and $T_f = 195.3$ K $= -77.8$°C.

C.2.4. For $T_0 = 20$°C: $P_i = 7.45$ bar and $T_f = 184.3$ K $= -88.8$°C.

C.2.5. *First method:* comparison $P_{sat}(T_f)$ and $P_f = P_0$.

Second method: evaluation of the transition temperature at P_0 and comparison with T_f.

C.2.6. *First method:* $P_{sat}^{CO_2}(T_f = 6°C) = 1.07\,\text{bar} > P_0$. As the solid-liquid frontier has a positive slope in P, T state-diagram, the final state of CO_2 is gaseous. $P_{sat}^{CO_2}(T_f = 20°C) = 0.41\,\text{bar} < P_0$. As the solid-gas frontier has a positive slope in P, T state-diagram, the final gaseous state hypothesis is inconsistent and a phase transition has occured in the latter case.

Second method: $T_{trans} = \frac{B}{A - \log_{10}\left(\frac{P_0}{P_0}\right)} - C$. $T_{trans} = 194.4$ K $= -78.8$°C. For $T_0 = 6$°C: $T_f = 195.3\,\text{K} > T_{trans}$; the final state of CO_2 is gaseous. For $T_0 = 20$°C: $T_f = 184.3\,\text{K} < T_{trans}$; the final gaseous state hypothesis is inconsistent and a phase transition has occured.

C.2.7. The true statements are: 1 and 4.

C.2.1. Final pressure of the expansion.	0.1
C.2.2. Litteral expression of T_f.	0.1
C.2.3. For $T_0 = 6$°C: $P_i = 4.69$ bar and $T_f = 195.3$ K ;	0.1
C.2.4. For $T_0 = 20$°C: $P_i = 7.45$ bar and $T_f = 184.3$ K;	0.1
C.2.5. Idea of comparison between P_{sat} and P_0 or evaluation of the transition temperature at P_0 and idea of comparison with T_f.	0.1
C.2.6. Numerical comparison.	0.1
C.2.7. True statements (all or nothing).	0.1

Figure 2: Two scoring criteria examples, for IPhO 2025 Theory Problem 1 Part C.1 and Theory Problem 3 Part C.2, correspondingly. As shown, there are 5 scoring criteria for Theory Problem 1 Part C.1, and 7 scoring criteria for Theory Problem 3 Part C.2.

As shown in Figure 2, for each problem, very detailed answers and scoring criteria are provided, making fine-grained scoring possible for answers. In Table 1 we count the number of scoring criteria for each part of each Theory Problem.

B Detailed Prompts

B.1 Image Analyzer Tool Prompt

The image analyzer tool utilizes an LLM provided with the image and question from the manager agent. Its task is to answer manager's questions based on provided information. It then returns an str-object of its measurements.

```
# input: manager_query: str, img_file: ImageFile, vision_expert_llm:
    LLMModel
IMG_SYSTEM_PROMPT = "You are an expert in dealing with image in
    Physics Olympiads."
messages = [
     ChatMessage(role=MessageRole.SYSTEM, content=IMG_SYSTEM_PROMPT),
     ChatMessage(role=MessageRole.USER, content=[
         {"type": "image", "image": img_file},
         {"type": "text", "text": question},
     ]),
]
output: str = vision_expert_llm.generate(messages)
```

B.2 Answer Reviewer Tool Prompt

The Answer Reviewer tool utilizes an LLM provided with: (1) manager agent's solution; (2) manager agent's notes; (3) the original problems (including texts and figures). It then returns an str-object representing its review results.

```
# input: agent_solution:str, agent_note:str, markdown_content:
    List[Dict[str, Any], review_expert_llm: LLMModel
# markdown_content includes markdown file text and image.
REVIEW_SYSTEM_PROMPT = (
    "You are an uncompromising Physics peer-reviewer. Your job is to
        find *every* logical, mathematical
error in the worker's answer. "
    "Check dimensional consistency, missing steps, incorrect sign
        conventions, numerical mistakes, and
unclear explanations. Focus especially on wrong answers, less on
    presentations."
    "Be extremely critical: if something is wrong, point it out and
        request clarification or correction.
Mainly focus on errors that would lead to a wrong result, rather than
    focusing extremely on presentation
or style."
    "It is possible that the worker's answer is not correct, so
        please be prepared to provide detailed
feedback. The worker's answer contains some error, so you must check
    and point it out. Also, if the
worker reads measurements from image, make sure to remind the worker
    that whenever it reads or measures
from image, it uses the ask_image_expert tool, or the readings might
    be very inaccurate.\n"
)

review_instruction = (
    f"Please review the following solution:\n\n"
    f"WORKER'S SOLUTION:\n{agent_solution}\n\n"
    f"WORKER'S NOTE: {agent_note}\n\n"
    f"Please provide detailed feedback on correctness. "
    f"Point out any errors, wrong steps, focus more on correctness
        rather than presentation."
    f"The original problem follows:"
)

combined_content : List[Dict[str, Any]] = [
    {"type": "text", "text": review_instruction}
] + markdown_content

messages = [
    ChatMessage(role=MessageRole.SYSTEM,
        content=REVIEW_SYSTEM_PROMPT),
    ChatMessage(role=MessageRole.USER, content=combined_content),
]
output: str = review_expert_llm.generate(messages)
```

C Analysis

C.1 Ablation Study

To study the impact of each tool on the final scores, we replicate the protocol described in Section 3, providing different tools for the agent system. We consider four settings: (1) Physics Supernova; (2) without *ImageAnalyzer* tool; (3) without *AnswerReviewer* tool; and (4) LLM only. For each theory problem, we perform 5 independent runs and report the mean and standard deviation. The agents are powered by Gemini 2.5 Pro. The results are shown in Table 3. Comparing Table 1 and Table3, we find:

Method	Theory 1	Theory 2	Theory 3	Theory Part
Total score	10.0	10.0	10.0	30.0
Physics Supernova	$\mathbf{9.02 \pm 0.11}$	$\mathbf{6.08 \pm 0.77}$	$\mathbf{8.40 \pm 0.27}$	$\mathbf{23.5 \pm 0.8}$
w.o.ImgTool	8.58 ± 0.19	5.98 ± 0.54	8.26 ± 0.24	22.8 ± 0.6
w.o.RevTool	8.62 ± 0.16	5.54 ± 0.61	8.26 ± 0.15	22.4 ± 0.6
LLM Only	8.46 ± 0.16	5.30 ± 0.99	7.66 ± 0.51	21.4 ± 1.1

Table 3: Ablation study results for four experiment settings: Physics Supernova, Physics Supernova without ***AnswerReviewer*** (RevTool), Physics Supernova without ***ImageAnalyzer*** (ImgTool), and using LLM only.

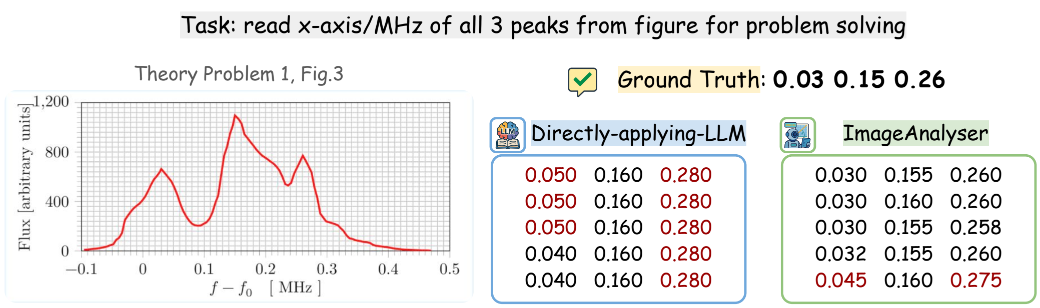


Figure 3: Effect of *ImageAnalyzer* on Theory Problem 1 Part C. The problem here requires accurate measurement from a figure (shown left). We show 5 repeated experiments of directly applying LLMs and using Image Analyzer Tool (shown right), with reading difference > 0.01MHZ colored in red. As shown, the improvement mainly comes from a reduction in measurement error.

ImageAnalyzer **helps with tasks requiring Accurate Image Measurements.** For example, as shown in Table 1, the theory problem requires accurate measurements from figures, and in this case, the Image Analyzer successfully improves the performance of the model. This indicates that delegating high-accuracy image analysis to *ImageAnalyzer* improves the overall score. We present a case study in Section C.2 for better illustration.

AnswerReviewer **raises overall scores via post-hoc review.** In most problems (especially non-easy ones), removing the Answer Reviewer reduces performance. For example, we see a performance drop for overall scores for all three theory problems after removing the Answer Reviewer. This implies that, equipped with *AnswerReviewer* for reviewing, locating errors and providing feedback, Physics Supernova can improve performance for many cases.

C.2 Case Study: Image Analyzer Tool

We zoom in on Theory Problem 1 Part C to see how Image Analyzer Tool helps improve scores. In particular, this problem requires the contestants to accurately read a figure to solve the problem. The specific task related to the image is shown in Figure 3.

As shown in Figure 3, using a specialized Image Reading LLM as Image Analyzer Tool excels directly exposing the manager agent to all images of a single problem, which reduces the mean absolute error (MAE) from 0.015 to 0.004. This example further shows that, by using Image Analyzer Tool, Physics Supernova improves its ability to accurately read from figures, thus improving its overall performance.

C.3 Domain-Specific Tools: WolframAlpha QA Tool

Physics research usually requires expert domain knowledge. In Physics Olympiads, some reference information (e.g., physical constants, specific physics formulas, etc.) is provided as information for

Problem Q4. Using the IAPWS-IF97 formulation for water/steam, compute the specific enthalpy of water at p = 15 MPa and T = 650 K (single-phase state as appropriate). Return a single number: the value in $kJkg^{-1}$, rounded to exactly 5 significant digits, in scientific notation. Do not include units or extra text.

GT Answer: 2.8686E+3 **w.o.wolfTool Answer:** 3.0462E+3 **w.wolfTool Answer:** 2.8690E+3

Problem Q5. Using NIST X-ray transition energies (or equivalent), determine the photon energy of the copper $\kappa\alpha_1$ (KL_3) line for elemental Cu at ambient conditions. Return a single number: the value in keV, rounded to exactly 5 significant digits, in scientific notation. Do not include units or extra text.

GT Answer: 8.0478E+0 **w.o.wolfTool Answer:** 8.0463E+0 **w.wolfTool Answer:** 8.0478E+0

Figure 4: Examples of 10 expert-knowledge-requiring problems we generate, listed in Appendix D. This example requires expert domain knowledge that can only be obtained through web search/expert database queries.

	w.o. WolframAlpha Tool	**w. WolframAlpha Tool**
# **3-digit Accurate Answers**	3/10	9/10
# **4-digit Accurate Answers**	2/10	6/10

Table 4: Number of accurate answers for expert-knowledge-requiring tasks shown in Appendix D. ('N-digit accurate' means that the answer differs from GT only by ± 1 on the $N-$th significant digit. For example, in Figure 4 Q4, the 'w.o.wolfTool Answer' is 1-digit accurate, while the 'w.wolfTool Answer' is 4-digit accurate, compared to the ground truth answer.)

contestants to use: this is not the case for more complicated Physics researches. In these expert-level cases, accessing domain knowledge is very important. Figure 4 provides examples of this case.

To better handle domain-specific queries in Physics Supernova, we further equip Physics Supernova with a question-answering (QA) tool for expert domain knowledge. To be specific, we utilize WolframAlpha [29]: a computational knowledge engine capable of providing accurate and concise results for science-related queries.

In order to study Physics Supernova's performance on more expert tasks with WolframAlpha QA Tool, we generate 10 problems which require expert knowledge (most efficiently obtained from standard references or WolframAlpha), listed in Appendix D. As shown in Figure 4 are examples of these problems. Table 4 reports the performance with and without the WolframAlpha tool. The result shows that WolframAlpha Tool improves the ability of Physics Supernova to solve problems that require expert physics knowledge.

In all: experiments presented in this Section C show that, by integrating appropriate tools into the AI agent system, we can significantly advance the capabilities of LLMs in solving complex and challenging physics problems. This enhancement improves the mastery of physics of the agent system without requiring any modifications to the underlying LLMs. In conclusion, our Physics Supernova emerges as a powerful, flexible, and extensible physics problem-solving framework built upon the agent-based paradigm.

D Expert-knowledge Requiring Tasks

We further generate several tasks requiring expert knowledge to test how integrating WolframAlpha would help Physics Supernova with accurate expert knowledge.

As shown in the following examples, **When given access to WolframAlpha Tools, the agent system provides more accurate answers.** In the experiments, we use Gemini 2.5 Pro as LLM and compare the result with and without WolframAlpha Tools, as shown below. The aggregated results are also shown in Table 4.

Problem Q1. Using the latest AME (Atomic Mass Evaluation) atomic masses, compute the Q-value of double beta decay $^{76}Ge \rightarrow^{76} Se + 2e^{-}$ (ground state ground state). Return a single number: the value in keV, rounded to exactly 5 significant digits, in scientific notation. Do not include units or extra text.

GT Answer: 2.0391E+3 **w.o.wolfTool Answer:** 2.0391E+3 **w.wolfTool Answer:** 2.0390E+3

Problem Q2. Using NIST XCOM (or an equivalent authoritative database), determine the mass attenuation coefficient μ/ρ of lead (Pb) for photons of energy 662.0 keV ($^{137}Cs\ \gamma$ line). Return a single number: the value in $cm^2 g^{-1}$, rounded to exactly 5 significant digits, in scientific notation. Do not include units or extra text.

GT Answer: 1.1105E-1 **w.o.wolfTool Answer:** 1.1352E-1 **w.wolfTool Answer:** 1.1150E-1

Problem Q3. Using the Ciddor (1996) refractive-index model for air, at wavelength $\lambda = 633nm$ (vacuum), P = 101325 Pa, T = 20 řC, RH = 50 , and CO_2 = 450 ppm, compute $n - 1$. Return a single number: the value (dimensionless), rounded to exactly 5 significant digits, in scientific notation. Do not include units or extra text.

GT Answer: 2.7132E-4 **w.o.wolfTool Answer:** 2.6894E-4 **w.wolfTool Answer:** 2.7139E-4

Problem Q4. Using the IAPWS-IF97 formulation for water/steam, compute the specific enthalpy of water at p = 15 MPa and T = 650 K (single-phase state as appropriate). Return a single number: the value in $kJkg^{-1}$, rounded to exactly 5 significant digits, in scientific notation. Do not include units or extra text.

GT Answer: 2.8686E+3 **w.o.wolfTool Answer:** 3.0462E+3 **w.wolfTool Answer:** 2.8690E+3

Problem Q5. Using NIST X-ray transition energies (or equivalent), determine the photon energy of the copper $\kappa\alpha_1\ (KL_3)$ line for elemental Cu at ambient conditions. Return a single number: the value in keV, rounded to exactly 5 significant digits, in scientific notation. Do not include units or extra text.

GT Answer: 8.0478E+0 **w.o.wolfTool Answer:** 8.0463E+0 **w.wolfTool Answer:** 8.0478E+0

Problem Q6. Using the IGRF 13th generation (epoch 2025.0), compute the total geomagnetic field magnitude at (40.0140ř N, 105.2705ř W, altitude 1624 m) on 2025-01-01 00:00 UTC. Return a single number: the value in nT, rounded to exactly 5 significant digits, in scientific notation. Do not include units or extra text.

GT Answer: 5.1321E+4 **w.o.wolfTool Answer:** 4.9726E+4 **w.wolfTool Answer:** 5.1300E+4

Problem Q7. Using CODATA-2022 fundamental constants, compute the rest frequency of the neutral hydrogen 21 cm hyperfine transition (ground-state spin-flip). Return a single number: the value in Hz, rounded to exactly 5 significant digits, in scientific notation. Do not include units or extra text.

GT Answer: 1.4204E+9 **w.o.wolfTool Answer:** 1.4228E+9 **w.wolfTool Answer:** 1.4204E+9

Problem Q8. Using the NIST ESTAR database (or equivalent), determine the mass stopping power of aluminum (Al) for electrons of kinetic energy 1.000 MeV. Return a single number: the value in $MeVcm^2g^{-1}$, rounded to exactly 5 significant digits, in scientific notation. Do not include units or extra text.

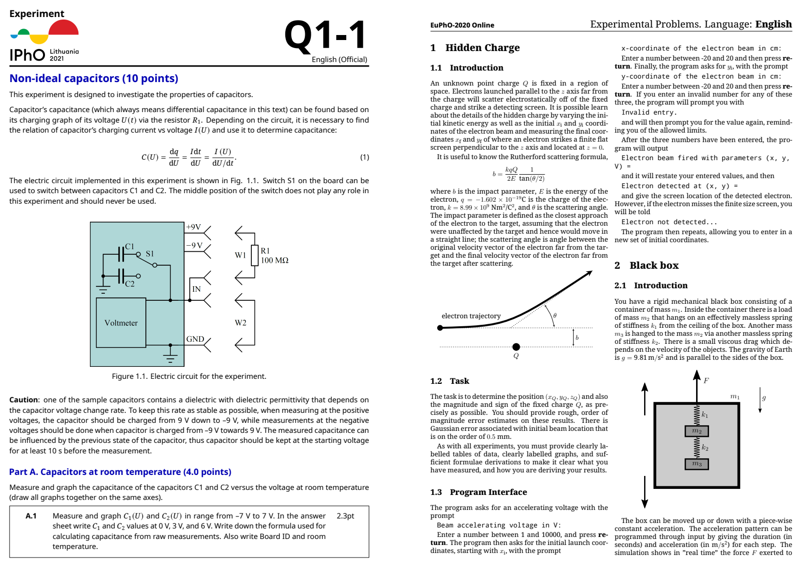
GT Answer: 1.4860E+0 **w.o.wolfTool Answer:** 1.5980E+0 **w.wolfTool Answer:** 1.5980E+0

Problem Q9. Using JANAF/NIST thermochemical data (ideal-gas heat capacities), determine the molar heat capacity at constant pressure, C_p, of nitrogen gas (N_2) at T = 1200 K (assume thermally perfect ideal gas, no dissociation). Return a single number: the value in $Jmol^{-1}K^{-1}$, rounded to exactly 5 significant digits, in scientific notation. Do not include units or extra text.

GT Answer: 3.3723E+1 **w.o.wolfTool Answer:** 3.3540E+1 **w.wolfTool Answer:** 3.3724E+1

Problem Q10. For Beijing, China (39.9042ř N, 116.4074ř E, elevation 50 m), determine the umbral magnitude of the next lunar eclipse after 2025-08-09 that is at least partially visible from that location. Return a single number: the umbral magnitude (dimensionless), rounded to exactly 5 significant digits, in scientific notation. Do not include units or extra text.

GT Answer: 1.3638E+0 **w.o.wolfTool Answer:** 1.1510E+0 **w.wolfTool Answer:** 1.3680E+0



Experiment

IPhO Lithuania 2021

Q1-1

English (Official)

Non-ideal capacitors (10 points)

This experiment is designed to investigate the properties of capacitors.

Capacitor's capacitance (which always means differential capacitance in this text) can be found based on its charging graph of its voltage $U(t)$ via the resistor R_1. Depending on the circuit, it is necessary to find the relation of capacitor's charging current vs voltage $I(U)$ and use it to determine capacitance:

$$C(U) = \frac{dq}{dU} = \frac{I dt}{dU} = \frac{I(U)}{dU/dt}. \quad (1)$$

The electric circuit implemented in this experiment is shown in Fig. 1.1. Switch S1 on the board can be used to switch between capacitors C1 and C2. The middle position of the switch does not play any role in this experiment and should never be used.

Figure 1.1. Electric circuit for the experiment.

Caution: one of the sample capacitors contains a dielectric with dielectric permittivity that depends on the capacitor voltage change rate. To keep this rate as stable as possible, when measuring at the positive voltages, the capacitor should be charged from 9 V down to –9 V, while measurements at the negative voltages should be done when capacitor is charged from –9 V towards 9 V. The measured capacitance can be influenced by the previous state of the capacitor, thus capacitor should be kept at the starting voltage for at least 10 s before the measurement.

Part A. Capacitors at room temperature (4.0 points)

Measure and graph the capacitance of the capacitors C1 and C2 versus the voltage at room temperature (draw all graphs together on the same axes).

A.1	Measure and graph $C_1(U)$ and $C_2(U)$ in range from –7 V to 7 V. In the answer sheet write C_1 and C_2 values at 0 V, 3 V, and 6 V. Write down the formula used for calculating capacitance from raw measurements. Also write Board ID and room temperature.	2.3pt

EuPhO-2020 Online | Experimental Problems. Language: **English**

1 Hidden Charge

1.1 Introduction

An unknown point charge Q is fixed in a region of space. Electrons launched parallel to the z axis far from the charge will scatter electrostatically off of the fixed charge and strike a detecting screen. It is possible learn about the details of the hidden charge by varying the initial kinetic energy as well as the initial x_i and y_i coordinates of the electron beam and measuring the final coordinates x_f and y_f of where an electron strikes a finite flat screen perpendicular to the z axis and located at $z = 0$.

It is useful to know the Rutherford scattering formula,

$$b = \frac{kqQ}{2E}\frac{1}{\tan(\theta/2)}$$

where b is the impact parameter, E is the energy of the electron, $q = -1.602 \times 10^{-19}$C is the charge of the electron, $k = 8.99 \times 10^9$ Nm²/C², and θ is the scattering angle. The impact parameter is defined as the closest approach of the electron to the target, assuming that the electron were unaffected by the target and hence would move in a straight line; the scattering angle is angle between the original velocity vector of the electron far from the target and the final velocity vector of the electron far from the target after scattering.

1.2 Task

The task is to determine the position (x_Q, y_Q, z_Q) and also the magnitude and sign of the fixed charge Q, as precisely as possible. You should provide rough, order of magnitude error estimates on these results. There is Gaussian error associated with initial beam location that is on the order of 0.5 mm.

As with all experiments, you must provide clearly labelled tables of data, clearly labelled graphs, and sufficient formulae derivations to make it clear what you have measured, and how you are deriving your results.

1.3 Program Interface

The program asks for an accelerating voltage with the prompt

```
Beam accelerating voltage in V:
```

Enter a number between 1 and 10000, and press **return**. The program then asks for the initial launch coordinates, starting with x_i, with the prompt

```
x-coordinate of the electron beam in cm:
```

Enter a number between -20 and 20 and then press **return**. Finally, the program asks for y_i, with the prompt

```
y-coordinate of the electron beam in cm:
```

Enter a number between -20 and 20 and then press **return**. If you enter an invalid number for any of these three, the program will prompt you with

```
Invalid entry.
```

and will then prompt you for the value again, reminding you of the allowed limits.

After the three numbers have been entered, the program will output

```
Electron beam fired with parameters (x, y, V) =
```

and it will restate your entered values, and then

```
Electron detected at (x, y) =
```

and give the screen location of the detected electron. However, if the electron misses the finite size screen, you will be told

```
Electron not detected...
```

The program then repeats, allowing you to enter in a new set of initial coordinates.

2 Black box

2.1 Introduction

You have a rigid mechanical black box consisting of a container of mass m_1. Inside the container there is a load of mass m_2 that hangs on an effectively massless spring of stiffness k_1 from the ceiling of the box. Another mass m_3 is hanged to the mass m_2 via another massless spring of stiffness k_2. There is a small viscous drag which depends on the velocity of the objects. The gravity of Earth is $g = 9.81$ m/s² and is parallel to the sides of the box.

The box can be moved up or down with a piece-wise constant acceleration. The acceleration pattern can be programmed through input by giving the duration (in seconds) and acceleration (in m/s²) for each step. The simulation shows in "real time" the force F exerted to

Figure 5: Left: an instrument-based experiment example (IPhO 2021 experiment problem 1); Right: a program-based experiment example (EuPhO 2020 experiment problem 1). For IPhO 2021 experiment problem 1, a circuit board with electronic components to be measured inside it is provided, where contestants have to conduct measurements for these components: **this instrument-based experiment requires contestants to appropriately conduct manipulations on real experiment instruments, which is not tested in program-based experiments.** For EuPhO 2020 experiment problem 1, a program simulates experiments about detecting unknown charge with electron beams, similar to the Rutherford scattering experiment: **this program-simulated experiment is more related to modern physics, and it is impractical in a typical Olympiad venue due to cost and safety constraints.**

E Example of Instrument-based and Program-based experiments

We provide two examples of Instrument-based and Program-based experiments, as shown in Figure 5: the instrument-based experiment is IPhO 2021 experiment problem 1; and the program-based experiment is EuPhO 2020 experiment problem 1.

As shown and described in the caption of Figure 5, the program-based experiment can be more related to modern physics and bypasses the difficulties of cost and safety issues, although they are less real compared to instrument-based experiments.

F Related Work

F.1 Task Solving Agents

Since LLMs were shown to be capable of performing a wide range of tasks with appropriate prompts [30], researchers have been trying to enhance their capability and integration with tasks through prompt-based methods and tools. The concept of agent systems was then developed, exemplified by ReAct [17], in which LLMs iterate between reasoning and tool calls to perform more complicated tasks [31]. Later, agent-specific codebases have appeared: autogen [32] provides framework for creating multi-agent systems; smolagents [24] focuses more on code-related problem solving with defined domains. Upon these systems are more complicated general-purpose agent systems: for example, tool-creating agents like Alita [19] focus on creating and using tools for complex tasks with self-improvement; products like Openai Deep Research [33] and manus [34] focus on providing reliable research results to users. There has also been work focusing on optimizing the cost efficiency and effectiveness of agent systems [35, 36, 37]. These agent systems are mainly designed for general purpose tasks, with a focus on general virtual assistant related tasks represented by GAIA [21].

In addition to these general purpose agents, there is a growing body of work on domain-specific agents. For example, for math problems, some work has discussed the potential pipeline-based LLM approaches to mathematical problem solving [22, 25, 38]; for the humanities domain, HistAgent [23] introduces domain-specific high quality benchmarks and specially designed agent systems targeted at historical reasoning; for human-computer interaction and mental health, EmoAgent [39] studies the impact of LLM-based chat systems on users' mental health, and proposes agent-based methods to monitor users' mental states and ensure safer human-AI interactions. These domain-specific agents show the potential of agent systems in solving professional tasks requiring specialized domain knowledge; however, previous work rarely focuses on solving expert physics problems with agent systems.

F.2 Olympiads as Benchmarks

Olympiad competitions (e.g. IMO, IPhO, etc.) are widely regarded as challenging even to domain experts, and have been seen as challenging benchmarks for LLMs. Recent work shows progress in Math Olympiad problems, with both natural-language-based pipelines [25] and formal-language-based provers [26, 40].

Physics Olympiads have gradually drawn more attention in recent years as benchmarks for LLMs [11, 10, 12, 41, 42]. OlympiadBench [11] collects Olympiad problems across subjects, including physics. In 2025, SeePhysics [10] collects physics problems with images. Most benchmarks use publicly available problems from past Olympiads (e.g., prior IPhO problems). In contrast, PhyBench [12] curates an 'entirely original' dataset, aiming to minimize data leakage; it excludes images. Some other datasets, represented by PhysicsArena [42] and PhysReason [43], feature solutions with detailed reasoning processes. These benchmarks, though some contain problems that require image-related physics abilities, some require a high level of professional knowledge, and some require complicated reasoning and calculation, are primarily designed for nonagentic single-LLM settings.

G Discussions

G.1 Physics Experimental Exams: Instrument-based Exam and Program-based Exam as Proxies for Research-level Physics Experiments

Experiments are foundational to physics research [44]. In Physics Olympiads,**experimental exams** have served as proxies for research-level physics investigations. There are two common formats of Physics Olympiad experimental exams: **instrument-based Experiments** and **program-based Experiments**[5]. For the former, contestants are provided with (sometimes simplified) instruments and

[5]Most Physics Olympiads use instrument-based experiments, for example, current IPhO, APhO, EuPhO, etc. During the pandemic, some Olympiads utilize program-based experiments, for example, APhO 2021, EuPhO 2020, EuPhO 2021, etc.

are asked to plan and carry out experiments and measurements; for the latter, contestants design and carry out simulated experiments with programs, followed by data analysis.

In this work, we mainly focus on the IPhO 2025 Theory Problems, rather than the instrument-based experimental problems. This limitation is caused in part by limited access to the experiment instruments. We hope with advances in robotics future LLM-based agents may also work on these experimental exams. Moreover, we argue that program-based exams could also serve as good proxies for benchmarking current and near-future AI's ability in research-level experiments. Although instrument-based experiments are closer to real-world researches, there are **two aspects** where program-based experiments are potentially better:

Program-based experimental exams can simulate more advanced and complex experiments compared to instrument-based exams. For instrument-based exams, contestants have to design experiments, conduct manual operations and measurements, process data and answer questions. For program-based experiments, contestants run simulated experiments through programs: While manual experimentation is no longer required, critical challenges like experimental design and data analysis continue to play a central role in program-based assessments. One big advantage is that the program-based exams can access experiments that are more complex or require more sophisticated designing, because they are not limited by cost, safety concerns and other limiting factors for instrument-based experiments. We provide two examples for instrument-based experiments and program-based experiments in Figure 5, shown in Appendix E.

Program-based experimental exams can shift the focus from testing robotic manipulation of instruments to evaluating the ability in "physics". There has been work using robotic systems for implementing experiments in physics [45]; however, in all, current AI systems still fall short in conducting robotic tasks for manipulating instruments in experiments. This critical challenge will be addressed with advances in robotics research. With program-based experiments, we hope that in the near future one can already better evaluate essential abilities of AI systems in experimental physics.

We also emphasize that while program-based experiments have potential advantages in benchmarking AI's capability in experimental physics, the instrument-based tests are certainly essential ultimately. As AI systems advance toward ASI and daily use, one should not overlook the importance of instrument-based tests as they (1) yield a smaller deviation from real-world research; (2) provide a better metric for characterizing robotic capability; and (3) evaluate the performance better under extreme or unexpected conditions like instrument failure, etc.

G.2 Verifiable Physics Reasoning

In this work, we utilize an answer reviewer tool to verify the deductions of the worker, which is based purely on natural language. A huge step in automatic math proofs are **verifiable LLM-generated proofs** written in Lean [46]. Some previous work proposes to use Lean-like tools for verifiable physics formula deductions [47, 48]. However, the process of deriving physics formulas from natural-language-based problems, whether grounded in theoretical models or experimental observations, currently lacks reliable automatic verification through comparable processes. This limitation remains an open area for further research. Promising directions for future exploration include: (1) developing methods to verify the abstraction and transformation between formulas, physical representations, and intuitive reasoning; (2) establishing a more rigorous and transparent calculation framework that supports verifiability; and (3) enhancing answer-review systems with tools that possess broader and deeper expertise in physics.

The first direction represents a big challenge in symbolic AI [49]. In relation to developing verifiable physics calculations, extensive work has been done on machine-checked mathematics in Lean-like languages [50]. For example, Deepseek-Prover, Kimina-Prover and AlphaProof [51, 52, 38] use RL-based methods to train LLMs that excel in generating lean-based proofs. Others discuss test-time scaling methods through pipeline workflows [53, 54, 55]. Future work for solving physics problems might adopt similar methods to improve the ability to generate reliable and verifiable solutions.

In summary, we suggest future work on using AI systems for physics problem-solving to focus on: (1) program-based or instrument-based experiments; and (2) verifiable and reliable solution generation.