

# SPHYR: SPATIAL-PHYSICAL REASONING BENCHMARK ON MATERIAL DISTRIBUTION

Anonymous authors

Paper under double-blind review

## ABSTRACT

We introduce a novel dataset designed to benchmark the physical and spatial reasoning capabilities of Large Language Models (LLM) based on topology optimization, a method for computing optimal material distributions within a design space under prescribed loads and supports. In this dataset, LLMs are provided with conditions such as 2D boundary, applied forces and supports, and must reason about the resulting optimal material distribution. The dataset includes a variety of tasks, ranging from filling in masked regions within partial structures to predicting complete material distributions. Solving these tasks requires understanding the flow of forces and the required material distribution under given constraints, without access to simulation tools or explicit physical models, challenging models to reason about structural stability and spatial organization. Our dataset targets the evaluation of spatial and physical reasoning abilities in 2D settings, offering a complementary perspective to traditional language and logic benchmarks <sup>1</sup>.

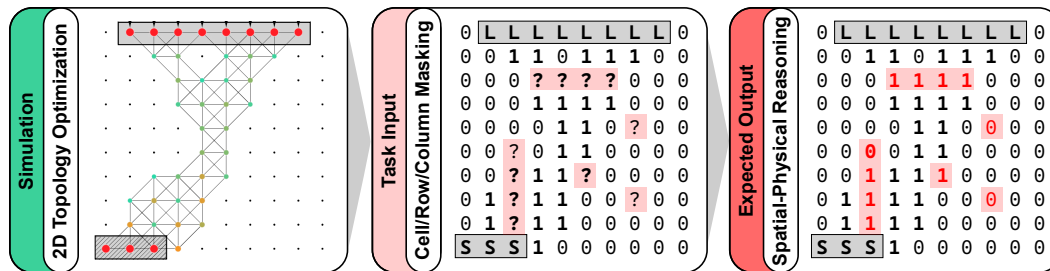


Figure 1: Topology Optimization is used to calculate material distribution. Masking individual cells, rows, columns or the complete distribution space offer interesting spatial physical reasoning tasks.

## 1 INTRODUCTION

Large language models (LLMs) have achieved strong performance on linguistic and logical tasks, but their ability to reason about physical systems and spatial structures remains underexplored Zhang et al. (2025). Existing benchmarks primarily probe either visual perception or text-based commonsense knowledge, but few explicitly test reasoning grounded in physical constraints.

For example, visual question-answering benchmarks such as CLEVR focus on object attributes and spatial relations in synthetic scenes Johnson et al. (2016), while intuitive physics datasets like IntPhys and Physion evaluate models’ ability to predict or assess the plausibility of physical events in videos Riochet et al. (2020); Bear et al. (2022). Interactive environments such as PHYRE Bakhtin et al. (2019) and stability-focused datasets like ShapeStacks Groth et al. (2018) further probe causal reasoning and contact mechanics, whereas text-based datasets such as PIQA Bisk et al. (2019) and PhysReason Zhang et al. (2025) target physical commonsense and multi-step problem solving in language form.

<sup>1</sup>Huggingface Dataset: anonymized

054 These benchmarks collectively advance our understanding of physical reasoning in models but share  
 055 a key limitation: none require reasoning about how material should be arranged within a structure to  
 056 satisfy load and support constraints. Such reasoning - central to engineering and design - demands  
 057 integrating spatial layout with physical principles such as force flow and stability, which remains a  
 058 blind spot in current LLM evaluation.

059 Beyond physical reasoning, recent work like ARC-AGI-2 Chollet et al. (2025) has introduced grid-  
 060 based tasks for testing abstract reasoning and generalization. While unrelated to physics, this work  
 061 highlights the value of structured 2D representations for isolating reasoning capabilities. We build  
 062 on this intuition but shift the focus from symbolic transformations to spatially grounded physical  
 063 reasoning.

064 To address this gap, we introduce **SPhyR**, a new benchmark for evaluating spatial and physical  
 065 reasoning in LLMs. SPhyR formulates topology optimization-inspired tasks in a grid-based format,  
 066 where models must infer how to distribute material to support specified forces and constraints.  
 067 By testing whether models can reason about load paths, stability, and structural connectivity from  
 068 descriptions alone, SPhyR bridges the gap between language-based reasoning and physically grounded  
 069 design tasks. We benchmark state-of-the-art LLMs on SPhyR and reveal fundamental limitations in  
 070 their ability to integrate spatial and physical reasoning.

## 072 2 RELATED WORK

074 **Benchmarks for Physical and Spatial Reasoning** A wide range of benchmarks probe models’  
 075 understanding of physical and spatial reasoning (Table 1). CLEVR Johnson et al. (2016) evaluates  
 076 visual reasoning about objects and spatial relations in synthetic scenes, while CLEVRER Yi et al.  
 077 (2020) extends this to temporal and causal reasoning in videos. IntPhys Riochet et al. (2020) and  
 078 Physion Bear et al. (2022) test whether models can predict or assess the plausibility of physical  
 079 events, while ShapeStacks Groth et al. (2018) targets block stability prediction.

080 In interactive settings, PHYRE Bakhtin et al. (2019) challenges agents to solve 2D physics puzzles  
 081 by reasoning about actions and causal effects. Language-based datasets such as PIQA Bisk et al.  
 082 (2019) and PhysReason Zhang et al. (2025) shift the focus from perception to textual physical  
 083 reasoning, evaluating knowledge of everyday object interactions and multi-step physics problem  
 084 solving, respectively.

085 While these benchmarks advance physical reasoning evaluation, they largely focus on event prediction  
 086 or commonsense reasoning. None require models to determine optimal material arrangements under  
 087 explicit load and support constraints - a capability crucial for real-world engineering reasoning.

090 Table 1: Comparison of existing benchmarks evaluating physical and spatial reasoning. Our proposed  
 091 dataset focuses specifically on material distribution reasoning under boundary conditions, combining  
 092 spatial and physical understanding in structured tasks.

093 Benchmark	094 Format	Physical Reasoning	Spatial Reasoning	Notes
095 CLEVR (2017)	Visual QA	✗	✓	Scene reasoning
096 CLEVRER (2020)	Video QA	✓	✓	Causal events
097 IntPhys (2018)	Video plausibility	✓	✓	Violation detection
098 Physion (2021)	Video prediction	✓	✓	Object behavior prediction
099 ShapeStacks (2016)	Image classification	✓	✓	Block stability
PHYRE (2019)	2D physics puzzles	✓	✓	Action planning
100 PIQA (2020)	Text QA	✓	✗	Physical commonsense
101 PhysReason (2023)	Text QA	✓	✗	Multi-step physics
102 <b>SPhyR</b>	Structured prediction	✓	✓	Material distribution

104 **Topology Optimization as a Benchmark** Topology optimization (TO) Bendsoe & Sigmund (2004)  
 105 is a well-established method for computing optimal material layouts in a domain under specified  
 106 forces and supports. Prior work on Machine Learning (ML) in this space has focused on accelerating  
 107 solvers or generating high-quality designs Banga et al. (2018); Rawat & Shen (2019). Our work

repurposes topology optimization as a reasoning benchmark rather than a design tool. By framing it as a grid-based prediction problem, SPhyR tests whether LLMs can infer material distributions solely from boundary conditions and physical constraints - without access to solvers or simulation engines. This setup complements existing physical reasoning benchmarks by embedding spatial and physical structure into tasks that require more than pattern recognition.

**Machine Learning for Topology Optimization** ML for TO has shown that CNNs, particularly U-Net encoder-decoder models, can predict optimized material layouts with high fidelity and orders-of-magnitude faster than conventional solvers Banga et al. (2018); Rawat & Shen (2019). Early work framed TO as an image segmentation problem Sosnovik & Oseledets (2017), while later studies demonstrated strong generalization across boundary conditions using U-Net-style CNNs Zhang et al. (2020). These approaches often incorporate physics-informed loss functions or compliance constraints, embedding explicit structural knowledge into training. Surveys highlight the success of such domain-specific methods Shin et al. (2023). In contrast, SPhyR evaluates general-purpose LLMs in a zero-shot setting, probing whether emergent, implicit physical knowledge acquired during broad pretraining can substitute for explicitly learned physics.

**Structured Reasoning Beyond Physics** Finally, our work connects to broader research on structured reasoning in grid-based environments. ARC-AGI-2 Chollet et al. (2025) tests abstract reasoning and generalization in symbolic, non-physical tasks. While ARC-AGI-2 and SPhyR share a structured representation, SPhyR introduces grounded physical constraints, bridging the gap between abstract symbolic reasoning and the physically grounded reasoning required for real-world design.

### 3 PROBLEM SETUP

**Topology Optimization Task** Topology optimization is a computational design method used to determine the optimal material distribution within a given domain, subject to boundary conditions such as external forces and fixed supports. The goal is to produce structures that maximize performance metrics, typically minimizing compliance (maximizing stiffness) while using the least amount of material. The resulting solutions exhibit highly structured patterns that balance material connectivity, load paths, and spatial efficiency.

In this work, we repurpose topology optimization problems as reasoning tasks for LLMs. Rather than solving the optimization numerically, models are challenged to predict or complete plausible material distributions based solely on the specified forces, supports, and boundary. This formulation requires models to infer structural principles such as load transfer, stability, and efficient use of material, without access to physical simulation engines.

**Input and Output Specification** Each task instance in our benchmark is defined by a set of boundary conditions and a corresponding material distribution. The **inputs** provided to the model are: **2D boundaries**: A discretized 2D grid representing the spatial extent of the structure, **fixed supports**: Locations within the boundaries that act as load bearing supports and **applied forces**: Locations within the boundaries specifying external loads. The **output** expected from the model is a partial or complete *material distribution* over the domain grid, indicating where material should be placed to form a material optimized, that is minimum material distributed, but stable structure under the given boundary conditions. All inputs and outputs are represented in structured formats suitable for LLMs, through textual descriptions and serialized grids. No direct access to simulation results or numerical solvers is provided.

**Reasoning Challenges** The tasks in our benchmark require a combination of physical and spatial reasoning that poses significant challenges for current large language models. First, models must infer how forces propagate through the structure, deciding where material is necessary to maintain stability and support loads. This involves understanding force paths, support connectivity, and load transfer-concepts that are rarely encountered in typical LLM training data. Second, models must reason spatially about the layout of material across a 2D grid. Predicting plausible completions requires local coherence (e.g., avoiding isolated material islands) as well as global structural organization (e.g., maintaining continuous load paths from forces to supports). Moreover, models must solve these tasks without explicit simulation tools or numerical methods. Instead, they must generalize from the provided boundary conditions and partial observations, synthesizing structures that satisfy implicit

physical constraints. These reasoning demands span from local (individual cells or lines) to global (complete structures), creating a rich and graded challenge space for evaluating LLM capabilities beyond language-based tasks.

**Task Variations** We define several task variations based on the nature and extent of the masked regions in the material distribution with two difficulty levels, easy and hard. Easy is distribution based on binary values such as material or no material, while hard is based on a continuous value range, 0 to 1. **N-Random Cell(s)**: Predict the material state of N randomly masked cell(s), where N is one of 1, 5 or 10. **N-Random Row(s)**: Predict the material state of N randomly masked row(s), where N is one of 1 or 3. **N-Random Columns(s)**: Predict the material state of N randomly masked columns(s), where N is one of 1 or 3. **Full Structure Prediction**: Predict the complete material distribution based only on boundary conditions. These variations allow us to systematically probe local and global reasoning abilities, from single-cell predictions to complex structural synthesis (Samples in Figure 2).

## 4 DATASET DESCRIPTION

The dataset was generated using McNeel Rhinoceros 8, a computer-aided design (CAD) software widely used in architecture, engineering, and industrial design, in combination with the Grasshopper visual programming environment and the Millipede plugin for topology optimization Michalatos & Kaijima (2024). Millipede enables the simulation of material distribution under given boundary conditions using density-based optimization methods.

We constructed a set of 2D samples by varying the positions of applied forces and supports. Applied forces and supports were systematically placed to create diverse loading scenarios. In this work we focused on load-support scenarios that can be found most likely in building structural scenarios, that is, load from the top, support on the bottom. We introduce one support and load location for each scenario, ranging from 3 to 6 cells in width. Each optimization was performed until 10 iterations, resulting in plausible material distributions optimized for stiffness and material efficiency. By leveraging the flexibility of Grasshopper and Millipede, we ensured substantial variation across samples, including different structural typologies and support-force relationships. This variability is crucial for creating tasks that require generalization beyond memorization of fixed patterns.

**Dataset Statistics** The dataset includes 2D structural optimization samples. Specifically, material distributions are generated on grids of size  $10 \times 10$ . These discretizations balance computational tractability with sufficient spatial complexity to require nontrivial reasoning. We generate a diverse set of boundary conditions and corresponding optimized material distributions. The dataset is organized into task-specific subsets, covering random cell completion, row and column completion, and full structure prediction tasks as described in Table 2. In total, the dataset contains **1296** scenarios for all task variations. Each sample is accompanied by structured representations of the boundary conditions (forces, supports) and the corresponding ground truth material distribution. The dataset is organized such that each task variant corresponds to a separate subject, allowing for isolated evaluation and targeted analysis. The subjects for easy and hard difficulties are: 1 Random Cell, 5 Random Cell, 10 Random Cell, 1 Random Row, 3 Random Row, 1 Random Column, 3 Random Column and Full.

**Input and Output Formats** Each sample in the dataset is represented as a structured input-output pair designed for compatibility with large language models. Samples are grouped into task-specific subjects, enabling targeted evaluation of different reasoning challenges.

The input consists of a natural language prompt that describes the task and defines the structural grid format. Within this grid, different symbols represent key physical roles:  $\perp$  indicates an applied load,  $S$  marks a support, and  $\nabla$  denotes a masked cell whose material state must be predicted. Known material or void regions are explicitly represented using binary values (1 for solid material, 0 for empty space). The prompt also includes clear instructions emphasizing structural plausibility and material efficiency, along with a grid where each row appears on a separate line and cell values are space-separated.

The expected output is a completed version of the same grid, where all  $\nabla$  cells are replaced by predicted values (1 or 0) while preserving the original structure and formatting. No explanation or commentary is included in the output-only the raw grid content.



	Difficulty: Easy Prompt (GRID)	Ground Truth	Difficulty: Hard Prompt (GRID)	Ground Truth
216				
217				
218	<b>1 Random Cell</b>			
219				
220				
221				
222	<b>5 Random Cell</b>			
223				
224				
225				
226				
227	<b>10 Random Cell</b>			
228				
229				
230				
231				
232				
233	<b>1 Random Row</b>			
234				
235				
236				
237				
238	<b>3 Random Row</b>			
239				
240				
241				
242				
243	<b>1 Random Row</b>			
244				
245				
246				
247				
248	<b>3 Random Row</b>			
249				
250				
251				
252				
253	<b>1 Random Column</b>			
254				
255				
256				
257				
258	<b>3 Random Column</b>			
259				
260				
261				
262				
263	<b>Full</b>			
264				
265				
266				
267				
268				
269				

Figure 2: Overview of task variations: predicting material distributions for N random cells, rows, columns, or a full structure 3.

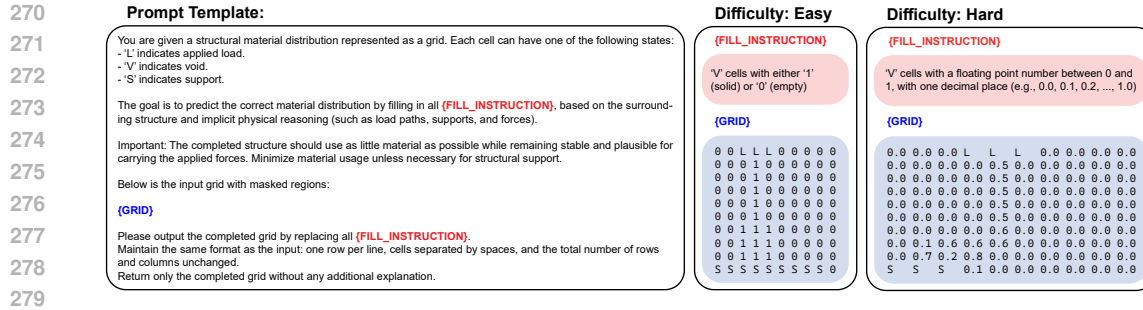


Figure 3: Prompt template used across all tasks and difficulty levels, showing instructions and grid format as served to models for evaluation.

Each subject is labeled with a difficulty level. In "easy" variants, the ground truth material distribution is binary, focusing on high-level structural placement and discrete spatial reasoning. In "hard" variants, the underlying distributions are continuous or involve more complex structural dependencies, requiring finer-grained predictions and deeper reasoning about stress propagation and global support. Please refer to Table 3 for our evaluation prompt template.

## 5 EVALUATION SETUP

### 5.1 EVALUATION METRICS

We evaluate model performance using three complementary metrics: **Exact Match**, **Score**, and **Normalized Score**. These metrics are designed to capture both perfect completions and partial correctness relative to the structural ground truth. **Exact Match** is a binary metric indicating whether the model’s output exactly matches the ground truth in all masked regions. A prediction receives an Exact Match score of 1 if no errors are present, and 0 otherwise. **Score** quantifies how close the predicted output is to the ground truth, relative to the number of differences between the raw input and the ground truth. Specifically, it is computed as the proportion of masked region errors corrected by the model:

$$\text{Score} = 1 - \frac{\text{differences count between model output and ground truth}}{\text{differences count between input and ground truth}}$$

A score of 1 indicates a perfect correction of all masked regions, while a score below 0 implies that the model introduced more errors than it corrected. **Normalized Score** is a clipped version of Score, with a lower bound of 0, ensuring interpretability on a  $[0, 1]$  scale even when the model degrades the input:

$$\text{Normalized Score} = \max(0, \text{Score}).$$

Together, these metrics capture whether a model achieves perfect reconstruction, how effectively it improves over the incomplete input, and how reliably it avoids introducing additional errors. In addition to these quantitative metrics, we also perform qualitative analysis by visually inspecting predicted outputs for structural plausibility, focusing on factors such as load path continuity, material efficiency, and connectivity.

### 5.2 EXPERIMENTS

To establish baseline performance, we evaluate a broad set of contemporary language models in a zero-shot setting. From OpenAI, we include GPT-3.5 Brown et al. (2020), GPT-4.1 OpenAI et al. (2024a), and GPT-4o OpenAI et al. (2024b), representing successive generations with improved reasoning and multimodal capabilities. From Anthropic, we test Claude 3.7 Sonnet anthropic (2025a) and Claude Opus 4 anthropic (2025b), the strongest in the Claude family. From Google DeepMind, we include Gemini 1.5 Pro Team et al. (2024) and Gemini 2.5 Pro Comanici et al. (2025), designed for complex multimodal reasoning. We also assess DeepSeek-R1 DeepSeek-AI et al. (2025), an open-source model for scientific and engineering tasks, and Perplexity Sonar Team (2025a) and Sonar Reasoning Team (2025b), tuned for information-seeking and multi-step reasoning. Models are prompted (G)

with structured descriptions of boundary conditions, forces, and supports, without simulation tools or external knowledge. A balanced subset of **100** examples spanning all task variations is used, and all models are evaluated under identical conditions via publicly available APIs (Table 4). Performance is measured using the metrics defined in Section 5.1.

## 6 RESULTS AND ANALYSIS

### 6.1 QUANTITATIVE RESULTS

Table 2: Exact Match, Overlap- and Normalized Overlap-Score across all difficulties, tasks and models.

Task	Metric	GPT 3.5 Turbo	GPT 4.1	GPT 4o	Claude 3.7 Sonnet	Claude Opus 4	Gemini 1.5 Pro	Gemini 2.5 Pro	DeepSeek-R1	Perplexity Sonar	Perplexity Sonar Reasoning
<b>Difficulty: Easy</b>											
1 Random Cell	Exact Match	37	26	1	75	82	65	38	58	49	0
	Score	-232.00	26.00	-1450.00	62.00	82.00	65.00	-702.00	58.00	-116.00	-4947.00
5 Random Cells	Exact Match	37.00	26.00	1.00	75.00	82.00	65.00	38.00	58.00	49.00	0.00
	Score	-25.77	23.45	-244.63	67.07	84.27	68.07	-61.32	60.30	-123.27	-917.05
10 Random Cells	Exact Match	20.85	23.65	0.80	67.07	84.27	68.07	39.15	60.30	35.25	0.00
	Score	0	0	0	4	15	3	8	2	1	0
1 Random Row	Exact Match	6.48	27.36	-111.70	63.27	78.72	66.03	12.23	58.66	20.72	-437.93
	Score	21.95	27.36	0.66	63.27	78.72	66.03	44.21	58.86	49.52	0.00
3 Random Rows	Exact Match	12	20	4	40	52	40	30	14	20	0
	Score	19.40	75.50	-61.50	87.50	90.50	87.20	41.20	54.40	-7.10	-414.50
1 Random Column	Exact Match	36.30	75.50	12.90	87.50	90.50	87.20	61.00	54.40	57.60	0.00
	Score	25.20	15.70	0	10	23	12	26	8	12	0
3 Random Columns	Exact Match	41.30	16.50	2.30	34.80	68.50	50.60	69.10	37.20	47.60	-407.00
	Score	36.08	41.57	19.45	60.50	74.05	67.73	69.08	51.13	53.53	-99.65
Full	Exact Match	40.25	41.57	26.07	60.50	74.05	67.73	69.73	51.13	54.82	0.00
	Score	46.56	64.69	57.81	63.39	70.09	75.74	71.59	48.49	57.83	37.49
Average	Exact Match	8.00	6.75	1.00	21.38	32.12	22.62	19.00	14.88	14.12	0.00
	Score	-11.94	43.64	-233.81	65.80	80.07	71.38	-51.30	55.25	2.06	-911.63
Overall Average	Exact Match	34.85	43.77	16.50	67.43	80.07	71.39	60.63	55.31	54.56	4.69
	Normalized Score										
<b>Difficulty: Hard</b>											
1 Random Cell	Exact Match	0	13	0	39	76	65	75	37	14	0
	Score	-1072.00	13.00	-1852.00	39.00	73.00	65.00	33.00	-6.00	-405.00	-4855.00
5 Random Cells	Exact Match	0.00	13.00	0.00	39.00	76.00	65.00	75.00	37.00	14.00	0.00
	Score	-405.65	3.80	-393.75	19.30	81.20	67.30	76.65	40.60	-99.55	-939.90
10 Random Cells	Exact Match	0.85	4.60	0.00	19.30	81.20	67.70	77.85	48.40	17.05	0.00
	Score	0	0	0	0	3	3	14	3	0	0
1 Random Row	Exact Match	0.11	3.26	-178.26	17.35	77.61	64.87	75.97	49.09	-65.35	-429.75
	Score	0.11	3.67	0.43	17.35	77.61	65.32	75.97	52.29	17.03	0.00
3 Random Rows	Exact Match	0	2	0	12	48	36	46	33	7	0
	Score	-174.30	39.10	-136.50	64.80	85.00	80.50	79.80	79.60	21.30	-400.80
1 Random Column	Exact Match	2.30	40.10	2.40	65.00	85.00	80.50	79.80	79.60	43.20	0.00
	Score	0	0	0	6	21	11	12	17	1	0
3 Random Columns	Exact Match	0	0	0	6	21	11	12	17	1	0
	Score	-45.00	29.85	-25.55	62.48	81.85	73.07	77.83	76.88	44.18	-95.22
Full	Exact Match	5.92	30.30	6.15	62.48	81.85	73.07	77.83	76.88	44.68	0.00
	Score	-174.10	1.70	-161.30	18.60	54.20	44.60	60.30	46.80	-2.00	-406.70
Average	Exact Match	2.00	3.00	1.70	18.60	54.30	46.80	60.30	47.90	11.80	0.00
	Score	0	0	0	3	7	5	3	5	2	0
Overall Average	Exact Match	-54.05	4.97	-34.48	36.80	69.00	60.98	58.02	62.28	21.35	-94.25
	Score	3.57	5.17	3.33	37.03	69.00	61.33	58.02	64.33	24.95	0.00
Overall Average	Exact Match	0	0	0	0	0	0	0	0	0	0
	Normalized Score	10.93	16.96	9.21	30.09	49.34	51.70	40.48	37.19	28.66	37.34
Overall Average	Exact Match	11.05	16.96	11.10	30.29	49.34	51.70	40.56	37.41	28.90	37.34
	Normalized Score										
Overall Average	Exact Match	4.00	4.31	0.50	15.00	30.06	20.88	22.81	15.25	8.88	0.00
	Score	-140.35	28.86	-290.20	50.93	75.74	67.44	5.73	51.78	-27.50	-904.83
Overall Average	Exact Match	19.04	29.18	9.82	51.78	75.95	67.66	64.40	55.39	39.88	4.68
	Normalized Score										

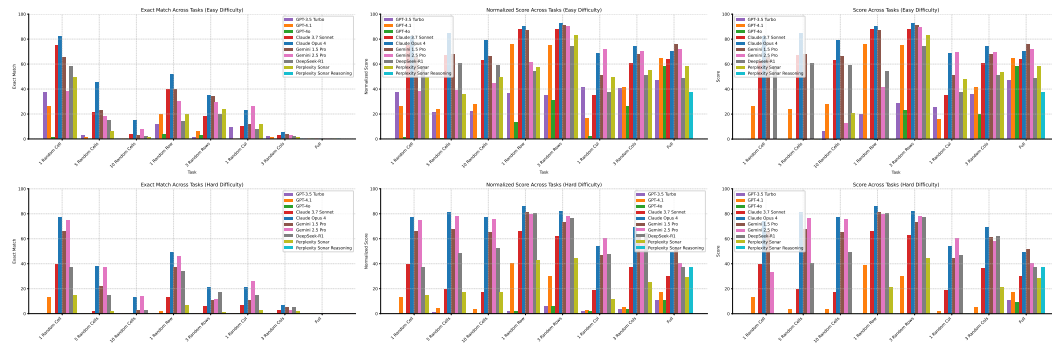


Figure 4: Exact Match, Overlap- and Normalized Overlap-Score plot for easy (top) and hard (bottom) difficulty across all tasks and models.

We evaluate model performance using Exact Match, Score, and Normalized Score (Section 5.1). Results show that performance generally decreases with task difficulty, with "easy" tasks yielding higher scores than "hard" ones. Exact matches are rare, but moderate Normalized Scores indicate that models often make meaningful partial corrections. Localized tasks, such as single-cell or row completions, are consistently easier than full-structure prediction, which remains highly challenging for all models. Row tasks tend to outperform column tasks, suggesting directional bias or easier contextual inference. Overall, current LLMs show limited ability to reason about global structure and physical constraints, though some local spatial reasoning is evident, especially in simpler scenarios. In addition to the zero-shot evaluations reported here, we conducted supplementary few-shot experiments, the results of which are provided in Appendix E.

### 6.1.1 ROTATION EXPERIMENTS

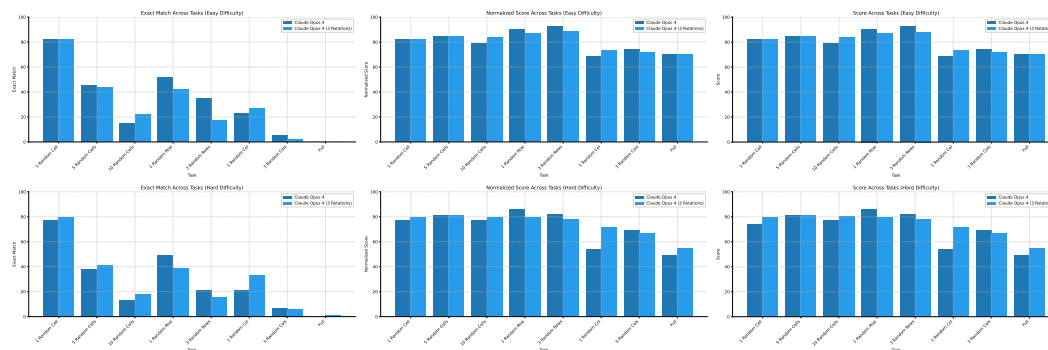


Figure 5: Claude Opus 4 performance on rotated setups (load left, support right) compared to the default top-bottom configuration for easy (top) and hard (bottom) difficulty across all tasks. Results for additional models are in the Appendix D.

We introduce a rotated setup with the load placed on the left and the support on the right to examine whether differences between row and column tasks arise from task structure or model bias. GPT-4.1, Claude Opus 4, Gemini 2.5 Pro, DeepSeek R-1, and Perplexity Sonar were evaluated on a subset of easy tasks and compared with the default top-bottom setup. To extend the analysis, the full task set was then run with Claude Opus 4, the strongest model under rotation (Figure 5). Additional plots and tables for other models can be found in Appendix D. Furthermore, we provide a complete prompt and completion sample in the Appendix D.1.

### 6.1.2 PHYSICS-ENHANCED AND -NEUTRAL PROMPT EXPERIMENTS

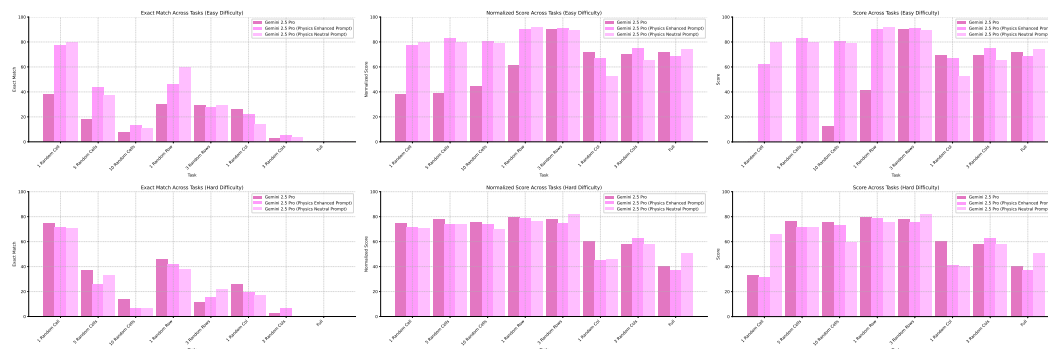


Figure 6: Gemini 2.5 Pro under default, physics-enhanced, and -neutral prompts; easy (top) and hard (bottom) across all tasks. See Appendix F for results table.

We investigated the role of prompt design by comparing physics-enriched and physics-neutral formulations. In the physics-enhanced setting, prompts were augmented with terminology from

432 structural mechanics. This improved performance on easier tasks with limited masking but led to  
433 underperformance on more challenging tasks relative to the default prompt. In the physics-neutral  
434 setting, all physics-related descriptions were removed, isolating spatial reasoning. Results from this  
435 setup provide a counterpoint to the physics-enhanced variant and help decouple spatial from physical  
436 contributions to model performance.

437 Figure 6 illustrates these effects for Gemini 2.5 Pro across default, physics-enhanced, and physics-  
438 neutral prompts, highlighting the trade-off between easier and harder task performance. Physics-  
439 neutral and -enhanced prompt and completion samples are provided in Appendix F.1 and F.2,  
440 respectively.

## 442 6.2 QUALITATIVE RESULTS

443  
444 To complement the quantitative evaluation, we analyze model outputs qualitatively by inspecting  
445 predicted material distributions for different task variations. Successful completions typically show  
446 structurally plausible material layouts that connect supports to load points while maintaining conti-  
447 nuity across the domain. In easier tasks, especially random cell and line completions, models often  
448 correctly infer local material presence based on nearby context. Notably, while Gemini 2.5 Pro  
449 ranks first in overall performance on overall tasks, it frequently modifies the existing (non-masked)  
450 structure more than necessary, indicating a tendency to overwrite rather than extend the original  
451 design. However, we observe several common failure modes across models: they sometimes predict  
452 disconnected structures, resulting in isolated material islands that would not carry load effectively;  
453 they may also overbuild regions, producing dense material clusters without regard to force flow; and,  
454 particularly in hard and 3D tasks, they frequently underbuild or omit critical connections, failing to  
455 maintain continuous paths from force application points to supports. Representative examples of  
456 model successes and failures can be found in the Appendix G.

## 457 7 DISCUSSION

458  
459 The quantitative and qualitative results highlight key challenges for large language models in spatial  
460 and physical reasoning tasks. While models demonstrate some ability to predict local material  
461 placements in simple scenarios, performance drops significantly for harder masked completions and  
462 full structure predictions. Typical errors, such as disconnected or overbuilt regions, suggest that  
463 current LLMs struggle to internalize fundamental physical concepts like load path continuity and  
464 material efficiency without explicit simulation or physics grounding. Rotation experiments confirmed  
465 that such limitations persist even when load and support orientations are reversed, indicating that  
466 row-column differences reflect deeper reasoning challenges rather than simple directional bias.  
467 Prompt experiments further showed that physics-enriched wording can improve performance on  
468 easier tasks but not on harder ones, while physics-neutral prompts perform comparably to the default,  
469 suggesting that models rely partly on surface cues without deeper grounding. Together, these findings  
470 underscore the value of SPhyR as a benchmark: it highlights the gap between the linguistic strengths  
471 of LLMs and their limited ability to reason about structured, physics-governed contexts.

## 472 8 CONCLUSION

473  
474 We introduce a new benchmark for evaluating the spatial and physical reasoning abilities of large  
475 language models, based on topology optimization tasks in the 2D domain. Our dataset requires  
476 models to infer plausible material distributions given only forces, supports, and boundary conditions,  
477 presenting a structured and physically grounded reasoning challenge. Through quantitative and  
478 qualitative analyses, we show that current LLMs, despite strong general reasoning capabilities,  
479 struggle with tasks requiring physical intuition and structural coherence. Our results highlight an  
480 important gap between language-based reasoning and physical-spatial understanding. Looking ahead,  
481 future work could expand the benchmark by increasing the grid size beyond the current  $10 \times 10$   
482 resolution and by extending tasks from the 2D to the 3D domain. A small set of 3D samples are  
483 available in Appendix C and in our GitHub repository, offering an initial starting point for extensions  
484 into 3D tasks. We hope this benchmark inspires further research at the intersection of machine  
485 reasoning, physical simulation, and structural design, pushing models toward deeper generalization  
beyond text and logic alone.

## REFERENCES

- 486 anthropic. Claude 3.7 Sonnet and Claude Code, February 2025a. URL [https://www.](https://www.anthropic.com/news/claude-3-7-sonnet)  
487 anthropic.com/news/claude-3-7-sonnet.
- 488 anthropic. Introducing Claude 4, May 2025b. URL [https://www.anthropic.com/news/](https://www.anthropic.com/news/claude-4)  
489 claude-4.
- 490 Anton Bakhtin, Laurens van der Maaten, Justin Johnson, Laura Gustafson, and Ross Girshick.  
491 PHYRE: A New Benchmark for Physical Reasoning, August 2019. URL [http://arxiv.org/](http://arxiv.org/abs/1908.05656)  
492 abs/1908.05656. arXiv:1908.05656 [cs].
- 493 Saurabh Banga, Harsh Gehani, Sanket Bhilare, Sagar Patel, and Levent Kara. 3D Topology Optimiza-  
494 tion using Convolutional Neural Networks, August 2018. URL [http://arxiv.org/abs/](http://arxiv.org/abs/1808.07440)  
495 1808.07440. arXiv:1808.07440 [cs].
- 496 Daniel M. Bear, Elias Wang, Damian Mrowca, Felix J. Binder, Hsiao-Yu Fish Tung, R. T. Pramod,  
497 Cameron Holdaway, Sirui Tao, Kevin Smith, Fan-Yun Sun, Li Fei-Fei, Nancy Kanwisher, Joshua B.  
498 Tenenbaum, Daniel L. K. Yamins, and Judith E. Fan. Physion: Evaluating Physical Prediction from  
499 Vision in Humans and Machines, June 2022. URL <http://arxiv.org/abs/2106.08261>.  
500 arXiv:2106.08261 [cs].
- 501 Martin P. Bendsøe and Ole Sigmund. *Topology Optimization*. Springer, Berlin, Heidelberg, 2004.  
502 ISBN 978-3-642-07698-5 978-3-662-05086-6. doi: 10.1007/978-3-662-05086-6. URL [http:](http://link.springer.com/10.1007/978-3-662-05086-6)  
503 [://link.springer.com/10.1007/978-3-662-05086-6](http://link.springer.com/10.1007/978-3-662-05086-6).
- 504 Yonatan Bisk, Rowan Zellers, Ronan Le Bras, Jianfeng Gao, and Yejin Choi. PIQA: Reasoning about  
505 Physical Commonsense in Natural Language, November 2019. URL [http://arxiv.org/](http://arxiv.org/abs/1911.11641)  
506 abs/1911.11641. arXiv:1911.11641 [cs].
- 507 Tom B. Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal,  
508 Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel  
509 Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel M. Ziegler,  
510 Jeffrey Wu, Clemens Winter, Christopher Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott  
511 Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya  
512 Sutskever, and Dario Amodei. Language Models are Few-Shot Learners, July 2020. URL  
513 <http://arxiv.org/abs/2005.14165>. arXiv:2005.14165 [cs].
- 514 Francois Chollet, Mike Knoop, Gregory Kamradt, Bryan Landers, and Henry Pinkard. ARC-AGI-2:  
515 A New Challenge for Frontier AI Reasoning Systems, May 2025. URL [http://arxiv.org/](http://arxiv.org/abs/2505.11831)  
516 abs/2505.11831. arXiv:2505.11831 [cs].
- 517 Gheorghe Comanici, Eric Bieber, Mike Schaekermann, Ice Pasupat, Noveen Sachdeva, Inderjit  
518 Dhillon, Marcel Blistein, Ori Ram, Dan Zhang, Evan Rosen, Luke Marris, Sam Petulla, Colin  
519 Gaffney, Asaf Aharoni, Nathan Lintz, Tiago Cardal Pais, Henrik Jacobsson, Idan Szpektor, Nan-  
520 Jiang Jiang, Krishna Haridasan, Ahmed Omran, Nikunj Saunshi, Dara Bahri, Gaurav Mishra, Eric  
521 Chu, Toby Boyd, Brad Hekman, Aaron Parisi, Chaoyi Zhang, Kornrathop Kawintiranon, Tania  
522 Bedrax-Weiss, Oliver Wang, Ya Xu, Ollie Purkiss, Uri Mendlovic, Ilai Deutel, Nam Nguyen, Adam  
523 Langley, Flip Korn, Lucia Rossazza, Alexandre Ramé, Sagar Waghmare, Helen Miller, Nathan  
524 Byrd, Ashrith Sheshan, Raia Hadsell Sangnie Bhardwaj, Pawel Janus, Tero Rissa, Dan Horgan,  
525 Sharon Silver, Ayzaan Wahid, Sergey Brin, Yves Raimond, Klemen Kloboves, Cindy Wang,  
526 Nitesh Bharadwaj Gundavarapu, Iliia Shumailov, Bo Wang, Mantas Pajarskas, Joe Heyward, Martin  
527 Nikoltchev, Maciej Kula, Hao Zhou, Zachary Garrett, Sushant Kafle, Sercan Arik, Ankita Goel,  
528 Mingyao Yang, Jiho Park, Koji Kojima, Parsa Mahmoudieh, Koray Kavukcuoglu, Grace Chen,  
529 Doug Fritz, Anton Bulyenov, Sudeshna Roy, Dimitris Paparas, Hadar Shemtov, Bo-Juen Chen,  
530 Robin Strudel, David Reitter, Aurko Roy, Andrey Vlasov, Changwan Ryu, Chas Lechner, Haichuan  
531 Yang, Zelda Mariet, Denis Vnukov, Tim Sohn, Amy Stuart, Wei Liang, Minmin Chen, Praynaa  
532 Rawlani, Christy Koh, J. D. Co-Reyes, Guangda Lai, Praseem Banzal, Dimitrios Vytiniotis, Jieru  
533 Mei, Mu Cai, Mohammed Badawi, Corey Fry, Ale Hartman, Daniel Zheng, Eric Jia, James Keeling,  
534 Annie Louis, Ying Chen, Efen Robles, Wei-Chih Hung, Howard Zhou, Nikita Saxena, Sonam  
535 Goenka, Olivia Ma, Zach Fisher, Mor Hazan Taege, Emily Graves, David Steiner, Yujia Li, Sarah  
536 Nguyen, Rahul Sukthankar, Joe Stanton, Ali Eslami, Gloria Shen, Berkin Akin, Alexey Guseynov,

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

594 Rouillard, Jimgeng Rao, Zhihao Li, Yuntao Xu, Flavien Prost, Chi Zou, Ziwei Ji, Alberto Magni,  
 595 Tyler Liechty, Dan A. Calian, Deepak Ramachandran, Igor Krivokon, Hui Huang, Terry Chen,  
 596 Anja Hauth, Anastasija Ilić, Weijuan Xi, Hyeontaek Lim, Vlad-Doru Ion, Pooya Moradi, Metin  
 597 Toksoz-Exley, Kalessha Bullard, Miltos Allamanis, Xiaomeng Yang, Sophie Wang, Zhi Hong,  
 598 Anita Gergely, Cheng Li, Bhavishya Mittal, Vitaly Kovalev, Victor Ungureanu, Jane Labanowski,  
 599 Jan Wassenberg, Nicolas Lacasse, Geoffrey Cideron, Petar Dević, Annie Marsden, Lynn Nguyen,  
 600 Michael Fink, Yin Zhong, Tatsuya Kiyono, Desi Ivanov, Sally Ma, Max Bain, Kiran Yalasang,  
 601 Jennifer She, Anastasia Petrushkina, Mayank Lunayach, Carla Bromberg, Sarah Hodgkinson,  
 602 Vilobh Meshram, Daniel Vlasic, Austin Kyker, Steve Xu, Jeff Stanway, Zuguang Yang, Kai Zhao,  
 603 Matthew Tung, Seth Odoom, Yasuhisa Fujii, Justin Gilmer, Eunyoung Kim, Felix Halim, Quoc Le,  
 604 Bernd Bohnet, Seliem El-Sayed, Behnam Neyshabur, Malcolm Reynolds, Dean Reich, Yang Xu,  
 605 Erica Moreira, Anuj Sharma, Zeyu Liu, Mohammad Javad Hosseini, Naina Raisinghani, Yi Su,  
 606 Ni Lao, Daniel Formoso, Marco Gelmi, Almog Gueta, Tapomay Dey, Elena Gribovskaya, Domagoj  
 607 Ćevid, Sidharth Mudgal, Garrett Bingham, Jianling Wang, Anurag Kumar, Alex Cullum, Feng Han,  
 608 Konstantinos Bousmalis, Diego Cedillo, Grace Chu, Vladimir Magay, Paul Michel, Ester Hlavnova,  
 609 Daniele Calandriello, Setareh Ariafar, Kaisheng Yao, Vikash Sehwal, Arpi Vezer, Agustin Dal  
 610 Lago, Zhenkai Zhu, Paul Kishan Rubenstein, Allen Porter, Anirudh Baddepudi, Oriana Riva,  
 611 Mihai Dorin Istin, Chih-Kuan Yeh, Zhi Li, Andrew Howard, Nilpa Jha, Jeremy Chen, Raoul de  
 612 Liedekerke, Zafarali Ahmed, Mikel Rodriguez, Tanuj Bhatia, Bangju Wang, Ali Elqursh, David  
 613 Klinghoffer, Peter Chen, Pushmeet Kohli, Te I, Weiyang Zhang, Zack Nado, Jilin Chen, Maxwell  
 614 Chen, George Zhang, Aayush Singh, Adam Hillier, Federico Lebron, Yiqing Tao, Ting Liu, Gabriel  
 615 Dulac-Arnold, Jingwei Zhang, Shashi Narayan, Buhuang Liu, Orhan Firat, Abhishek Bhowmick,  
 616 Bingyuan Liu, Hao Zhang, Zizhao Zhang, Georges Rotival, Nathan Howard, Anu Sinha, Alexander  
 617 Grushetsky, Benjamin Beyret, Keerthana Gopalakrishnan, James Zhao, Kyle He, Szabolcs Payrits,  
 618 Zaid Nabulsi, Zhaoyi Zhang, Weijie Chen, Edward Lee, Nova Fallen, Sreenivas Gollapudi, Aurick  
 619 Zhou, Filip Pavetić, Thomas Köppe, Shiyu Huang, Rama Pasumarthi, Nick Fernando, Felix  
 620 Fischer, Daria Ćurko, Yang Gao, James Svensson, Austin Stone, Haroon Qureshi, Abhishek  
 621 Sinha, Apoorv Kulshreshtha, Martin Matysiak, Jieming Mao, Carl Saroufim, Aleksandra Faust,  
 622 Qingnan Duan, Gil Fidel, Kaan Katircioglu, Raphaël Lopez Kaufman, Dhruv Shah, Weize Kong,  
 623 Abhishek Bapna, Gellért Weisz, Emma Dunleavy, Praneet Dutta, Tianqi Liu, Rahma Chaabouni,  
 624 Carolina Parada, Marcus Wu, Alexandra Belias, Alessandro Bissacco, Stanislav Fort, Li Xiao,  
 625 Fantine Huot, Chris Knutsen, Yochai Blau, Gang Li, Jennifer Prendki, Juliette Love, Yinlan  
 626 Chow, Pichi Charoenpanit, Hidetoshi Shimokawa, Vincent Coriou, Karol Gregor, Tomas Izo, Arjun  
 627 Akula, Mario Pinto, Chris Hahn, Dominik Paulus, Jiaxian Guo, Neha Sharma, Cho-Jui Hsieh,  
 628 Aadae Chukwuka, Kazuma Hashimoto, Nathalie Rauschmayr, Ling Wu, Christof Angermueller,  
 629 Yulong Wang, Sebastian Gerlach, Michael Pliskin, Daniil Mirylenka, Min Ma, Lexi Baugher,  
 630 Bryan Gale, Shaan Bijwadia, Nemanja Rakićević, David Wood, Jane Park, Chung-Ching Chang,  
 631 Babi Seal, Chris Tar, Kacper Krasowiak, Yiwen Song, Georgi Stephanov, Gary Wang, Marcello  
 632 Maggioni, Stein Xudong Lin, Felix Wu, Shachi Paul, Zixuan Jiang, Shubham Agrawal, Bilal Piot,  
 633 Alex Feng, Cheolmin Kim, Tulsee Doshi, Jonathan Lai, Chuqiao Xu, Sharad Vikram, Ciprian  
 634 Chelba, Sebastian Krause, Vincent Zhuang, Jack Rae, Timo Denk, Adrian Collister, Lotte Weerts,  
 635 Xianghong Luo, Yifeng Lu, Håvard Garnes, Nitish Gupta, Terry Spitz, Avinatan Hassidim, Lihao  
 636 Liang, Izhak Shafran, Peter Humphreys, Kenny Vassigh, Phil Wallis, Virat Shejwalkar, Nicolas  
 637 Perez-Nieves, Rachel Hornung, Melissa Tan, Beka Westberg, Andy Ly, Richard Zhang, Brian  
 638 Farris, Jongbin Park, Alec Kosik, Zeynep Cankara, Andrii Maksai, Yunhan Xu, Albin Cassirer,  
 639 Sergi Caelles, Abbas Abdolmaleki, Mencher Chiang, Alex Fabrikant, Shravya Shetty, Luheng  
 640 He, Mai Giménez, Hadi Hashemi, Sheena Panthaplackel, Yana Kulizhskaya, Salil Deshmukh,  
 641 Daniele Pighin, Robin Alazard, Disha Jindal, Seb Noury, Pradeep Kumar S, Siyang Qin, Xerxes  
 642 Dotiwala, Stephen Spencer, Mohammad Babaeizadeh, Blake JianHang Chen, Vaibhav Mehta,  
 643 Jennie Lees, Andrew Leach, Penporn Koanantakool, Ilia Akolzin, Ramona Comanescu, Junwhan  
 644 Ahn, Alexey Svyatkovskiy, Basil Mustafa, David D'Ambrosio, Shiva Mohan Reddy Garlapati,  
 645 Pascal Lamblin, Alekh Agarwal, Shuang Song, Pier Giuseppe Sessa, Pauline Coquinot, John  
 646 Maggs, Hussain Masoom, Divya Pitta, Yaqing Wang, Patrick Morris-Suzuki, Billy Porter, Johnson  
 647 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 Vušak, Blake Hechtman,  
 Sheela Goenka, Avital Zipori, Kareem Ayoub, Ashok Papat, Trilok Acharya, Luo Yu, Dawn  
 Bloxwich, Hugo Song, Paul Roit, Haiqiong Li, Aviel Boag, Nigamaa Nayakanti, Bilva Chandra,



648 Tianli Ding, Aahil Mehta, Cath Hope, Jiageng Zhang, Idan Heimlich Shtacher, Kartikeya Badola,  
649 Ryo Nakashima, Andrei Sozanschi, Iulia Comşa, Ante Žužul, Emily Caveness, Julian Odell,  
650 Matthew Watson, Dario de Cesare, Phillip Lippe, Derek Lockhart, Siddharth Verma, Huizhong  
651 Chen, Sean Sun, Lin Zhuo, Aditya Shah, Prakhar Gupta, Alex Muzio, Ning Niu, Amir Zait,  
652 Abhinav Singh, Meenu Gaba, Fan Ye, Prajit Ramachandran, Mohammad Saleh, Raluca Ada Popa,  
653 Ayush Dubey, Frederick Liu, Sara Javanmardi, Mark Epstein, Ross Hemsley, Richard Green,  
654 Nishant Ranka, Eden Cohen, Chuyuan Kelly Fu, Sanjay Ghemawat, Jed Borovik, James Martens,  
655 Anthony Chen, Pranav Shyam, André Susano Pinto, Ming-Hsuan Yang, Alexandru Țifrea, David  
656 Du, Boqing Gong, Ayushi Agarwal, Seungyeon Kim, Christian Frank, Saloni Shah, Xiaodan Song,  
657 Zhiwei Deng, Ales Mikhalap, Kleopatra Chatzprimou, Timothy Chung, Toni Creswell, Susan  
658 Zhang, Yennie Jun, Carl Lebsack, Will Truong, Slavica Andačić, Itay Yona, Marco Fornoni, Rong  
659 Rong, Serge Toropov, Afzal Shama Soudagar, Andrew Audibert, Salah Zaiem, Zaheer Abbas,  
660 Andrei Rusu, Sahitya Potluri, Shitao Weng, Anastasios Kementsietsidis, Anton Tsitsulin, Daiyi  
661 Peng, Natalie Ha, Sanil Jain, Tejasi Latkar, Simeon Ivanov, Cory McLean, Anirudh GP, Rajesh  
662 Venkataraman, Canoe Liu, Dilip Krishnan, Joel D'sa, Roey Yogev, Paul Collins, Benjamin Lee,  
663 Lewis Ho, Carl Doersch, Gal Yona, Shawn Gao, Felipe Tiengo Ferreira, Adnan Ozturel, Hannah  
664 Muckenhirn, Ce Zheng, Gargi Balasubramaniam, Mudit Bansal, George van den Driessche, Sivan  
665 Eiger, Salem Haykal, Vedant Misra, Abhimanyu Goyal, Danilo Martins, Gary Leung, Jonas  
666 Valfridsson, Four Flynn, Will Bishop, Chenxi Pang, Yoni Halpern, Honglin Yu, Lawrence Moore,  
667 Yuvein, Zhu, Sridhar Thiagarajan, Yoel Drori, Zhisheng Xiao, Lucio Dery, Rolf Jagerman, Jing  
668 Lu, Eric Ge, Vaibhav Aggarwal, Arjun Khare, Vinh Tran, Oded Elyada, Ferran Alet, James Rubin,  
669 Ian Chou, David Tian, Libin Bai, Lawrence Chan, Lukasz Lew, Karolis Misiunas, Taylan Bilal,  
670 Aniket Ray, Sindhu Raghuram, Alex Castro-Ros, Viral Carpenter, C. J. Zheng, Michael Kilgore,  
671 Josef Broder, Emily Xue, Praveen Kallakuri, Dheeru Dua, Nancy Yuen, Steve Chien, John Schultz,  
672 Saurabh Agrawal, Reut Tsarfaty, Jingcao Hu, Ajay Kannan, Dror Marcus, Nisarg Kothari, Baochen  
673 Sun, Ben Horn, Matko Bošnjak, Ferjad Naeem, Dean Hirsch, Lewis Chiang, Boya Fang, Jie Han,  
674 Qifei Wang, Ben Hora, Antoine He, Mario Lučić, Beer Changpinyo, Anshuman Tripathi, John  
675 Youssef, Chester Kwak, Philippe Schlattner, Cat Graves, Rémi Leblond, Wenjun Zeng, Anders  
676 Andreassen, Gabriel Rasskin, Yue Song, Eddie Cao, Junhyuk Oh, Matt Hoffman, Wojtek Skut,  
677 Yichi Zhang, Jon Stritar, Xingyu Cai, Saarthak Khanna, Kathie Wang, Shriya Sharma, Christian  
678 Reisswig, Younghoon Jun, Aman Prasad, Tatiana Sholokhova, Preeti Singh, Adi Gerzi Rosenthal,  
679 Anian Ruoss, Françoise Beaufays, Sean Kirmani, Dongkai Chen, Johan Schalkwyk, Jonathan  
680 Herzig, Been Kim, Josh Jacob, Damien Vincent, Adrian N. Reyes, Ivana Balazevic, Léonard  
681 Hussenot, Jon Schneider, Parker Barnes, Luis Castro, Spandana Raj Babbula, Simon Green,  
682 Serkan Cabi, Nico Duduta, Danny Driess, Rich Galt, Noam Velan, Junjie Wang, Hongyang Jiao,  
683 Matthew Mauger, Du Phan, Miteyan Patel, Vlado Galic, Jerry Chang, Eyal Marcus, Matt Harvey,  
684 Julian Salazar, Elahe Dabir, Suraj Satishkumar Sheth, Amol Mandhane, Hanie Sedghi, Jeremiah  
685 Willcock, Amir Zandieh, Shruthi Prabhakara, Aida Amini, Antoine Miech, Victor Stone, Massimo  
686 Nicosia, Paul Niemczyk, Ying Xiao, Lucy Kim, Sławek Kwasiborski, Vikas Verma, Ada Maksutaj  
687 Oflazer, Christoph Hirschall, Peter Sung, Lu Liu, Richard Everett, Michiel Bakker, Ágoston  
688 Weisz, Yufei Wang, Vivek Sampathkumar, Uri Shaham, Bibo Xu, Yasemin Altun, Mingqiu Wang,  
689 Takaaki Saeki, Guanjie Chen, Emanuel Taropa, Shanthal Vasanth, Sophia Austin, Lu Huang,  
690 Goran Petrovic, Qingyun Dou, Daniel Golovin, Grigory Rozhdestvenskiy, Allie Culp, Will Wu,  
691 Motoki Sano, Divya Jain, Julia Proskurnia, Sébastien Cevey, Alejandro Cruzado Ruiz, Piyush Patil,  
692 Mahdi Mirzazadeh, Eric Ni, Javier Snaider, Lijie Fan, Alexandre Fréchet, A. J. Pierigiovanni,  
693 Shariq Iqbal, Kenton Lee, Claudio Fantacci, Jinwei Xing, Lisa Wang, Alex Irpan, David Raposo,  
694 Yi Luan, Zhuoyuan Chen, Harish Ganapathy, Kevin Hui, Jiazhong Nie, Isabelle Guyon, Heming  
695 Ge, Roopali Vij, Hui Zheng, Dayeong Lee, Alfonso Castaño, Khuslen Baatarsukh, Gabriel  
696 Ibagon, Alexandra Chronopoulou, Nicholas FitzGerald, Shashank Viswanadha, Safeen Huda,  
697 Rivka Moroshko, Georgi Stoyanov, Prateek Kolhar, Alain Vaucher, Ishaan Watts, Adhi Kuncoro,  
698 Henryk Michalewski, Satish Kambala, Bat-Orgil Batsaikhan, Alek Andreev, Irina Jurenka, Maigo  
699 Le, Qihang Chen, Wael Al Jishi, Sarah Chakera, Zhe Chen, Aditya Kini, Vikas Yadav, Aditya  
700 Siddhant, Iliia Labzovsky, Balaji Lakshminarayanan, Carrie Grimes Bostock, Pankil Botadra,  
701 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

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

756 Ben Limonchik, Wei Fan, Jim Stephan, Ian Stewart-Binks, Ryan Kappedal, Tong He, Sarah Cogan,  
757 Romina Datta, Tong Zhou, Jiayu Ye, Leandro Kieliger, Ana Ramalho, Kyle Kastner, Fabian  
758 Mentzer, Wei-Jen Ko, Arun Suggala, Tianhao Zhou, Shiraz Butt, Hana Strejček, Lior Belenki,  
759 Subhashini Venugopalan, Mingyang Ling, Evgenii Eltyshev, Yunxiao Deng, Geza Kovacs, Mukund  
760 Raghavachari, Hanjun Dai, Tal Schuster, Steven Schwarcz, Richard Nguyen, Arthur Nguyen, Gavin  
761 Buttimore, Shrestha Basu Mallick, Sudeep Gandhe, Seth Benjamin, Michal Jastrzebski, Le Yan,  
762 Sugato Basu, Chris Apps, Isabel Edkins, James Allingham, Immanuel Odisho, Tomas Kocisky,  
763 Jewel Zhao, Linting Xue, Apoorv Reddy, Chrysovalantis Anastasiou, Aviel Atias, Sam Redmond,  
764 Kieran Milan, Nicolas Heess, Herman Schmit, Allan Dafoe, Daniel Andor, Tynan Gangwani,  
765 Anca Dragan, Sheng Zhang, Ashyana Kachra, Gang Wu, Siyang Xue, Kevin Aydin, Siqi Liu,  
766 Yuxiang Zhou, Mahan Malihi, Austin Wu, Siddharth Gopal, Candice Schumann, Peter Stys,  
767 Alek Wang, Mirek Olšák, Dangyi Liu, Christian Schallhart, Yiran Mao, Demetra Brady, Hao  
768 Xu, Tomas Mery, Chawin Sitawarin, Siva Velusamy, Tom Cobley, Alex Zhai, Christian Walder,  
769 Nitzan Katz, Ganesh Jawahar, Chinmay Kulkarni, Antoine Yang, Adam Paszke, Yinan Wang,  
770 Bogdan Damoc, Zalán Borsos, Ray Smith, Jinning Li, Mansi Gupta, Andrei Kapishnikov, Sushant  
771 Prakash, Florian Luisier, Rishabh Agarwal, Will Grathwohl, Kuangyuan Chen, Kehang Han,  
772 Nikhil Mehta, Andrew Over, Shekoofeh Azizi, Lei Meng, Niccolò Dal Santo, Kelvin Zheng, Jane  
773 Shapiro, Igor Petrovski, Jeffrey Hui, Amin Ghafouri, Jasper Snoek, James Qin, Mandy Jordan,  
774 Caitlin Sikora, Jonathan Malmaud, Yuheng Kuang, Aga Świetlik, Ruoxin Sang, Chongyang Shi,  
775 Leon Li, Andrew Rosenberg, Shubin Zhao, Andy Crawford, Jan-Thorsten Peter, Yun Lei, Xavier  
776 Garcia, Long Le, Todd Wang, Julien Amelot, Dave Orr, Praneeth Kacham, Dana Alon, Gladys  
777 Tyen, Abhinav Arora, James Lyon, Alex Kurakin, Mimi Ly, Theo Guidroz, Zhipeng Yan, Rina  
778 Panigrahy, Pingmei Xu, Thais Kagohara, Yong Cheng, Eric Noland, Jinhyuk Lee, Jonathan Lee,  
779 Cathy Yip, Maria Wang, Efrat Nehoran, Alexander Bykovsky, Zhihao Shan, Ankit Bhagatwala,  
780 Chaochao Yan, Jie Tan, Guillermo Garrido, Dan Ethier, Nate Hurley, Grace Vesom, Xu Chen,  
781 Siyuan Qiao, Abhishek Nayyar, Julian Walker, Paramjit Sandhu, Mihaela Rosca, Danny Swisher,  
782 Mikhail Dektiarev, Josh Dillon, George-Cristian Muraru, Manuel Tragut, Artiom Myaskovsky,  
783 David Reid, Marko Velic, Owen Xiao, Jasmine George, Mark Brand, Jing Li, Wenhao Yu, Shane  
784 Gu, Xiang Deng, François-Xavier Aubet, Soheil Hassas Yeganeh, Fred Alcober, Celine Smith,  
785 Trevor Cohn, Kay McKinney, Michael Tschannen, Ramesh Sampath, Gowoon Cheon, Liangchen  
786 Luo, Luyang Liu, Jordi Orbay, Hui Peng, Gabriela Botea, Xiaofan Zhang, Charles Yoon, Cesar  
787 Magalhaes, Paweł Stradomski, Ian Mackinnon, Steven Hemingray, Kumar Venkatesan, Rhys  
788 May, Jaeyoun Kim, Alex Druinsky, Jingchen Ye, Zheng Xu, Terry Huang, Jad Al Abdallah, Adil  
789 Dostmohamed, Rachana Fellingner, Tsendsuren Munkhdalai, Akanksha Maurya, Peter Garst, Yin  
790 Zhang, Maxim Krikun, Simon Bucher, Aditya Srikanth Veerubhotla, Yaxin Liu, Sheng Li, Nishesh  
791 Gupta, Jakub Adamek, Hanwen Chen, Bennett Orlando, Aleksandr Zaks, Joost van Amersfoort,  
792 Josh Camp, Hui Wan, HyunJeong Choe, Zhichun Wu, Kate Olszewska, Weiren Yu, Archita Vadali,  
793 Martin Scholz, Daniel De Freitas, Jason Lin, Amy Hua, Xin Liu, Frank Ding, Yichao Zhou, Boone  
794 Severson, Katerina Tsihlias, Samuel Yang, Tammo Spalink, Varun Yerram, Helena Pankov, Rory  
795 Blevins, Ben Vargas, Sarthak Jauhari, Matt Miecnikowski, Ming Zhang, Sandeep Kumar, Clement  
796 Farabet, Charline Le Lan, Sebastian Flennerhag, Yonatan Bitton, Ada Ma, Arthur Bražinskas,  
797 Eli Collins, Niharika Ahuja, Sneha Kudugunta, Anna Bortsova, Minh Giang, Wanzheng Zhu,  
798 Ed Chi, Scott Lundberg, Alexey Stern, Subha Puttagunta, Jing Xiong, Xiao Wu, Yash Pande,  
799 Amit Jhinal, Daniel Murphy, Jon Clark, Marc Brockschmidt, Maxine Deines, Kevin R. McKee,  
800 Dan Bahir, Jiajun Shen, Minh Truong, Daniel McDuff, Andrea Gesmundo, Edouard Rosseel,  
801 Bowen Liang, Ken Caluwaerts, Jessica Hamrick, Joseph Kready, Mary Cassin, Rishikesh Ingale,  
802 Li Lao, Scott Pollom, Yifan Ding, Wei He, Lizzetth Bellot, Joana Iljazi, Ramya Sree Boppana,  
803 Shan Han, Tara Thompson, Amr Khalifa, Anna Bulanova, Blagoj Mitrevski, Bo Pang, Emma  
804 Cooney, Tian Shi, Rey Coaguila, Tamar Yakar, Marc aurelio Ranzato, Nikola Momchev, Chris  
805 Rawles, Zachary Charles, Young Maeng, Yuan Zhang, Rishabh Bansal, Xiaokai Zhao, Brian  
806 Albert, Yuan Yuan, Sudheendra Vijayanarasimhan, Roy Hirsch, Vinay Ramasesh, Kiran Vodrahalli,  
807 Xingyu Wang, Arushi Gupta, D. J. Strouse, Jianmo Ni, Roma Patel, Gabe Taubman, Zhouyuan  
808 Huo, Dero Gharibian, Marianne Monteiro, Hoi Lam, Shobha Vasudevan, Aditi Chaudhary, Isabela  
809 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,

810 Sameera Ponda, Simon Tong, Michelle Liu, Dan Malkin, Khalid Salama, Yuankai Chen, Rohan  
811 Anil, Anand Rao, Rigel Swavely, Misha Bilenko, Nina Anderson, Tat Tan, Jing Xie, Xing Wu,  
812 Lijun Yu, Oriol Vinyals, Andrey Ryabtsev, Rumen Dangovski, Kate Baumli, Daniel Keysers,  
813 Christian Wright, Zoe Ashwood, Betty Chan, Artem Shtefan, Yaohui Guo, Ankur Bapna, Radu  
814 Soricut, Steven Pecht, Sabela Ramos, Rui Wang, Jiahao Cai, Trieu Trinh, Paul Barham, Linda  
815 Friso, Eli Stickgold, Xiangzhuo Ding, Siamak Shakeri, Diego Ardila, Eleftheria Briakou, Phil  
816 Culliton, Adam Raveret, Jingyu Cui, David Saxton, Subhrajit Roy, Javad Azizi, Pengcheng Yin,  
817 Lucia Loher, Andrew Bunner, Min Choi, Faruk Ahmed, Eric Li, Yin Li, Shengyang Dai, Michael  
818 Elabd, Sriram Ganapathy, Shivani Agrawal, Yiqing Hua, Paige Kunkle, Sujeewan Rajayogam, Arun  
819 Ahuja, Arthur Conmy, Alex Vasiloff, Parker Beak, Christopher Yew, Jayaram Mudigonda, Bartek  
820 Wydrowski, Jon Blanton, Zhengdong Wang, Yann Dauphin, Zhuo Xu, Martin Polacek, Xi Chen,  
821 Hexiang Hu, Pauline Sho, Markus Kunesch, Mehdi Hafezi Manshadi, Eliza Rutherford, Bo Li,  
822 Sissie Hsiao, Iain Barr, Alex Tudor, Matija Kecman, Arsha Nagrani, Vladimir Pchelin, Martin  
823 Sundermeyer, Aishwarya P. S., Abhijit Karmarkar, Yi Gao, Grishma Chole, Olivier Bachem, Isabel  
824 Gao, Arturo BC, Matt Dibb, Mauro Verzetti, Felix Hernandez-Campos, Yana Lunts, Matthew  
825 Johnson, Julia Di Trapani, Raphael Koster, Idan Brusilovsky, Binbin Xiong, Megha Mohabey, Han  
826 Ke, Joe Zou, Tea Sabolić, Víctor Campos, John Palowitch, Alex Morris, Linhai Qiu, Pranavaraj  
827 Ponnuramu, Fangtao Li, Vivek Sharma, Kiranbir Sodhia, Kaan Tekelioglu, Aleksandr Chuklin,  
828 Madhavi Yenugula, Erika Gemzer, Theofilos Strinopoulos, Sam El-Husseini, Huiyu Wang, Yan  
829 Zhong, Edouard Leurent, Paul Natsev, Weijun Wang, Dre Mahaarachchi, Tao Zhu, Songyou Peng,  
830 Sami Alabed, Cheng-Chun Lee, Anthony Brohan, Arthur Szlam, G. S. Oh, Anton Kovsharov, Jenny  
831 Lee, Renee Wong, Megan Barnes, Gregory Thornton, Felix Gimeno, Omer Levy, Martin Sevenich,  
832 Melvin Johnson, Jonathan Mallinson, Robert Dadashi, Ziyue Wang, Qingchun Ren, Preethi Lahoti,  
833 Arka Dhar, Josh Feldman, Dan Zheng, Thatcher Ulrich, Liviu Panait, Michiel Blokzijl, Cip  
834 Baetu, Josip Matak, Jitendra Harlalka, Maulik Shah, Tal Marian, Daniel von Dincklage, Cosmo  
835 Du, Ruy Ley-Wild, Bethanie Brownfield, Max Schumacher, Yury Stuken, Shadi Noghabi, Sonal  
836 Gupta, Xiaoqi Ren, Eric Malmi, Felix Weissenberger, Blanca Huergo, Maria Bauza, Thomas  
837 Lampe, Arthur Douillard, Mojtaba Seyedhosseini, Roy Frostig, Zoubin Ghahramani, Kelvin  
838 Nguyen, Kashyap Krishnakumar, Chengxi Ye, Rahul Gupta, Alireza Nazari, Robert Geirhos, Pete  
839 Shaw, Ahmed Eleryan, Dima Damen, Jennimaria Palomaki, Ted Xiao, Qiyin Wu, Quan Yuan,  
840 Phoenix Meadowlark, Matthew Bilotti, Raymond Lin, Mukund Sridhar, Yannick Schroecker,  
841 Da-Woon Chung, Jincheng Luo, Trevor Strohman, Tianlin Liu, Anne Zheng, Jesse Emond, Wei  
842 Wang, Andrew Lampinen, Toshiyuki Fukuzawa, Folawiyi Campbell-Ajala, Monica Roy, James  
843 Lee-Thorp, Lily Wang, Iftekhar Naim, Tony, Nguy\~ên, Guy Bensky, Aditya Gupta, Dominika  
844 Rogozińska, Justin Fu, Thanumalayan Sankaranarayanan Pillai, Petar Veličković, Shahar Drath,  
845 Philipp Neubeck, Vaibhav Tulsyan, Arseniy Klimovskiy, Don Metzler, Sage Stevens, Angel Yeh,  
846 Junwei Yuan, Tianhe Yu, Kelvin Zhang, Alec Go, Vincent Tsang, Ying Xu, Andy Wan, Isaac  
847 Galatzer-Levy, Sam Sobell, Abodunrinwa Toki, Elizabeth Salesky, Wenlei Zhou, Diego Antognini,  
848 Sholto Douglas, Shimu Wu, Adam Lelkes, Frank Kim, Paul Cavallaro, Ana Salazar, Yuchi Liu,  
849 James Besley, Tiziana Refice, Yiling Jia, Zhang Li, Michal Sokolik, Arvind Kannan, Jon Simon,  
850 Jo Chick, Avia Aharon, Meet Gandhi, Mayank Daswani, Keyvan Amiri, Vighnesh Birodkar, Abe  
851 Ittycheriah, Peter Grabowski, Oscar Chang, Charles Sutton, Zhixin, Lai, Umesh Telang, Susie  
852 Sargsyan, Tao Jiang, Raphael Hoffmann, Nicole Brichtova, Matteo Hessel, Jonathan Halcrow,  
853 Sammy Jerome, Geoff Brown, Alex Tomala, Elena Buchatskaya, Dian Yu, Sachit Menon, Pol  
854 Moreno, Yuguo Liao, Vicky Zayats, Luming Tang, S. Q. Mah, Ashish Shenoy, Alex Siegman,  
855 Majid Hadian, Okwan Kwon, Tao Tu, Nima Khajehpour, Ryan Foley, Parisa Haghani, Zhongru  
856 Wu, Vaishakh Keshava, Khyatti Gupta, Tony Bruguier, Rui Yao, Danny Karmon, Luisa Zintgraf,  
857 Zhicheng Wang, Enrique Piqueras, Junehyuk Jung, Jenny Brennan, Diego Machado, Marissa  
858 Giustina, M. H. Tessler, Kamyu Lee, Qiao Zhang, Joss Moore, Kaspar Daugaard, Alexander  
859 Frömmgen, Jennifer Beattie, Fred Zhang, Daniel Kasenberg, Ty Geri, Danfeng Qin, Gaurav Singh  
860 Tomar, Tom Ouyang, Tianli Yu, Luwei Zhou, Rajiv Mathews, Andy Davis, Yaoyiran Li, Jai  
861 Gupta, Damion Yates, Linda Deng, Elizabeth Kemp, Ga-Young Joung, Sergei Vassilvitskii, Mandy  
862 Guo, Pallavi LV, Dave Dopson, Sami Lachgar, Lara McConnaughey, Himadri Choudhury, Dragos  
863 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  
Žanić, Tom Schaul, Manolis Delakis, Tejas Iyer, Guanyu Wang, Harman Singh, Abdelrahman  
Abdelhamed, Tara Thomas, Siddhartha Brahma, Hilal Dib, Naveen Kumar, Wenxuan Zhou, Liang

864 Bai, Pushkar Mishra, Jiao Sun, Valentin Anklin, Roykronk Sukkerd, Lauren Agubuzu, Anton  
865 Briukhov, Anmol Gulati, Maximilian Sieb, Fabio Pardo, Sara Nasso, Junquan Chen, Kexin Zhu,  
866 Tiberiu Sosea, Alex Goldin, Keith Rush, Spurthi Amba Hombaiah, Andreas Noever, Allan Zhou,  
867 Sam Haves, Mary Phuong, Jake Ades, Yi-ting Chen, Lin Yang, Joseph Pagadora, Stan Bileschi,  
868 Victor Cotruta, Rachel Saputro, Arijit Pramanik, Sean Ammirati, Dan Garrette, Kevin Villela, Tim  
869 Blyth, Canfer Akbulut, Neha Jha, Alban Rrustemi, Arissa Wongpanich, Chirag Nagpal, Yonghui  
870 Wu, Morgane Rivièrè, Sergey Kishchenko, Pranesh Srinivasan, Alice Chen, Animesh Sinha, Trang  
871 Pham, Bill Jia, Tom Hennigan, Anton Bakalov, Nithya Attaluri, Drew Garmon, Daniel Rodriguez,  
872 Dawid Wegner, Wenhao Jia, Evan Senter, Noah Fiedel, Denis Petek, Yuchuan Liu, Cassidy Hardin,  
873 Harshal Tushar Lehri, Joao Carreira, Sara Smoot, Marcel Prasetya, Nami Akazawa, Anca Stefanoiu,  
874 Chia-Hua Ho, Anelia Angelova, Kate Lin, Min Kim, Charles Chen, Marcin Sieniek, Alice Li,  
875 Tongfei Guo, Sorin Baltateanu, Pouya Tafti, Michael Wunder, Nadav Olmert, Divyansh Shukla,  
876 Jingwei Shen, Neel Kovelamudi, Balaji Venkatraman, Seth Neel, Romal Thoppilan, Jerome Connor,  
877 Frederik Benzing, Axel Stjerngren, Golnaz Ghiasi, Alex Polozov, Joshua Howland, Theophane  
878 Weber, Justin Chiu, Ganesh Poomal Girirajan, Andreas Terzis, Pidong Wang, Fangda Li, Yoav Ben  
879 Shalom, Dinesh Tewari, Matthew Denton, Roece Aharoni, Norbert Kalb, Heri Zhao, Junlin Zhang,  
880 Angelos Filos, Matthew Rahtz, Lalit Jain, Connie Fan, Vitor Rodrigues, Ruth Wang, Richard  
881 Shin, Jacob Austin, Roman Ring, Mariella Sanchez-Vargas, Mehadi Hassen, Ido Kessler, Uri Alon,  
882 Gufeng Zhang, Wenhui Chen, Yenai Ma, Xiance Si, Le Hou, Azalia Mirhoseini, Marc Wilson,  
883 Geoff Bacon, Becca Roelofs, Lei Shu, Gautam Vasudevan, Jonas Adler, Artur Dwornik, Tayfun  
884 Terzi, Matt Lawlor, Harry Askham, Mike Bernico, Xuanyi Dong, Chris Hidey, Kevin Kilgour, Gaël  
885 Liu, Surya Bhupatiraju, Luke Leonhard, Siqi Zuo, Partha Talukdar, Qing Wei, Aliaksei Severyn,  
886 Vít Listík, Jong Lee, Aditya Tripathi, S. K. Park, Yossi Matias, Hao Liu, Alex Ruiz, Rajesh  
887 Jayaram, Jackson Tolins, Pierre Marcenac, Yiming Wang, Bryan Seybold, Henry Prior, Deepak  
888 Sharma, Jack Weber, Mikhail Sirotenko, Yunhsuan Sung, Dayou Du, Ellie Pavlick, Stefan Zinke,  
889 Markus Freitag, Max Dylla, Montse Gonzalez Arenas, Natan Potikha, Omer Goldman, Connie  
890 Tao, Rachita Chhaparia, Maria Voitovich, Pawan Dogra, Andrija Ražnatović, Zak Tsai, Chong  
891 You, Oleaser Johnson, George Tucker, Chenjie Gu, Jae Yoo, Maryam Majzoubi, Valentin Gabeur,  
892 Bahram Raad, Rocky Rhodes, Kashyap Kolipaka, Heidi Howard, Geta Sampemane, Benny Li,  
893 Chulayuth Asawaroengchai, Duy Nguyen, Chiyuan Zhang, Timothee Cour, Xinxin Yu, Zhao Fu,  
894 Joe Jiang, Po-Sen Huang, Gabriela Surita, Iñaki Iturrate, Yael Karov, Michael Collins, Martin  
895 Baeuml, Fabian Fuchs, Shilpa Shetty, Swaroop Ramaswamy, Sayna Ebrahimi, Qiuchen Guo,  
896 Jeremy Shar, Gabe Barth-Maron, Sravanti Addepalli, Bryan Richter, Chin-Yi Cheng, Eugénie  
897 Rives, Fei Zheng, Johannes Griesser, Nishanth Dikkala, Yoel Zeldes, Ilkin Safarli, Dipanjan Das,  
898 Himanshu Srivastava, Sadh MNM Khan, Xin Li, Aditya Pandey, Larisa Markeeva, Dan Belov, Qiqi  
899 Yan, Mikołaj Rybiński, Tao Chen, Megha Nawhal, Michael Quinn, Vineetha Govindaraj, Sarah  
900 York, Reed Roberts, Roopal Garg, Namrata Godbole, Jake Abernethy, Anil Das, Lam Nguyen  
901 Thiet, Jonathan Tompson, John Nham, Neera Vats, Ben Caine, Wesley Helmholz, Francesco  
902 Pongetti, Yeongil Ko, James An, Clara Huiyi Hu, Yu-Cheng Ling, Julia Pawar, Robert Leland,  
903 Keisuke Kinoshita, Waleed Khawaja, Marco Selvi, Eugene Ie, Danila Sinopalnikov, Lev Proleev,  
904 Nilesch Tripuraneni, Michele Bevilacqua, Seungji Lee, Clayton Sanford, Dan Suh, Dustin Tran,  
905 Jeff Dean, Simon Baumgartner, Jens Heitkaemper, Sagar Gubbi, Kristina Toutanova, Yichong Xu,  
906 Chandu Thekkath, Keran Rong, Palak Jain, Annie Xie, Yan Virin, Yang Li, Lubo Litchev, Richard  
907 Powell, Tarun Bharti, Adam Kraft, Nan Hua, Marissa Ikonomidis, Ayal Hitron, Sanjiv Kumar,  
908 Loic Matthey, Sophie Bridgers, Lauren Lax, Ishaan Malhi, Ondrej Skopek, Ashish Gupta, Jiawei  
909 Cao, Michelle Rasquinha, Siim Pöder, Wojciech Stokowiec, Nicholas Roth, Guowang Li, Michaël  
910 Sander, Joshua Kessinger, Vihan Jain, Edward Loper, Wonpyo Park, Michal Yarom, Liqun Cheng,  
911 Guru Guruganesh, Kanishka Rao, Yan Li, Catarina Barros, Mikhail Sushkov, Chun-Sung Ferng,  
912 Rohin Shah, Ophir Aharoni, Ravin Kumar, Tim McConnell, Peiran Li, Chen Wang, Fernando  
913 Pereira, Craig Swanson, Fayaz Jamil, Yan Xiong, Anitha Vijayakumar, Prakash Shroff, Kedar  
914 Soparkar, Jindong Gu, Livio Baldini Soares, Eric Wang, Kushal Majmundar, Aurora Wei, Kai  
915 Bailey, Nora Kassner, Chizu Kawamoto, Goran Žužić, Victor Gomes, Abhirut Gupta, Michael  
916 Guzman, Ishita Dasgupta, Xinyi Bai, Zhufeng Pan, Francesco Piccinno, Hadas Natalie Vogel,  
917 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, Aroonlok  
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

918 Panagopoulos, A. J. Piergiovanni, Yunfei Bai, George Powell, Deep Karkhanis, Trevor Yacovone,  
919 Petr Mitrichev, Joe Kovac, Dave Uthus, Amir Yazdanbakhsh, David Amos, Steven Zheng, Bing  
920 Zhang, Jin Miao, Bhuvana Ramabhadran, Soroush Radpour, Shantanu Thakoor, Josh Newlan, Oran  
921 Lang, Orion Jankowski, Shikhar Bharadwaj, Jean-Michel Sarr, Shereen Ashraf, Sneha Mondal, Jun  
922 Yan, Ankit Singh Rawat, Sarmishta Velury, Greg Kochanski, Tom Eccles, Franz Och, Abhanshu  
923 Sharma, Ethan Mahintorabi, Alex Gurney, Carrie Muir, Vered Cohen, Saksham Thakur, Adam  
924 Bloniarz, Asier Mujika, Alexander Pritzel, Paul Caron, Altaf Rahman, Fiona Lang, Yasumasa Onoe,  
925 Petar Sirkovic, Jay Hoover, Ying Jian, Pablo Duque, Arun Narayanan, David Soergel, Alex Haig,  
926 Loren Maggiore, Shyamal Buch, Josef Dean, Ilya Figotin, Igor Karpov, Shaleen Gupta, Denny  
927 Zhou, Muhuan Huang, Ashwin Vaswani, Christopher Semturs, Kaushik Shivakumar, Yu Watanabe,  
928 Vinodh Kumar Rajendran, Eva Lu, Yanhan Hou, Wenting Ye, Shikhar Vashishth, Nana Nti, Vytenis  
929 Sakenas, Darren Ni, Doug DeCarlo, Michael Bendersky, Sumit Bagri, Nacho Cano, Elijah Peake,  
930 Simon Tokumine, Varun Godbole, Carlos Guía, Tanya Lando, Vittorio Selo, Seher Ellis, Danny  
931 Tarlow, Daniel Gillick, Alessandro Epasto, Siddhartha Reddy Jonnalagadda, Meng Wei, Meiyang  
932 Xie, Ankur Taly, Michela Paganini, Mukund Sundararajan, Daniel Toyama, Ting Yu, Dessie  
933 Petrova, Aneesh Pappu, Rohan Agrawal, Senaka Buthpitiya, Justin Frye, Thomas Buschmann,  
934 Remi Crocker, Marco Tagliasacchi, Mengchao Wang, Da Huang, Sagi Perel, Brian Wieder, Hideto  
935 Kazawa, Weiyue Wang, Jeremy Cole, Himanshu Gupta, Ben Golan, Seojin Bang, Nitish Kulkarni,  
936 Ken Franko, Casper Liu, Doug Reid, Sid Dalmia, Jay Whang, Kevin Cen, Prasha Sundaram, Johan  
937 Ferret, Berivan Isik, Lucian Ionita, Guan Sun, Anna Shekhawat, Muqthar Mohammad, Philip  
938 Pham, Ronny Huang, Karthik Raman, Xingyi Zhou, Ross McIlroy, Austin Myers, Sheng Peng,  
939 Jacob Scott, Paul Covington, Sofia Erell, Pratik Joshi, João Gabriel Oliveira, Natasha Noy, Tajwar  
940 Nasir, Jake Walker, Vera Axelrod, Tim Dozat, Pu Han, Chun-Te Chu, Eugene Weinstein, Anand  
941 Shukla, Shreyas Chandrakaladharan, Petra Poklukar, Bonnie Li, Ye Jin, Prem Eruvbetine, Steven  
942 Hansen, Avigail Dabush, Alon Jacovi, Samrat Phatale, Chen Zhu, Steven Baker, Mo Shomrat, Yang  
943 Xiao, Jean Pouget-Abadie, Mingyang Zhang, Fanny Wei, Yang Song, Helen King, Yiling Huang,  
944 Yun Zhu, Ruoxi Sun, Juliana Vicente Franco, Chu-Cheng Lin, Sho Arora, Hui, Li, Vivian Xia,  
945 Luke Vilnis, Mariano Schain, Kaiz Alarakyia, Laurel Prince, Aaron Phillips, Caleb Habtegebriel,  
946 Luyao Xu, Huan Gui, Santiago Ontanon, Lora Aroyo, Karan Gill, Peggy Lu, Yash Katariya,  
947 Dhruv Madeka, Shankar Krishnan, Shubha Srinivas Raghvendra, James Freedman, Yi Tay, Gaurav  
948 Menghani, Peter Choy, Nishita Shetty, Dan Abolafia, Doron Kukliansky, Edward Chou, Jared  
949 Lichtarge, Ken Burke, Ben Coleman, Dee Guo, Larry Jin, Indro Bhattacharya, Victoria Langston,  
950 Yiming Li, Suyog Kotecha, Alex Yakubovich, Xinyun Chen, Petre Petrov, Tolly Powell, Yanzhang  
951 He, Corbin Quick, Kanav Garg, Dawsen Hwang, Yang Lu, Srinadh Bhojanapalli, Kristian Kjems,  
952 Ramin Mehran, Aaron Archer, Hado van Hasselt, Ashwin Balakrishna, J. K. Kearns, Meiqi Guo,  
953 Jason Riesa, Mikita Sazanovich, Xu Gao, Chris Sauer, Chengrun Yang, XiangHai Sheng, Thomas  
954 Jimma, Wouter Van Gansbeke, Vitaly Nikolaev, Wei Wei, Katie Millican, Ruizhe Zhao, Justin  
955 Snyder, Levent Bolelli, Maura O'Brien, Shawn Xu, Fei Xia, Wentao Yuan, Arvind Neelakantan,  
956 David Barker, Sachin Yadav, Hannah Kirkwood, Farooq Ahmad, Joel Wee, Jordan Grimstad, Boyu  
957 Wang, Matthew Wiethoff, Shane Settle, Miaosen Wang, Charles Blundell, Jingjing Chen, Chris  
958 Duvarney, Grace Hu, Olaf Ronneberger, Alex Lee, Yuanzhen Li, Abhishek Chakladar, Alena  
959 Butryna, Georgios Evangelopoulos, Guillaume Desjardins, Jonni Kanerva, Henry Wang, Averi  
960 Nowak, Nick Li, Alyssa Loo, Art Khurshudov, Laurent El Shafey, Nagabhushan Baddi, Karel Lenc,  
961 Yasaman Razeghi, Tom Lieber, Amer Sinha, Xiao Ma, Yao Su, James Huang, Asahi Ushio, Hanna  
962 Klimczak-Plucińska, Kareem Mohamed, J. D. Chen, Simon Osindero, Stav Ginzburg, Lampros  
963 Lamprou, Vasilisa Bashlovkina, Duc-Hieu Tran, Ali Khodaei, Ankit Anand, Yixian Di, Ramy  
964 Eskander, Manish Reddy Vuyyuru, Jasmine Liu, Aishwarya Kamath, Roman Goldenberg, Mathias  
965 Bellaiche, Juliette Pluto, Bill Rosgen, Hassan Mansoor, William Wong, Suhas Ganesh, Eric Bailey,  
966 Scott Baird, Dan Deutsch, Jinoo Baek, Xuhui Jia, Chansoo Lee, Abe Friesen, Nathaniel Braun, Kate  
967 Lee, Amayika Panda, Steven M. Hernandez, Duncan Williams, Jianqiao Liu, Ethan Liang, Arnaud  
968 Autef, Emily Pitler, Deepali Jain, Phoebe Kirk, Oskar Bunyan, Jaume Sanchez Elias, Tongxin Yin,  
969 Machel Reid, Aedan Pope, Nikita Putikhin, Bidisha Samanta, Sergio Guadarrama, Dahun Kim,  
970 Simon Rowe, Marcella Valentine, Geng Yan, Alex Salcianu, David Silver, Gan Song, Richa Singh,  
971 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

972 Iinuma, Dustin Zelle, Rohit Sathyanarayana, Andrea D’olimpio, Morgan Redshaw, Matt Ginsberg,  
 973 Ashwin Murthy, Mark Geller, Tatiana Matejovicova, Ayan Chakrabarti, Ryan Julian, Christine  
 974 Chan, Qiong Hu, Daniel Jarrett, Manu Agarwal, Jeshwanth Challagundla, Tao Li, Sandeep Tata,  
 975 Wen Ding, Maya Meng, Zhuyun Dai, Giulia Vezzani, Shefali Garg, Jannis Bulian, Mary Jasarevic,  
 976 Honglong Cai, Harish Rajamani, Adam Santoro, Florian Hartmann, Chen Liang, Bartek Perz,  
 977 Apoorv Jindal, Fan Bu, Sungyong Seo, Ryan Poplin, Adrian Goedeckemeyer, Badih Ghazi, Nikhil  
 978 Khadke, Leon Liu, Kevin Mather, Mingda Zhang, Ali Shah, Alex Chen, Jinliang Wei, Keshav  
 979 Shivam, Yuan Cao, Donghyun Cho, Angelo Scorza Scarpati, Michael Moffitt, Clara Barbu, Ivan  
 980 Jurin, Ming-Wei Chang, Hongbin Liu, Hao Zheng, Shachi Dave, Christine Kaeser-Chen, Xiaobin  
 981 Yu, Alvin Abdagic, Lucas Gonzalez, Yanping Huang, Peilin Zhong, Cordelia Schmid, Bryce  
 982 Petrini, Alex Wertheim, Jifan Zhu, Hoang Nguyen, Kaiyang Ji, Yanqi Zhou, Tao Zhou, Fangxiaoyu  
 983 Feng, Regev Cohen, David Rim, Shubham Milind Phal, Petko Georgiev, Ariel Brand, Yue Ma,  
 984 Wei Li, Somit Gupta, Chao Wang, Pavel Dubov, Jean Tarbouriech, Kingshuk Majumder, Huijian  
 985 Li, Norman Rink, Apurv Suman, Yang Guo, Yinghao Sun, Arun Nair, Xiaowei Xu, Mohamed  
 986 Elhawaty, Rodrigo Cabrera, Guangxing Han, Julian Eisenschlos, Junwen Bai, Yuqi Li, Yamini  
 987 Bansal, Thibault Sellam, Mina Khan, Hung Nguyen, Justin Mao-Jones, Nikos Parotsidis, Jake  
 988 Marcus, Cindy Fan, Roland Zimmermann, Yony Kochinski, Laura Graesser, Feryal Behbahani,  
 989 Alvaro Caceres, Michael Riley, Patrick Kane, Sandra Lefdal, Rob Willoughby, Paul Vicol, Lun  
 990 Wang, Shujian Zhang, Ashleah Gill, Yu Liang, Gautam Prasad, Soroosh Mariooryad, Mehran  
 991 Kazemi, Zifeng Wang, Kritika Muralidharan, Paul Voigtlaender, Jeffrey Zhao, Huanjie Zhou,  
 992 Nina D’Souza, Aditi Mavalankar, Séb Arnold, Nick Young, Obaid Sarvana, Chace Lee, Milad  
 993 Nasr, Tingting Zou, Seokhwan Kim, Lukas Haas, Kaushal Patel, Neslihan Bulut, David Parkinson,  
 994 Courtney Biles, Dmitry Kalashnikov, Chi Ming To, Aviral Kumar, Jessica Austin, Alex Greve,  
 995 Lei Zhang, Megha Goel, Yeqing Li, Sergey Yaroshenko, Max Chang, Abhishek Jindal, Geoff  
 996 Clark, Hagai Taitelbaum, Dale Johnson, Ofir Roval, Jeongwoo Ko, Anhad Mohanane, Christian  
 997 Schuler, Shenil Dodhia, Ruichao Li, Kazuki Osawa, Claire Cui, Peng Xu, Rushin Shah, Tao Huang,  
 998 Ela Gruzewska, Nathan Clement, Mudit Verma, Olcan Sercinoglu, Hai Qian, Viral Shah, Masa  
 999 Yamaguchi, Abhinav Modi, Takahiro Kosakai, Thomas Strohmman, Junhao Zeng, Beliz Gunel, Jun  
 1000 Qian, Austin Tarango, Krzysztof Jastrzebski, Robert David, Jyn Shan, Parker Schuh, Kunal Lad,  
 1001 Willi Gierke, Mukundan Madhavan, Xinyi Chen, Mark Kurzeja, Rebeca Santamaria-Fernandez,  
 1002 Dawn Chen, Alexandra Cordell, Yuri Chervonyi, Frankie Garcia, Nithish Kannan, Vincent Perot,  
 1003 Nan Ding, Shlomi Cohen-Ganor, Victor Lavrenko, Junru Wu, Georgie Evans, Cicero Nogueira dos  
 1004 Santos, Madhavi Sewak, Ashley Brown, Andrew Hard, Joan Puigcerver, Zeyu Zheng, Yizhong  
 1005 Liang, Evgeny Gladchenko, Reeve Ingle, Uri First, Pierre Sermanet, Charlotte Magister, Mihajlo  
 1006 Velimirović, Sashank Reddi, Susanna Ricco, Eirikur Agustsson, Hartwig Adam, Nir Levine, David  
 1007 Gaddy, Dan Holtmann-Rice, Xuanhui Wang, Ashutosh Sathe, Abhijit Guha Roy, Blaž Bratanič,  
 1008 Alen Carin, Harsh Mehta, Silvano Bonacina, Nicola De Cao, Mara Finkelstein, Verena Rieser,  
 1009 Xinyi Wu, Florent Alché, Dylan Scandinaro, Li Li, Nino Vieillard, Nikhil Sethi, Garrett Tanzer,  
 1010 Zhi Xing, Shibo Wang, Parul Bhatia, Gui Citovsky, Thomas Anthony, Sharon Lin, Tianze Shi,  
 1011 Shoshana Jakobovits, Gena Gibson, Raj Apte, Lisa Lee, Mingqing Chen, Arunkumar Byravan,  
 1012 Petros Maniatis, Kellie Webster, Andrew Dai, Pu-Chin Chen, Jiaqi Pan, Asya Fadeeva, Zach  
 1013 Gleicher, Thang Luong, and Niket Kumar Bhumihar. Gemini 2.5: Pushing the Frontier with  
 1014 Advanced Reasoning, Multimodality, Long Context, and Next Generation Agentic Capabilities,  
 1015 July 2025. URL <http://arxiv.org/abs/2507.06261>. arXiv:2507.06261 [cs].  
 1016  
 1017 DeepSeek-AI, Daya Guo, Dejian Yang, Haowei Zhang, Junxiao Song, Ruoyu Zhang, Runxin Xu,  
 1018 Qihao Zhu, Shirong Ma, Peiyi Wang, Xiao Bi, Xiaokang Zhang, Xingkai Yu, Yu Wu, Z. F. Wu,  
 1019 Zhibin Gou, Zhihong Shao, Zhuoshu Li, Ziyi Gao, Aixin Liu, Bing Xue, Bingxuan Wang, Bochao  
 1020 Wu, Bei Feng, Chengda Lu, Chenggang Zhao, Chengqi Deng, Chenyu Zhang, Chong Ruan,  
 1021 Damai Dai, Deli Chen, Dongjie Ji, Erhang Li, Fangyun Lin, Fucong Dai, Fuli Luo, Guangbo Hao,  
 1022 Guanting Chen, Guowei Li, H. Zhang, Han Bao, Hanwei Xu, Haocheng Wang, Honghui Ding,  
 1023 Huajian Xin, Huazuo Gao, Hui Qu, Hui Li, Jianzhong Guo, Jiashi Li, Jiawei Wang, Jingchang  
 1024 Chen, Jingyang Yuan, Junjie Qiu, Junlong Li, J. L. Cai, Jiaqi Ni, Jian Liang, Jin Chen, Kai Dong,  
 1025 Kai Hu, Kaige Gao, Kang Guan, Kexin Huang, Kuai Yu, Lean Wang, Lecong Zhang, Liang Zhao,  
 Litong Wang, Liyue Zhang, Lei Xu, Leyi Xia, Mingchuan Zhang, Minghua Zhang, Minghui Tang,  
 Meng Li, Miaojuan Wang, Mingming Li, Ning Tian, Panpan Huang, Peng Zhang, Qiancheng Wang,  
 Qinyu Chen, Qiushi Du, Ruiqi Ge, Ruisong Zhang, Ruizhe Pan, Runji Wang, R. J. Chen, R. L.  
 Jin, Ruyi Chen, Shanghao Lu, Shangyan Zhou, Shanhuang Chen, Shengfeng Ye, Shiyu Wang,  
 Shuiping Yu, Shunfeng Zhou, Shuting Pan, S. S. Li, Shuang Zhou, Shaoqing Wu, Shengfeng

- 1026 Ye, Tao Yun, Tian Pei, Tianyu Sun, T. Wang, Wangding Zeng, Wanjia Zhao, Wen Liu, Wenfeng  
1027 Liang, Wenjun Gao, Wenqin Yu, Wentao Zhang, W. L. Xiao, Wei An, Xiaodong Liu, Xiaohan  
1028 Wang, Xiaokang Chen, Xiaotao Nie, Xin Cheng, Xin Liu, Xin Xie, Xingchao Liu, Xinyu Yang,  
1029 Xinyuan Li, Xuecheng Su, Xuheng Lin, X. Q. Li, Xiangyue Jin, Xiaojin Shen, Xiaosha Chen,  
1030 Xiaowen Sun, Xiaoxiang Wang, Xinnan Song, Xinyi Zhou, Xianzu Wang, Xinxia Shan, Y. K. Li,  
1031 Y. Q. Wang, Y. X. Wei, Yang Zhang, Yanhong Xu, Yao Li, Yao Zhao, Yaofeng Sun, Yaohui Wang,  
1032 Yi Yu, Yichao Zhang, Yifan Shi, Yiliang Xiong, Ying He, Yishi Piao, Yisong Wang, Yixuan Tan,  
1033 Yiyang Ma, Yiyuan Liu, Yongqiang Guo, Yuan Ou, Yuduan Wang, Yue Gong, Yuheng Zou, Yujia  
1034 He, Yunfan Xiong, Yuxiang Luo, Yuxiang You, Yuxuan Liu, Yuyang Zhou, Y. X. Zhu, Yanhong  
1035 Xu, Yanping Huang, Yaohui Li, Yi Zheng, Yuchen Zhu, Yunxian Ma, Ying Tang, Yukun Zha,  
1036 Yuting Yan, Z. Z. Ren, Zehui Ren, Zhangli Sha, Zhe Fu, Zhean Xu, Zhenda Xie, Zhengyan Zhang,  
1037 Zhewen Hao, Zhicheng Ma, Zhigang Yan, Zhiyu Wu, Zihui Gu, Zijia Zhu, Zijun Liu, Zilin Li,  
1038 Ziwei Xie, Ziyang Song, Zizheng Pan, Zhen Huang, Zhipeng Xu, Zhongyu Zhang, and Zhen  
1039 Zhang. DeepSeek-R1: Incentivizing Reasoning Capability in LLMs via Reinforcement Learning,  
1040 January 2025. URL <http://arxiv.org/abs/2501.12948>. arXiv:2501.12948 [cs].
- 1041 Oliver Groth, Fabian B. Fuchs, Ingmar Posner, and Andrea Vedaldi. ShapeStacks: Learning Vision-  
1042 Based Physical Intuition for Generalised Object Stacking, July 2018. URL <http://arxiv.org/abs/1804.08018>. arXiv:1804.08018 [cs].
- 1043
- 1044 Justin Johnson, Bharath Hariharan, Laurens van der Maaten, Li Fei-Fei, C. Lawrence Zitnick,  
1045 and Ross Girshick. CLEVR: A Diagnostic Dataset for Compositional Language and Elementa-  
1046 ry Visual Reasoning, December 2016. URL <http://arxiv.org/abs/1612.06890>.  
1047 arXiv:1612.06890 [cs].
- 1048
- 1049 Panagiotis Michalatos and Sawako Kaijima. Millipede Plugin Grasshopper 3D, 2024. URL <https://www.creativemutation.com/millipede>.
- 1050
- 1051 OpenAI, Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni  
1052 Aleman, Diogo Almeida, Janko Altenschmidt, Sam Altman, Shyamal Anadkat, Red Avila, Igor  
1053 Babuschkin, Suchir Balaji, Valerie Balcom, Paul Baltescu, Haiming Bao, Mohammad Bavarian,  
1054 Jeff Belgum, Irwan Bello, Jake Berdine, Gabriel Bernadett-Shapiro, Christopher Berner, Lenny  
1055 Bogdonoff, Oleg Boiko, Madelaine Boyd, Anna-Luisa Brakman, Greg Brockman, Tim Brooks,  
1056 Miles Brundage, Kevin Button, Trevor Cai, Rosie Campbell, Andrew Cann, Brittany Carey, Chelsea  
1057 Carlson, Rory Carmichael, Brooke Chan, Che Chang, Fotis Chantzis, Derek Chen, Sully Chen,  
1058 Ruby Chen, Jason Chen, Mark Chen, Ben Chess, Chester Cho, Casey Chu, Hyung Won Chung,  
1059 Dave Cummings, Jeremiah Currier, Yunxing Dai, Cory Decareaux, Thomas Degry, Noah Deutsch,  
1060 Damien Deville, Arka Dhar, David Dohan, Steve Dowling, Sheila Dunning, Adrien Ecoffet, Atty  
1061 Eleti, Tyna Eloundou, David Farhi, Liam Fedus, Niko Felix, Simón Posada Fishman, Juston Forte,  
1062 Isabella Fulford, Leo Gao, Elie Georges, Christian Gibson, Vik Goel, Tarun Gogineni, Gabriel  
1063 Goh, Rapha Gontijo-Lopes, Jonathan Gordon, Morgan Grafstein, Scott Gray, Ryan Greene, Joshua  
1064 Gross, Shixiang Shane Gu, Yufei Guo, Chris Hallacy, Jesse Han, Jeff Harris, Yuchen He, Mike  
1065 Heaton, Johannes Heidecke, Chris Hesse, Alan Hickey, Wade Hickey, Peter Hoeschele, Brandon  
1066 Houghton, Kenny Hsu, Shengli Hu, Xin Hu, Joost Huizinga, Shantanu Jain, Shawn Jain, Joanne  
1067 Jang, Angela Jiang, Roger Jiang, Haozhun Jin, Denny Jin, Shino Jomoto, Billie Jonn, Heewoo  
1068 Jun, Tomer Kaftan, Łukasz Kaiser, Ali Kamali, Ingmar Kanitscheider, Nitish Shirish Keskar,  
1069 Tabarak Khan, Logan Kilpatrick, Jong Wook Kim, Christina Kim, Yongjik Kim, Jan Hendrik  
1070 Kirchner, Jamie Kiros, Matt Knight, Daniel Kokotajlo, Łukasz Kondraciuk, Andrew Kondrich,  
1071 Aris Konstantinidis, Kyle Kosic, Gretchen Krueger, Vishal Kuo, Michael Lampe, Ikai Lan, Teddy  
1072 Lee, Jan Leike, Jade Leung, Daniel Levy, Chak Ming Li, Rachel Lim, Molly Lin, Stephanie  
1073 Lin, Mateusz Litwin, Theresa Lopez, Ryan Lowe, Patricia Lue, Anna Makanju, Kim Malfacini,  
1074 Sam Manning, Todor Markov, Yaniv Markovski, Bianca Martin, Katie Mayer, Andrew Mayne,  
1075 Bob McGrew, Scott Mayer McKinney, Christine McLeavey, Paul McMillan, Jake McNeil, David  
1076 Medina, Aalok Mehta, Jacob Menick, Luke Metz, Andrey Mishchenko, Pamela Mishkin, Vinnie  
1077 Monaco, Evan Morikawa, Daniel Mossing, Tong Mu, Mira Murati, Oleg Murk, David Mély,  
1078 Ashvin Nair, Reiichiro Nakano, Rajeev Nayak, Arvind Neelakantan, Richard Ngo, Hyeonwoo  
1079 Noh, Long Ouyang, Cullen O’Keefe, Jakub Pachocki, Alex Paino, Joe Palermo, Ashley Pantuliano,  
Giambattista Parascandolo, Joel Parish, Emy Parparita, Alex Passos, Mikhail Pavlov, Andrew  
Peng, Adam Perelman, Filipe de Avila Belbute Peres, Michael Petrov, Henrique Ponde de Oliveira  
Pinto, Michael, Pokorny, Michelle Pokrass, Vitchyr H. Pong, Tolly Powell, Alethea Power, Boris



1080 Power, Elizabeth Proehl, Raul Puri, Alec Radford, Jack Rae, Aditya Ramesh, Cameron Raymond,  
1081 Francis Real, Kendra Rimbach, Carl Ross, Bob Rotsted, Henri Roussez, Nick Ryder, Mario  
1082 Saltarelli, Ted Sanders, Shibani Santurkar, Girish Sastry, Heather Schmidt, David Schnurr, John  
1083 Schulman, Daniel Selsam, Kyla Sheppard, Toki Sherbakov, Jessica Shieh, Sarah Shoker, Pranav  
1084 Shyam, Szymon Sidor, Eric Sigler, Maddie Simens, Jordan Sitkin, Katarina Slama, Ian Sohl,  
1085 Benjamin Sokolowsky, Yang Song, Natalie Staudacher, Felipe Petroski Such, Natalie Summers,  
1086 Ilya Sutskever, Jie Tang, Nikolas Tezak, Madeleine B. Thompson, Phil Tillet, Amin Tootoonchian,  
1087 Elizabeth Tseng, Preston Tuggle, Nick Turley, Jerry Tworek, Juan Felipe Cerón Uribe, Andrea  
1088 Vallone, Arun Vijayvergiya, Chelsea Voss, Carroll Wainwright, Justin Jay Wang, Alvin Wang, Ben  
1089 Wang, Jonathan Ward, Jason Wei, C. J. Weinmann, Akila Welihinda, Peter Welinder, Jiayi Weng,  
1090 Lilian Weng, Matt Wiethoff, Dave Willner, Clemens Winter, Samuel Wolrich, Hannah Wong,  
1091 Lauren Workman, Sherwin Wu, Jeff Wu, Michael Wu, Kai Xiao, Tao Xu, Sarah Yoo, Kevin Yu,  
1092 Qiming Yuan, Wojciech Zaremba, Rowan Zellers, Chong Zhang, Marvin Zhang, Shengjia Zhao,  
1093 Tianhao Zheng, Juntang Zhuang, William Zhuk, and Barret Zoph. GPT-4 Technical Report, March  
1094 2024a. URL <http://arxiv.org/abs/2303.08774>. arXiv:2303.08774 [cs].

1095 OpenAI, Aaron Hurst, Adam Lerer, Adam P. Goucher, Adam Perelman, Aditya Ramesh, Aidan  
1096 Clark, A. J. Ostrow, Akila Welihinda, Alan Hayes, Alec Radford, Aleksander Mądry, Alex  
1097 Baker-Whitcomb, Alex Beutel, Alex Borzunov, Alex Carney, Alex Chow, Alex Kirillov, Alex  
1098 Nichol, Alex Paino, Alex Renzin, Alex Tachard Passos, Alexander Kirillov, Alexi Christakis,  
1099 Alexis Conneau, Ali Kamali, Allan Jabri, Allison Moyer, Allison Tam, Amadou Crookes, Amin  
1100 Tootoonchian, Amin Tootoonchian, Ananya Kumar, Andrea Vallone, Andrej Karpathy, Andrew  
1101 Braunstein, Andrew Cann, Andrew Codispoti, Andrew Galu, Andrew Kondrich, Andrew Tulloch,  
1102 Andrey Mishchenko, Angela Baek, Angela Jiang, Antoine Pelisse, Antonia Woodford, Anuj  
1103 Gosalia, Arka Dhar, Ashley Pantuliano, Avi Nayak, Avital Oliver, Barret Zoph, Behrooz Ghorbani,  
1104 Ben Leimberger, Ben Rossen, Ben Sokolowsky, Ben Wang, Benjamin Zweig, Beth Hoover, Blake  
1105 Samic, Bob McGrew, Bobby Spero, Bogo Gertler, Bowen Cheng, Brad Lightcap, Brandon Walkin,  
1106 Brendan Quinn, Brian Guarraci, Brian Hsu, Bright Kellogg, Brydon Eastman, Camillo Lugaresi,  
1107 Carroll Wainwright, Cary Bassin, Cary Hudson, Casey Chu, Chad Nelson, Chak Li, Chan Jun  
1108 Shern, Channing Conger, Charlotte Barette, Chelsea Voss, Chen Ding, Cheng Lu, Chong Zhang,  
1109 Chris Beaumont, Chris Hallacy, Chris Koch, Christian Gibson, Christina Kim, Christine Choi,  
1110 Christine McLeavey, Christopher Hesse, Claudia Fischer, Clemens Winter, Coley Czarnecki,  
1111 Colin Jarvis, Colin Wei, Constantin Koumouzelis, Dane Sherburn, Daniel Kappler, Daniel Levin,  
1112 Daniel Levy, David Carr, David Farhi, David Mely, David Robinson, David Sasaki, Denny Jin,  
1113 Dev Valladares, Dimitris Tsipras, Doug Li, Duc Phong Nguyen, Duncan Findlay, Edele Oiwoh,  
1114 Edmund Wong, Ehsan Asdar, Elizabeth Proehl, Elizabeth Yang, Eric Antonow, Eric Kramer, Eric  
1115 Peterson, Eric Sigler, Eric Wallace, Eugene Brevdo, Evan Mays, Farzad Khorasani, Felipe Petroski  
1116 Such, Filippo Raso, Francis Zhang, Fred von Lohmann, Freddie Sulit, Gabriel Goh, Gene Oden,  
1117 Geoff Salmon, Giulio Starace, Greg Brockman, Hadi Salman, Haiming Bao, Haitang Hu, Hannah  
1118 Wong, Haoyu Wang, Heather Schmidt, Heather Whitney, Heewoo Jun, Hendrik Kirchner, Henrique  
1119 Ponde de Oliveira Pinto, Hongyu Ren, Huiwen Chang, Hyung Won Chung, Ian Kivlichan, Ian  
1120 O’Connell, Ian O’Connell, Ian Osband, Ian Silber, Ian Sohl, Ibrahim Okuyucu, Ikai Lan, Ilya  
1121 Kostrikov, Ilya Sutskever, Ingmar Kanitscheider, Ishaan Gulrajani, Jacob Coxon, Jacob Menick,  
1122 Jakub Pachocki, James Aung, James Betker, James Crooks, James Lennon, Jamie Kiros, Jan  
1123 Leike, Jane Park, Jason Kwon, Jason Phang, Jason Teplitz, Jason Wei, Jason Wolfe, Jay Chen,  
1124 Jeff Harris, Jenia Varavva, Jessica Gan Lee, Jessica Shieh, Ji Lin, Jiahui Yu, Jiayi Weng, Jie  
1125 Tang, Jieqi Yu, Joanne Jang, Joaquin Quinonero Candela, Joe Beutler, Joe Landers, Joel Parish,  
1126 Johannes Heidecke, John Schulman, Jonathan Lachman, Jonathan McKay, Jonathan Uesato,  
1127 Jonathan Ward, Jong Wook Kim, Joost Huizinga, Jordan Sitkin, Jos Kraaijeveld, Josh Gross, Josh  
1128 Kaplan, Josh Snyder, Joshua Achiam, Joy Jiao, Joyce Lee, Juntang Zhuang, Justyn Harriman, Kai  
1129 Fricke, Kai Hayashi, Karan Singhal, Katy Shi, Kavin Karthik, Kayla Wood, Kendra Rimbach,  
1130 Kenny Hsu, Kenny Nguyen, Keren Gu-Lemberg, Kevin Button, Kevin Liu, Kiel Howe, Krithika  
1131 Muthukumar, Kyle Luther, Lama Ahmad, Larry Kai, Lauren Itow, Lauren Workman, Leher Pathak,  
1132 Leo Chen, Li Jing, Lia Guy, Liam Fedus, Liang Zhou, Lien Mamitsuka, Lilian Weng, Lindsay  
1133 McCallum, Lindsey Held, Long Ouyang, Louis Feuvrier, Lu Zhang, Lukas Kondraciuk, Lukasz  
Kaiser, Luke Hewitt, Luke Metz, Lyric Doshi, Mada Aflak, Maddie Simens, Madelaine Boyd,  
Madelaine Thompson, Marat Dukhan, Mark Chen, Mark Gray, Mark Hudnall, Marvin Zhang,  
Marwan Aljubeih, Mateusz Litwin, Matthew Zeng, Max Johnson, Maya Shetty, Mayank Gupta,  
Meghan Shah, Mehmet Yatbaz, Meng Jia Yang, Mengchao Zhong, Mia Glaese, Mianna Chen,

- 1134 Michael Janner, Michael Lampe, Michael Petrov, Michael Wu, Michele Wang, Michelle Fradin,  
 1135 Michelle Pokrass, Miguel Castro, Miguel Oom Temudo de Castro, Mikhail Pavlov, Miles Brundage,  
 1136 Miles Wang, Minal Khan, Mira Murati, Mo Bavarian, Molly Lin, Murat Yesildal, Nacho Soto,  
 1137 Natalia Gimelshein, Natalie Cone, Natalie Staudacher, Natalie Summers, Natan LaFontaine, Neil  
 1138 Chowdhury, Nick Ryder, Nick Stathas, Nick Turley, Nik Tezak, Niko Felix, Nithanth Kudige,  
 1139 Nitish Keskar, Noah Deutsch, Noel Bundick, Nora Puckett, Ofir Nachum, Ola Okelola, Oleg Boiko,  
 1140 Oleg Murk, Oliver Jaffe, Olivia Watkins, Olivier Godement, Owen Campbell-Moore, Patrick Chao,  
 1141 Paul McMillan, Pavel Belov, Peng Su, Peter Bak, Peter Bakkum, Peter Deng, Peter Dolan, Peter  
 1142 Hoeschele, Peter Welinder, Phil Tillet, Philip Pronin, Philippe Tillet, Prafulla Dhariwal, Qiming  
 1143 Yuan, Rachel Dias, Rachel Lim, Rahul Arora, Rajan Troll, Randall Lin, Rapha Gontijo Lopes, Raul  
 1144 Puri, Reah Miyara, Reimar Leike, Renaud Gaubert, Reza Zamani, Ricky Wang, Rob Donnelly, Rob  
 1145 Honsby, Rocky Smith, Rohan Sahai, Rohit Ramchandani, Romain Huet, Rory Carmichael, Rowan  
 1146 Zellers, Roy Chen, Ruby Chen, Ruslan Nigmatullin, Ryan Cheu, Saachi Jain, Sam Altman, Sam  
 1147 Schoenholz, Sam Toizer, Samuel Miserendino, Sandhini Agarwal, Sara Culver, Scott Ethersmith,  
 1148 Scott Gray, Sean Grove, Sean Metzger, Shamez Hermani, Shantanu Jain, Shengjia Zhao, Sherwin  
 1149 Wu, Shino Jomoto, Shirong Wu, Shuaiqi, Xia, Sonia Phene, Spencer Papay, Srinivas Narayanan,  
 1150 Steve Coffey, Steve Lee, Stewart Hall, Suchir Balaji, Tal Broda, Tal Stramer, Tao Xu, Tarun  
 1151 Gogineni, Taya Christianson, Ted Sanders, Tejal Patwardhan, Thomas Cunninghamman, Thomas  
 1152 Degry, Thomas Dimson, Thomas Raoux, Thomas Shadwell, Tianhao Zheng, Todd Underwood,  
 1153 Todor Markov, Toki Sherbakov, Tom Rubin, Tom Stasi, Tomer Kaftan, Tristan Heywood, Troy  
 1154 Peterson, Tyce Walters, Tyna Eloundou, Valerie Qi, Veit Moeller, Vinnie Monaco, Vishal Kuo,  
 1155 Vlad Fomenko, Wayne Chang, Weiyi Zheng, Wenda Zhou, Wesam Manassra, Will Sheu, Wojciech  
 1156 Zaremba, Yash Patil, Yilei Qian, Yongjik Kim, Youlong Cheng, Yu Zhang, Yuchen He, Yuchen  
 1157 Zhang, Yujia Jin, Yunxing Dai, and Yury Malkov. GPT-4o System Card, October 2024b. URL  
<http://arxiv.org/abs/2410.21276>. arXiv:2410.21276 [cs].
- 1158 Sharad Rawat and M.-H. Herman Shen. A Novel Topology Optimization Approach using Con-  
 1159 ditional Deep Learning, January 2019. URL <http://arxiv.org/abs/1901.04859>.  
 1160 arXiv:1901.04859 [cs].
- 1161 Ronan Riochet, Mario Ynocente Castro, Mathieu Bernard, Adam Lerer, Rob Fergus, Véronique Izard,  
 1162 and Emmanuel Dupoux. IntPhys: A Framework and Benchmark for Visual Intuitive Physics Rea-  
 1163 soning, February 2020. URL <http://arxiv.org/abs/1803.07616>. arXiv:1803.07616  
 1164 [cs].
- 1165 Seungyeon Shin, Dongju Shin, and Namwoo Kang. Topology Optimization via Machine Learning  
 1166 and Deep Learning: A Review, June 2023. URL <http://arxiv.org/abs/2210.10782>.  
 1167 arXiv:2210.10782 [cs].
- 1168 Ivan Sosnovik and Ivan Oseledets. Neural networks for topology optimization, September 2017.  
 1169 URL <http://arxiv.org/abs/1709.09578>. arXiv:1709.09578 [cs].
- 1170 Gemini Team, Petko Georgiev, Ving Ian Lei, Ryan Burnell, Libin Bai, Anmol Gulati, Garrett Tanzer,  
 1171 Damien Vincent, Zhufeng Pan, Shibo Wang, Soroosh Mariooryad, Yifan Ding, Xinyang Geng, Fred  
 1172 Alcober, Roy Frostig, Mark Omernick, Lexi Walker, Cosmin Paduraru, Christina Sorokin, Andrea  
 1173 Tacchetti, Colin Gaffney, Samira Daruki, Olcan Sercinoglu, Zach Gleicher, Juliette Love, Paul  
 1174 Voigtlaender, Rohan Jain, Gabriela Surita, Kareem Mohamed, Rory Blevins, Junwhan Ahn, Tao  
 1175 Zhu, Kornraphop Kawintiranon, Orhan Firat, Yiming Gu, Yujing Zhang, Matthew Rahtz, Manaal  
 1176 Faruqui, Natalie Clay, Justin Gilmer, J. D. Co-Reyes, Ivo Penchev, Rui Zhu, Nobuyuki Morioka,  
 1177 Kevin Hui, Krishna Haridasan, Victor Campos, Mahdis Mahdieh, Mandy Guo, Samer Hassan,  
 1178 Kevin Kilgour, Arpi Vezer, Heng-Tze Cheng, Raoul de Liedekerke, Siddharth Goyal, Paul Barham,  
 1179 D. J. Strouse, Seb Noury, Jonas Adler, Mukund Sundararajan, Sharad Vikram, Dmitry Lepikhin,  
 1180 Michela Paganini, Xavier Garcia, Fan Yang, Dasha Valter, Maja Trebacz, Kiran Vodrahalli,  
 1181 Chulayuth Asawaroengchai, Roman Ring, Norbert Kalb, Livio Baldini Soares, Siddhartha Brahma,  
 1182 David Steiner, Tianhe Yu, Fabian Mentzer, Antoine He, Lucas Gonzalez, Bibo Xu, Raphael Lopez  
 1183 Kaufman, Laurent El Shafey, Junhyuk Oh, Tom Hennigan, George van den Driessche, Seth Odoom,  
 1184 Mario Lucic, Becca Roelofs, Sid Lall, Amit Marathe, Betty Chan, Santiago Ontanon, Luheng He,  
 1185 Denis Teplyashin, Jonathan Lai, Phil Crone, Bogdan Damoc, Lewis Ho, Sebastian Riedel, Karel  
 1186 Lenc, Chih-Kuan Yeh, Aakanksha Chowdhery, Yang Xu, Mehran Kazemi, Ehsan Amid, Anastasia  
 1187 Petrushkina, Kevin Swersky, Ali Khodaei, Gowoon Chen, Chris Larkin, Mario Pinto, Geng Yan,

1188 Adria Puigdomenech Badia, Piyush Patil, Steven Hansen, Dave Orr, Sebastien M. R. Arnold,  
 1189 Jordan Grimstad, Andrew Dai, Sholto Douglas, Rishika Sinha, Vikas Yadav, Xi Chen, Elena  
 1190 Gribovskaya, Jacob Austin, Jeffrey Zhao, Kaushal Patel, Paul Komarek, Sophia Austin, Sebastian  
 1191 Borgeaud, Linda Friso, Abhimanyu Goyal, Ben Caine, Kris Cao, Da-Woon Chung, Matthew  
 1192 Lamm, Gabe Barth-Maron, Thais Kagohara, Kate Olszewska, Mia Chen, Kaushik Shivakumar,  
 1193 Rishabh Agarwal, Harshal Godhia, Ravi Rajwar, Javier Snaider, Xerxes Dotiwalla, Yuan Liu,  
 1194 Aditya Barua, Victor Ungureanu, Yuan Zhang, Bat-Orgil Batsaikhan, Mateo Wirth, James Qin, Ivo  
 1195 Danihelka, Tulsee Doshi, Martin Chadwick, Jilin Chen, Sanil Jain, Quoc Le, Arjun Kar, Madhu  
 1196 Gurumurthy, Cheng Li, Ruoxin Sang, Fangyu Liu, Lampros Lamprou, Rich Munoz, Nathan Lintz,  
 1197 Harsh Mehta, Heidi Howard, Malcolm Reynolds, Lora Aroyo, Quan Wang, Lorenzo Blanco, Albin  
 1198 Cassirer, Jordan Griffith, Dipanjan Das, Stephan Lee, Jakub Sygnowski, Zach Fisher, James Besley,  
 1199 Richard Powell, Zafarali Ahmed, Dominik Paulus, David Reitter, Zalan Borsos, Rishabh Joshi,  
 1200 Aedan Pope, Steven Hand, Vittorio Selo, Vihan Jain, Nikhil Sethi, Megha Goel, Takaki Makino,  
 1201 Rhys May, Zhen Yang, Johan Schalkwyk, Christina Butterfield, Anja Hauth, Alex Goldin, Will  
 1202 Hawkins, Evan Senter, Sergey Brin, Oliver Woodman, Marvin Ritter, Eric Noland, Minh Giang,  
 1203 Vijay Bolina, Lisa Lee, Tim Blyth, Ian Mackinnon, Machel Reid, Obaid Sarvana, David Silver,  
 1204 Alexander Chen, Lily Wang, Loren Maggiore, Oscar Chang, Nithya Attaluri, Gregory Thornton,  
 1205 Chung-Cheng Chiu, Oskar Bunyan, Nir Levine, Timothy Chung, Evgenii Eltyshev, Xiance Si,  
 1206 Timothy Lillicrap, Demetra Brady, Vaibhav Aggarwal, Boxi Wu, Yuanzhong Xu, Ross McIlroy,  
 1207 Kartikeya Badola, Paramjit Sandhu, Erica Moreira, Wojciech Stokowiec, Ross Hemsley, Dong Li,  
 1208 Alex Tudor, Pranav Shyam, Elahe Rahimtoroghi, Salem Haykal, Pablo Sprechmann, Xiang Zhou,  
 1209 Diana Mincu, Yujia Li, Ravi Addanki, Kalpesh Krishna, Xiao Wu, Alexandre Frechette, Matan  
 1210 Eyal, Allan Dafoe, Dave Lacey, Jay Whang, Thi Avrahami, Ye Zhang, Emanuel Taropa, Hanzhao  
 1211 Lin, Daniel Toyama, Eliza Rutherford, Motoki Sano, HyunJeong Choe, Alex Tomala, Chalence  
 1212 Safranek-Shrader, Nora Kassner, Mantas Pajarskas, Matt Harvey, Sean Sechrist, Meire Fortunato,  
 1213 Christina Lyu, Gamaleldin Elsayed, Chenkai Kuang, James Lottes, Eric Chu, Chao Jia, Chih-Wei  
 1214 Chen, Peter Humphreys, Kate Baumli, Connie Tao, Rajkumar Samuel, Cicero Nogueira dos  
 1215 Santos, Anders Andreassen, Nemanja Rakićević, Dominik Grewe, Aviral Kumar, Stephanie  
 1216 Winkler, Jonathan Caton, Andrew Brock, Sid Dalmia, Hannah Sheahan, Iain Barr, Yingjie Miao,  
 1217 Paul Natsev, Jacob Devlin, Feryal Behbahani, Flavien Prost, Yanhua Sun, Artiom Myaskovsky,  
 1218 Thanumalayan Sankaranarayanan Pillai, Dan Hurt, Angeliki Lazaridou, Xi Xiong, Ce Zheng, Fabio  
 1219 Pardo, Xiaowei Li, Dan Horgan, Joe Stanton, Moran Ambar, Fei Xia, Alejandro Lince, Mingqiu  
 1220 Wang, Basil Mustafa, Albert Webson, Hyo Lee, Rohan Anil, Martin Wicke, Timothy Dozat,  
 1221 Abhishek Sinha, Enrique Piqueras, Elahe Dabir, Shyam Upadhyay, Anudhyan Boral, Lisa Anne  
 1222 Hendricks, Corey Fry, Josip Djolonga, Yi Su, Jake Walker, Jane Labanowski, Ronny Huang, Vedant  
 1223 Misra, Jeremy Chen, R. J. Skerry-Ryan, Avi Singh, Shruti Rijhwani, Dian Yu, Alex Castro-Ros,  
 1224 Beer Changpinyo, Romina Datta, Sumit Bagri, Arnar Mar Hrafnkelsson, Marcello Maggioni,  
 1225 Daniel Zheng, Yury Sulsky, Shaobo Hou, Tom Le Paine, Antoine Yang, Jason Riesa, Dominika  
 1226 Rogozinska, Dror Marcus, Dalia El Badawy, Qiao Zhang, Luyu Wang, Helen Miller, Jeremy  
 1227 Greer, Lars Lowe Sjos, Azade Nova, Heiga Zen, Rahma Chaabouni, Mihaela Rosca, Jiepu Jiang,  
 1228 Charlie Chen, Ruibo Liu, Tara Sainath, Maxim Krikun, Alex Polozov, Jean-Baptiste Lespiau,  
 1229 Josh Newlan, Zeynep Cankara, Soo Kwak, Yunhan Xu, Phil Chen, Andy Coenen, Clemens  
 1230 Meyer, Katerina Tsihlias, Ada Ma, Juraj Gottweis, Jinwei Xing, Chenjie Gu, Jin Miao, Christian  
 1231 Frank, Zeynep Cankara, Sanjay Ganapathy, Ishita Dasgupta, Steph Hughes-Fitt, Heng Chen,  
 1232 David Reid, Keran Rong, Hongmin Fan, Joost van Amersfoort, Vincent Zhuang, Aaron Cohen,  
 1233 Shixiang Shane Gu, Anhad Mohananey, Anastasija Ilic, Taylor Tobin, John Wieting, Anna Bortsova,  
 1234 Phoebe Thacker, Emma Wang, Emily Caveness, Justin Chiu, Eren Sezener, Alex Kaskasoli,  
 1235 Steven Baker, Katie Millican, Mohamed Elhawaty, Kostas Aisopos, Carl Lebsack, Nathan Byrd,  
 1236 Hanjun Dai, Wenhao Jia, Matthew Wiethoff, Elnaz Davoodi, Albert Weston, Lakshman Yagati,  
 1237 Arun Ahuja, Isabel Gao, Golan Pundak, Susan Zhang, Michael Azzam, Khe Chai Sim, Sergi  
 1238 Caelles, James Keeling, Abhanshu Sharma, Andy Swing, YaGuang Li, Chenxi Liu, Carrie Grimes  
 1239 Bostock, Yamini Bansal, Zachary Nado, Ankesh Anand, Josh Lipschultz, Abhijit Karmarkar,  
 1240 Lev Proleev, Abe Ittycheriah, Soheil Hassas Yeganeh, George Polovets, Aleksandra Faust, Jiao  
 1241 Sun, Alban Rustemi, Pen Li, Rakesh Shivanna, Jeremiah Liu, Chris Welty, Federico Lebron,  
 Anirudh Baddepudi, Sebastian Krause, Emilio Parisotto, Radu Soricut, Zheng Xu, Dawn Bloxwich,  
 Melvin Johnson, Behnam Neyshabur, Justin Mao-Jones, Renshen Wang, Vinay Ramasesh, Zaheer  
 Abbas, Arthur Guez, Constant Segal, Duc Dung Nguyen, James Svensson, Le Hou, Sarah York,  
 Kieran Milan, Sophie Bridgers, Wiktoria Gworek, Marco Tagliasacchi, James Lee-Thorp, Michael  
 Chang, Alexey Guseynov, Ale Jakse Hartman, Michael Kwong, Ruizhe Zhao, Sheleem Kashem,

1242 Elizabeth Cole, Antoine Miech, Richard Tanburn, Mary Phuong, Filip Pavetic, Sebastien Cevey,  
 1243 Ramona Comanescu, Richard Ives, Sherry Yang, Cosmo Du, Bo Li, Zizhao Zhang, Mariko Iinuma,  
 1244 Clara Huiyi Hu, Aurko Roy, Shaan Bijwadia, Zhenkai Zhu, Danilo Martins, Rachel Saputro, Anita  
 1245 Gergely, Steven Zheng, Dawei Jia, Ioannis Antonoglou, Adam Sadovsky, Shane Gu, Yingying  
 1246 Bi, Alek Andreev, Sina Samangooei, Mina Khan, Tomas Kocisky, Angelos Filos, Chintu Kumar,  
 1247 Colton Bishop, Adams Yu, Sarah Hodkinson, Sid Mittal, Premal Shah, Alexandre Moufarek, Yong  
 1248 Cheng, Adam Bloniarz, Jaehoon Lee, Pedram Pejman, Paul Michel, Stephen Spencer, Vladimir  
 1249 Feinberg, Xuehan Xiong, Nikolay Savinov, Charlotte Smith, Siamak Shakeri, Dustin Tran, Mary  
 1250 Chesus, Bernd Bohnet, George Tucker, Tamara von Glehn, Carrie Muir, Yiran Mao, Hideto Kazawa,  
 1251 Ambrose Slone, Kedar Soparkar, Disha Shrivastava, James Cobon-Kerr, Michael Sharman, Jay  
 1252 Pavagadhi, Carlos Araya, Karolis Misiunas, Nimesh Ghelani, Michael Laskin, David Barker,  
 1253 Qiujia Li, Anton Briukhov, Neil Houlsby, Mia Glaese, Balaji Lakshminarayanan, Nathan Schucher,  
 1254 Yunhao Tang, Eli Collins, Hyeontaek Lim, Fangxiaoyu Feng, Adria Recasens, Guangda Lai,  
 1255 Alberto Magni, Nicola De Cao, Aditya Siddhant, Zoe Ashwood, Jordi Orbay, Mostafa Dehghani,  
 1256 Jenny Brennan, Yifan He, Kelvin Xu, Yang Gao, Carl Saroufim, James Molloy, Xinyi Wu, Seb  
 1257 Arnold, Solomon Chang, Julian Schrittwieser, Elena Buchatskaya, Soroush Radpour, Martin  
 1258 Polacek, Skye Giordano, Ankur Bapna, Simon Tokumine, Vincent Hellendoorn, Thibault Sottiaux,  
 1259 Sarah Cogan, Aliaksei Severyn, Mohammad Saleh, Shantanu Thakoor, Laurent Shefey, Siyan  
 1260 Qiao, Meenu Gaba, Shuo-yiin Chang, Craig Swanson, Biao Zhang, Benjamin Lee, Paul Kishan  
 1261 Rubenstein, Gan Song, Tom Kwiatkowski, Anna Koop, Ajay Kannan, David Kao, Parker Schuh,  
 1262 Axel Stjerngren, Golnaz Ghiasi, Gena Gibson, Luke Vilnis, Ye Yuan, Felipe Tiengo Ferreira,  
 1263 Aishwarya Kamath, Ted Klimenko, Ken Franko, Kefan Xiao, Indro Bhattacharya, Miteyan Patel,  
 1264 Rui Wang, Alex Morris, Robin Strudel, Vivek Sharma, Peter Choy, Sayed Hadi Hashemi, Jessica  
 1265 Landon, Mara Finkelstein, Priya Jhakra, Justin Frye, Megan Barnes, Matthew Mauger, Dennis  
 1266 Daun, Khuslen Baatarsukh, Matthew Tung, Wael Farhan, Henryk Michalewski, Fabio Viola, Felix  
 1267 de Chaumont Quitry, Charline Le Lan, Tom Hudson, Qingze Wang, Felix Fischer, Ivy Zheng,  
 1268 Elspeth White, Anca Dragan, Jean-baptiste Alayrac, Eric Ni, Alexander Pritzel, Adam Iwanicki,  
 1269 Michael Isard, Anna Bulanova, Lukas Zilka, Ethan Dyer, Devendra Sachan, Srivatsan Srinivasan,  
 1270 Hannah Muckenhirn, Honglong Cai, Amol Mandhane, Mukarram Tariq, Jack W. Rae, Gary Wang,  
 1271 Kareem Ayoub, Nicholas FitzGerald, Yao Zhao, Woohyun Han, Chris Alberti, Dan Garrette,  
 1272 Kashyap Krishnakumar, Mai Gimenez, Anselm Levskaya, Daniel Sohn, Josip Matak, Inaki Iturrate,  
 1273 Michael B. Chang, Jackie Xiang, Yuan Cao, Nishant Ranka, Geoff Brown, Adrian Hutter, Vahab  
 1274 Mirrokni, Nanxin Chen, Kaisheng Yao, Zoltan Egyed, Francois Galilee, Tyler Liechty, Praveen  
 1275 Kallakuri, Evan Palmer, Sanjay Ghemawat, Jasmine Liu, David Tao, Chloe Thornton, Tim Green,  
 1276 Mimi Jasarevic, Sharon Lin, Victor Cotruta, Yi-Xuan Tan, Noah Fiedel, Hongkun Yu, Ed Chi,  
 1277 Alexander Neitz, Jens Heitkaemper, Anu Sinha, Denny Zhou, Yi Sun, Charbel Kaed, Brice Hulse,  
 1278 Swaroop Mishra, Maria Georgaki, Sneha Kudugunta, Clement Farabet, Izhak Shafran, Daniel  
 1279 Vlasic, Anton Tsitsulin, Rajagopal Ananthanarayanan, Alen Carin, Guolong Su, Pei Sun, Shashank  
 1280 V, Gabriel Carvajal, Josef Broder, Iulia Comsa, Alena Repina, William Wong, Warren Weilun Chen,  
 1281 Peter Hawkins, Egor Filonov, Lucia Loher, Christoph Hirsenschall, Weiyi Wang, Jingchen Ye, Andrea  
 1282 Burns, Hardie Cate, Diana Gage Wright, Federico Piccinini, Lei Zhang, Chu-Cheng Lin, Ionel  
 1283 Gog, Yana Kulizhskaya, Ashwin Sreevatsa, Shuang Song, Luis C. Cobo, Anand Iyer, Chetan Tekur,  
 1284 Guillermo Garrido, Zhu Yun Xiao, Rupert Kemp, Huaixiu Steven Zheng, Hui Li, Ananth Agarwal,  
 1285 Christel Ngani, Kati Goshvadi, Rebeca Santamaria-Fernandez, Wojciech Fica, Xinyun Chen,  
 1286 Chris Gorgolewski, Sean Sun, Roopal Garg, Xinyu Ye, S. M. Ali Eslami, Nan Hua, Jon Simon,  
 1287 Pratik Joshi, Yelin Kim, Ian Tenney, Sahitya Potluri, Lam Nguyen Thiet, Quan Yuan, Florian  
 1288 Luisier, Alexandra Chronopoulou, Salvatore Scellato, Praveen Srinivasan, Minmin Chen, Vinod  
 1289 Koverkathu, Valentin Dalibard, Yaming Xu, Brennan Saeta, Keith Anderson, Thibault Sellam,  
 1290 Nick Fernando, Fantine Huot, Junehyuk Jung, Mani Varadarajan, Michael Quinn, Amit Raul,  
 1291 Maigo Le, Ruslan Habalov, Jon Clark, Komal Jalan, Kalesha Bullard, Achintya Singhal, Thang  
 1292 Luong, Boyu Wang, Sujeewan Rajayogam, Julian Eisenschlos, Johnson Jia, Daniel Finchelstein,  
 1293 Alex Yakubovich, Daniel Balle, Michael Fink, Sameer Agarwal, Jing Li, Dj Dvijotham, Shalini  
 1294 Pal, Kai Kang, Jaclyn Konzelmann, Jennifer Beattie, Olivier Dousse, Diane Wu, Remi Crocker,  
 1295 Chen Elkind, Siddhartha Reddy Jonnalagadda, Jong Lee, Dan Holtmann-Rice, Krystal Kallarackal,  
 Rosanne Liu, Denis Vnukov, Neera Vats, Luca Invernizzi, Mohsen Jafari, Huanjie Zhou, Lilly  
 Taylor, Jennifer Prendki, Marcus Wu, Tom Eccles, Tianqi Liu, Kavya Koppurapu, Françoise  
 Beaufays, Christof Angermueller, Andreea Marzoca, Shourya Sarcar, Hilal Dib, Jeff Stanway,  
 Frank Perbet, Nejc Trdin, Rachel Sterneck, Andrey Khorlin, Dinghua Li, Xihui Wu, Sonam Goenka,  
 David Madras, Sasha Goldshtein, Willi Gierke, Tong Zhou, Yaxin Liu, Yannie Liang, Anais White,

1296 Yunjie Li, Shreya Singh, Sanaz Bahargam, Mark Epstein, Sujoy Basu, Li Lao, Adnan Ozturel, Carl  
 1297 Crous, Alex Zhai, Han Lu, Zora Tung, Neeraj Gaur, Alanna Walton, Lucas Dixon, Ming Zhang,  
 1298 Amir Globerson, Grant Uy, Andrew Bolt, Olivia Wiles, Milad Nasr, Ilia Shumailov, Marco Selvi,  
 1299 Francesco Piccinno, Ricardo Aguilar, Sara McCarthy, Misha Khalman, Mrinal Shukla, Vlado Galic,  
 1300 John Carpenter, Kevin Villela, Haibin Zhang, Harry Richardson, James Martens, Matko Bosnjak,  
 1301 Shreyas Rammohan Belle, Jeff Seibert, Mahmoud Alnahlawi, Brian McWilliams, Sankalp Singh,  
 1302 Annie Louis, Wen Ding, Dan Popovici, Lenin Simicich, Laura Knight, Pulkit Mehta, Nishesh  
 1303 Gupta, Chongyang Shi, Saaber Fatehi, Jovana Mitrovic, Alex Grills, Joseph Pagadora, Tsendsuren  
 1304 Munkhdalai, Dessie Petrova, Danielle Eisenbud, Zhishuai Zhang, Damion Yates, Bhavishya  
 1305 Mittal, Nilesh Tripuraneni, Yannis Assael, Thomas Brovelli, Prateek Jain, Mihajlo Velimirovic,  
 1306 Canfer Akbulut, Jiaqi Mu, Wolfgang Macherey, Ravin Kumar, Jun Xu, Haroon Qureshi, Gheorghe  
 1307 Comanici, Jeremy Wiesner, Zhitao Gong, Anton Rudderick, Matthias Bauer, Nick Felt, Anirudh  
 1308 GP, Anurag Arnab, Dustin Zelle, Jonas Rothfuss, Bill Rosgen, Ashish Shenoy, Bryan Seybold,  
 1309 Xinjian Li, Jayaram Mudigonda, Goker Erdogan, Jiawei Xia, Jiri Simsa, Andrea Michi, Yi Yao,  
 1310 Christopher Yew, Steven Kan, Isaac Caswell, Carey Radebaugh, Andre Elisseeff, Pedro Valenzuela,  
 1311 Kay McKinney, Kim Paterson, Albert Cui, Eri Latorre-Chimoto, Solomon Kim, William Zeng, Ken  
 1312 Durden, Priya Ponnappalli, Tiberiu Sosea, Christopher A. Choquette-Choo, James Manyika, Brona  
 1313 Robenek, Harsha Vashisht, Sebastien Pereira, Hoi Lam, Marko Velic, Denese Owusu-Afriyie,  
 1314 Katherine Lee, Tolga Bolukbasi, Alicia Parrish, Shawn Lu, Jane Park, Balaji Venkatraman, Alice  
 1315 Talbert, Lambert Rosique, Yuchung Cheng, Andrei Sozanschi, Adam Paszke, Praveen Kumar,  
 1316 Jessica Austin, Lu Li, Khalid Salama, Bartek Perz, Wooyeol Kim, Nandita Dukkupati, Anthony  
 1317 Baryshnikov, Christos Kaplanis, XiangHai Sheng, Yuri Chervonyi, Caglar Unlu, Diego de Las  
 1318 Casas, Harry Askham, Kathryn Tunyasuvunakool, Felix Gimeno, Siim Poder, Chester Kwak,  
 1319 Matt Miecniowski, Vahab Mirrokni, Alek Dimitriev, Aaron Parisi, Dangyi Liu, Tomy Tsai, Toby  
 1320 Shevlane, Christina Kouridi, Drew Garmon, Adrian Goedeckemeyer, Adam R. Brown, Anitha  
 1321 Vijayakumar, Ali Elqursh, Sadegh Jazayeri, Jin Huang, Sara Mc Carthy, Jay Hoover, Lucy Kim,  
 1322 Sandeep Kumar, Wei Chen, Courtney Biles, Garrett Bingham, Evan Rosen, Lisa Wang, Qijun Tan,  
 1323 David Engel, Francesco Pongetti, Dario de Cesare, Dongseong Hwang, Lily Yu, Jennifer Pullman,  
 1324 Srini Narayanan, Kyle Levin, Siddharth Gopal, Megan Li, Asaf Aharoni, Trieu Trinh, Jessica Lo,  
 1325 Norman Casagrande, Roopali Vij, Loic Matthey, Bramandia Ramadhana, Austin Matthews, C. J.  
 1326 Carey, Matthew Johnson, Kremena Goranova, Rohin Shah, Shereen Ashraf, Kingshuk Dasgupta,  
 1327 Rasmus Larsen, Yicheng Wang, Manish Reddy Vuyyuru, Chong Jiang, Joana Ijazi, Kazuki Osawa,  
 1328 Celine Smith, Ramya Sree Boppana, Taylan Bilal, Yuma Koizumi, Ying Xu, Yasemin Altun, Nir  
 1329 Shabat, Ben Bariach, Alex Korchemny, Kiam Choo, Olaf Ronneberger, Chimezie Iwuanyanwu,  
 1330 Shubin Zhao, David Soergel, Cho-Jui Hsieh, Irene Cai, Shariq Iqbal, Martin Sundermeyer, Zhe  
 1331 Chen, Elie Bursztein, Chaitanya Malaviya, Fadi Biadsy, Prakash Shroff, Inderjit Dhillon, Tejasi  
 1332 Latkar, Chris Dyer, Hannah Forbes, Massimo Nicosia, Vitaly Nikolaev, Somer Greene, Marin  
 1333 Georgiev, Pidong Wang, Nina Martin, Hanie Sedghi, John Zhang, Praseem Banzal, Doug Fritz,  
 1334 Vikram Rao, Xuezhi Wang, Jiageng Zhang, Viorica Patraucean, Dayou Du, Igor Mordatch, Ivan  
 1335 Jurin, Lewis Liu, Ayush Dubey, Abhi Mohan, Janek Nowakowski, Vlad-Doru Ion, Nan Wei, Reiko  
 1336 Tojo, Maria Abi Raad, Drew A. Hudson, Vaishakh Keshava, Shubham Agrawal, Kevin Ramirez,  
 1337 Zhichun Wu, Hoang Nguyen, Ji Liu, Madhavi Sewak, Bryce Petriani, DongHyun Choi, Ivan Philips,  
 1338 Ziyue Wang, Ioana Bica, Ankush Garg, Jarek Wilkiewicz, Priyanka Agrawal, Xiaowei Li, Danhao  
 1339 Guo, Emily Xue, Naseer Shaik, Andrew Leach, Sadh MNM Khan, Julia Wiesinger, Sammy Jerome,  
 1340 Abhishek Chakladar, Alek Wenjiao Wang, Tina Ornduff, Folake Abu, Alireza Ghaffarkhah, Marcus  
 1341 Wainwright, Mario Cortes, Frederick Liu, Joshua Maynez, Andreas Terzis, Pouya Samangouei,  
 1342 Riham Mansour, Tomasz Kępa, François-Xavier Aubet, Anton Algymer, Dan Banica, Agoston  
 1343 Weisz, Andras Orban, Alexandre Senges, Ewa Andrejczuk, Mark Geller, Niccolo Dal Santo,  
 1344 Valentin Anklin, Majd Al Merey, Martin Baeuml, Trevor Strohmman, Junwen Bai, Slav Petrov,  
 1345 Yonghui Wu, Demis Hassabis, Koray Kavukcuoglu, Jeff Dean, and Oriol Vinyals. Gemini 1.5:  
 1346 Unlocking multimodal understanding across millions of tokens of context, December 2024. URL  
 1347 <http://arxiv.org/abs/2403.05530>. arXiv:2403.05530 [cs].

1345 Perplexity Team. Meet new Sonar: A Blazing Fast Model Optimized for Perplexity Search, February  
 1346 2025a. URL <https://www.perplexity.ai/hub>.

1348 Perplexity Team. RL Training For Math Reasoning, May 2025b. URL <https://www.perplexity.ai/hub>.

1350 Kexin Yi, Chuang Gan, Yunzhu Li, Pushmeet Kohli, Jiajun Wu, Antonio Torralba, and Joshua B.  
1351 Tenenbaum. CLEVRER: CoLLision Events for Video REpresentation and Reasoning, March 2020.  
1352 URL <http://arxiv.org/abs/1910.01442>. arXiv:1910.01442 [cs].  
1353  
1354 Xinyu Zhang, Yuxuan Dong, Yanrui Wu, Jiaying Huang, Chengyou Jia, Basura Fernando, Mike Zheng  
1355 Shou, Lingling Zhang, and Jun Liu. PhysReason: A Comprehensive Benchmark towards  
1356 Physics-Based Reasoning, February 2025. URL <http://arxiv.org/abs/2502.12054>.  
1357 arXiv:2502.12054 [cs].  
1358 Yiquan Zhang, Bo Peng, Xiaoyi Zhou, Cheng Xiang, and Dalei Wang. A deep Convolutional  
1359 Neural Network for topology optimization with strong generalization ability, March 2020. URL  
1360 <http://arxiv.org/abs/1901.07761>. arXiv:1901.07761 [cs].  
1361  
1362  
1363  
1364  
1365  
1366  
1367  
1368  
1369  
1370  
1371  
1372  
1373  
1374  
1375  
1376  
1377  
1378  
1379  
1380  
1381  
1382  
1383  
1384  
1385  
1386  
1387  
1388  
1389  
1390  
1391  
1392  
1393  
1394  
1395  
1396  
1397  
1398  
1399  
1400  
1401  
1402  
1403

A APPENDIX

B TOPOLOGY OPTIMIZATION SAMPLES: 2D

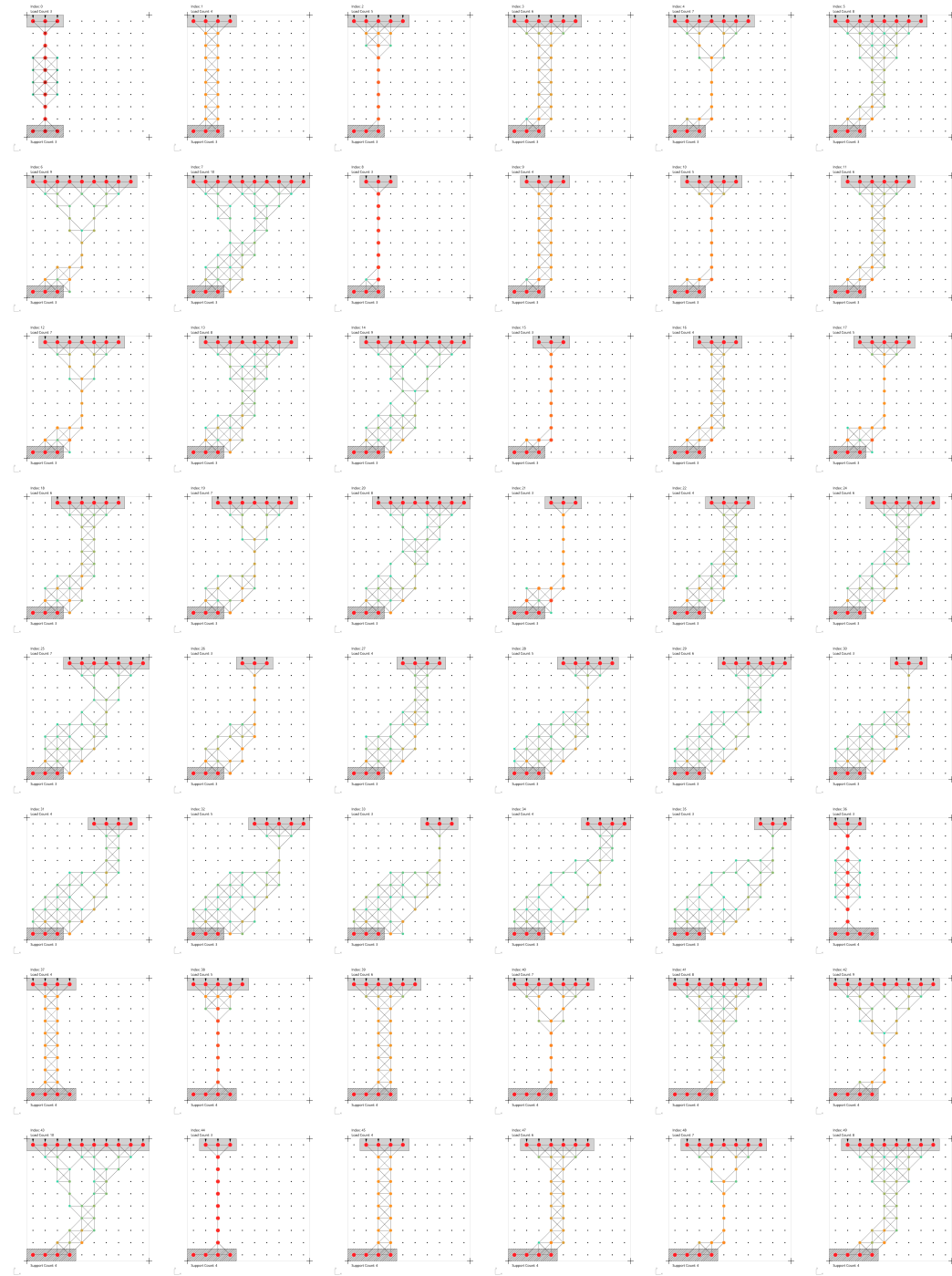


Figure 7: Example 2D topology optimization samples from the SPhyR dataset.

### C TOPOLOGY OPTIMIZATION SAMPLES: 3D

1458  
1459  
1460  
1461  
1462  
1463  
1464  
1465  
1466  
1467  
1468  
1469  
1470  
1471  
1472  
1473  
1474  
1475  
1476  
1477  
1478  
1479  
1480  
1481  
1482  
1483  
1484  
1485  
1486  
1487  
1488  
1489  
1490  
1491  
1492  
1493  
1494  
1495  
1496  
1497  
1498  
1499  
1500  
1501  
1502  
1503  
1504  
1505  
1506  
1507  
1508  
1509  
1510  
1511

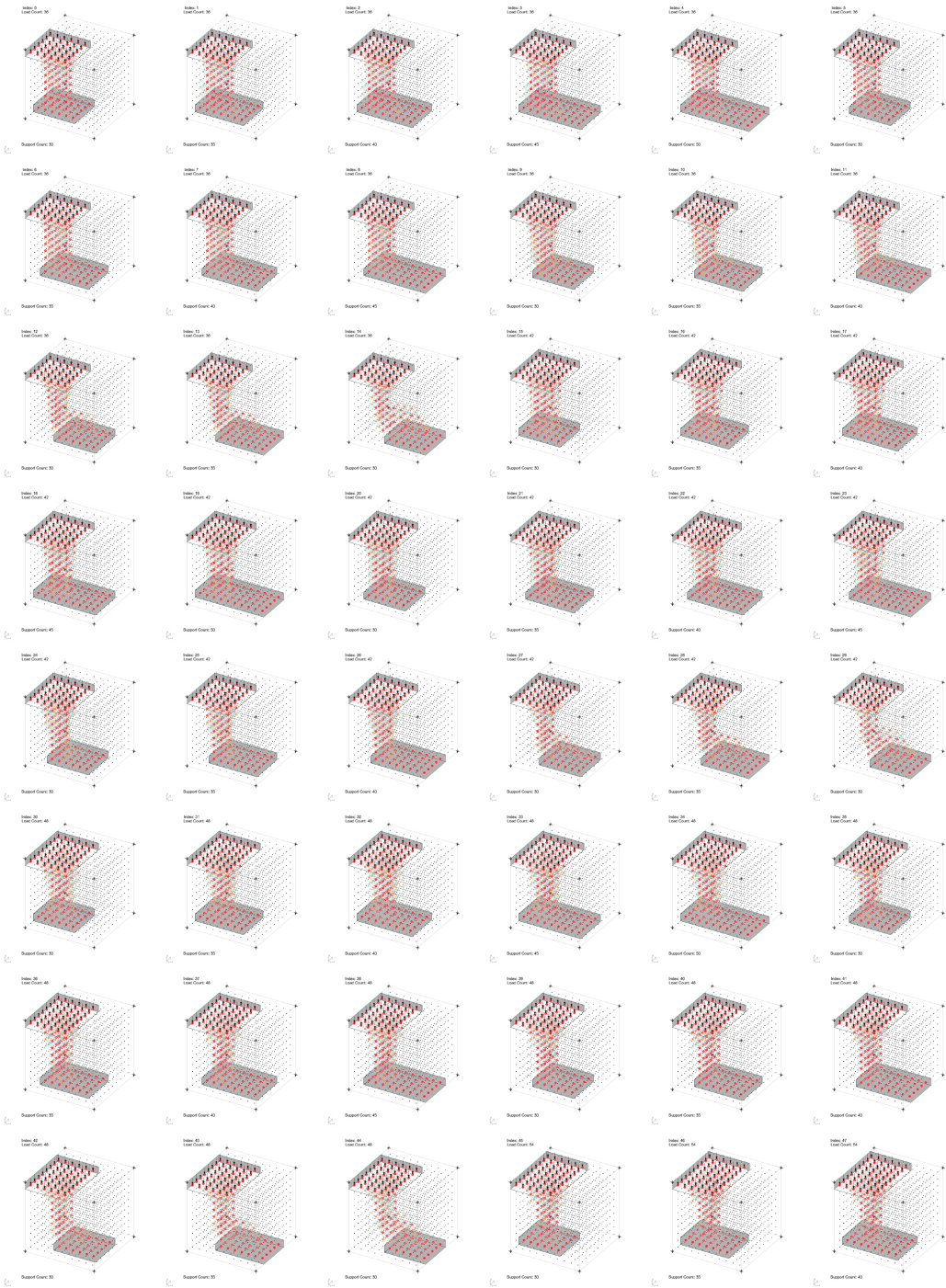


Figure 8: Example 3D topology optimization samples included for future benchmark extensions.



1512 **D ADDITIONAL ROTATION EXPERIMENT RESULTS**

1513

1514

1515

1516

1517

**D.1 ROTATED (3) SAMPLE PROMPT & COMPLETION: 1 RANDOM CELL EASY**

1518 **Prompt:** You are given a structural material distribution represented as a grid. Each cell can have one of  
1519 the following states:

1520 - 'L' indicates applied load.

1521 - 'V' indicates void.

1522 - 'S' indicates support.

1523

1524

1525

1526

The goal is to predict the correct material distribution by filling in all 'V' cells with either '1' (solid) or '0' (empty), based on the surrounding structure and implicit physical reasoning (such as load paths, supports, and forces).

1527

1528

1529

Important: The completed structure should use as little material as possible while remaining stable and plausible for carrying the applied forces. Minimize material usage unless necessary for structural support.

1530

Below is the input grid with masked regions:

1531

1532

1533

1534

1535

1536

1537

1538

1539

1540

1541

```
0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 S
0 0 0 0 0 0 0 0 0 S
0 0 0 0 0 0 0 0 0 S
0 0 0 0 0 0 0 0 0 S
0 0 0 0 0 0 0 0 0 S
L 0 0 0 0 0 1 1 1 S
L 1 V 1 1 1 1 1 1 S
L 0 0 0 0 0 1 1 1 S
0 0 0 0 0 0 0 0 0 S
0 0 0 0 0 0 0 0 0 S
```

1542

1543

1544

Please output the completed grid by replacing all 'V' cells with either '1' (solid) or '0' (empty).

Maintain the same format as the input: one row per line, cells separated by spaces, and the total number of rows and columns unchanged.

1545

Return only the completed grid without any additional explanation.

1546

**Model Completion (differences in red):**

1547

1548

1549

1550

1551

1552

1553

1554

1555

1556

1557

1558

1559

1560

1561

1562

1563

1564

1565

```
0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 S
0 0 0 0 0 0 0 0 0 S
0 0 0 0 0 0 0 0 0 S
0 0 0 0 0 0 0 0 0 S
0 0 0 0 0 0 0 0 0 S
L 0 0 0 0 0 1 1 1 S
L 1 1 1 1 1 1 1 1 S
L 0 0 0 0 0 1 1 1 S
0 0 0 0 0 0 0 0 0 S
0 0 0 0 0 0 0 0 0 S
0 0 0 0 0 0 0 0 0 S
0 0 0 0 0 0 0 0 0 S
L 0 0 0 0 0 1 1 1 S
L 1 1 1 1 1 1 1 1 S
L 0 0 0 0 0 1 1 1 S
0 0 0 0 0 0 0 0 0 S
0 0 0 0 0 0 0 0 0 S
```

1556

**Ground Truth:**

1557

1558

1559

1560

1561

1562

1563

1564

1565

D.2 ROTATED (3) RESULTS FOR MODEL SUB-SET AND TASK SUB-SET

Table 3: Exact Match, Overlap- and Normalized Overlap-Scores on sub-set of easy difficulty tasks.

Task	Metric	GPT 4.1	GPT 4.1 (3 Rotations)	Claude Opus 4	Claude Opus 4 (3 Rotations)	Gemini 2.5 Pro	Gemini 2.5 Pro (3 Rotations)	DeepSeek-R1	DeepSeek-R1 (3 Rotations)	Perplexity Sonar	Perplexity Sonar (3 Rotations)
Difficulty: Easy											
10 Random Cells	Exact Match	0	0	15	22	8	10	2	5	1	2
	Score	27.36	27.77	78.72	83.51	12.23	74.10	58.66	62.48	20.72	53.10
3 Random Rows	Normalized Score	27.36	27.89	78.72	83.51	44.21	74.40	58.86	62.48	49.52	53.60
	Exact Match	6	1	35	17	29	3	20	9	24	0
3 Random Columns	Score	74.88	49.35	92.45	88.13	89.73	80.18	73.80	64.13	83.13	57.53
	Normalized Score	74.88	49.35	92.45	88.13	89.73	80.18	73.80	64.13	83.13	57.57
Full	Exact Match	1	3	5	2	3	2	2	1	1	1
	Score	41.57	70.85	74.05	71.63	69.08	57.33	51.13	49.20	53.53	60.07
Average	Normalized Score	41.57	70.85	74.05	71.63	69.73	57.63	51.13	49.43	54.82	60.10
	Exact Match	0	0	0	0	0	0	0	0	0	0
Average	Score	64.69	68.04	70.09	69.79	71.59	67.38	48.49	54.38	57.83	67.55
	Normalized Score	64.69	68.04	70.09	69.79	71.59	67.38	48.49	54.38	57.83	67.55
Average	Exact Match	1.75	1.00	13.75	10.25	10.00	3.75	6.00	3.50	6.50	0.75
	Score	52.12	54.00	78.83	78.27	60.66	69.75	58.02	57.55	53.80	59.56
Average	Normalized Score	52.12	54.03	78.83	78.27	68.82	69.90	58.07	57.60	61.32	59.70

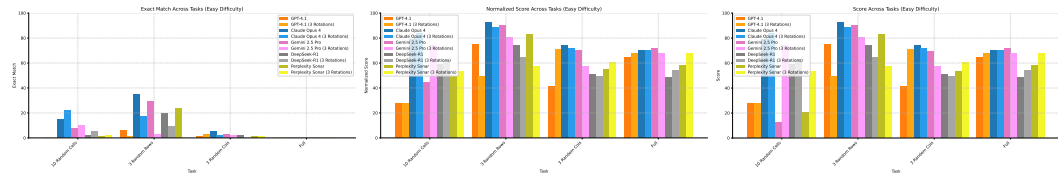


Figure 9: Exact Match, Overlap- and Normalized Overlap-Scores on sub-set of easy difficulty tasks graph plot.

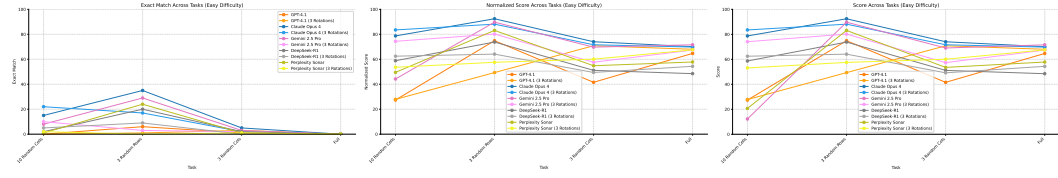


Figure 10: Exact Match, Overlap- and Normalized Overlap-Scores on sub-set of easy difficulty tasks line plot.

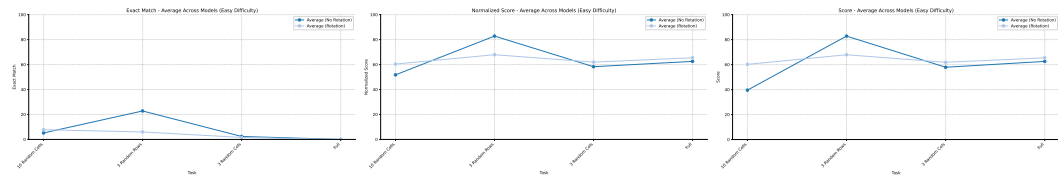


Figure 11: Average Exact Match, Overlap- and Normalized Overlap-Scores on sub-set of easy difficulty tasks line plot.

## D.3 ROTATED (3) RESULTS FOR BEST PERFORMING MODEL: CLAUDE OPUS 4 ON ALL TASKS

Table 4: Exact Match, Overlap- and Normalized Overlap-Scores across all difficulties and tasks for Claude Opus 4 comparing rotation and non-rotation setup.

Task	Metric	Claude Opus 4	Claude Opus 4 (3 Rotations)
Difficulty: Easy			
1 Random Cell	Exact Match	82	82
	Score	82.00	82.00
	Normalized Score	82.00	82.00
5 Random Cells	Exact Match	45	44
	Score	84.27	84.25
	Normalized Score	84.27	84.25
10 Random Cells	Exact Match	15	22
	Score	78.72	83.51
	Normalized Score	78.72	83.51
1 Random Row	Exact Match	52	42
	Score	90.50	86.70
	Normalized Score	90.50	86.70
3 Random Rows	Exact Match	35	17
	Score	92.45	88.13
	Normalized Score	92.45	88.13
1 Random Column	Exact Match	23	27
	Score	68.50	73.00
	Normalized Score	68.50	73.00
3 Random Columns	Exact Match	5	2
	Score	74.05	71.63
	Normalized Score	74.05	71.63
Full	Exact Match	0	0
	Score	70.09	69.79
	Normalized Score	70.09	69.79
Average	Exact Match	32.12	29.50
	Score	80.07	79.88
	Normalized Score	80.07	79.88
Difficulty: Hard			
1 Random Cell	Exact Match	76	80
	Score	73.00	80.00
	Normalized Score	76.00	80.00
5 Random Cells	Exact Match	38	41
	Score	81.20	81.55
	Normalized Score	81.20	81.55
10 Random Cells	Exact Match	13	18
	Score	77.61	80.08
	Normalized Score	77.61	80.08
1 Random Row	Exact Match	48	39
	Score	85.00	79.60
	Normalized Score	85.00	79.60
3 Random Rows	Exact Match	21	15
	Score	81.85	77.12
	Normalized Score	81.85	77.12
1 Random Column	Exact Match	21	33
	Score	54.20	71.90
	Normalized Score	54.30	71.90
3 Random Columns	Exact Match	7	6
	Score	69.00	66.92
	Normalized Score	69.00	66.92
Full	Exact Match	0	0
	Score	49.34	53.51
	Normalized Score	49.34	53.51
Average	Exact Match	28.00	29.00
	Score	71.40	73.83
	Normalized Score	71.79	73.83
<b>Overall Average</b>	<b>Exact Match</b>	30.06	29.25
	<b>Score</b>	75.74	76.86
	<b>Normalized Score</b>	75.93	76.86

## E ADDITIONAL FEW-SHOT EXPERIMENT RESULTS

## E.1 FEW-SHOT (1, 3) RESULTS FOR BEST PERFORMING MODEL: CLAUDE OPUS 4 ON ALL TASKS

Table 5: Exact Match, Overlap- and Normalized Overlap-Scores across all difficulties and tasks for Claude Opus 4 comparing zero-few-shot, 1-shot and 3-shot prompt setups.

Task	Metric	Claude Opus 4	Claude Opus 4 (Few Shot 1)	Claude Opus 4 (Few Shot 3)
Difficulty: Easy				
1 Random Cell	Exact Match	82	90	90
	Score	82.00	90.00	90.00
	Normalized Score	82.00	90.00	90.00
5 Random Cells	Exact Match	45	49	63
	Score	84.27	87.10	90.27
	Normalized Score	84.27	87.10	90.27
10 Random Cells	Exact Match	15	28	44
	Score	78.72	84.96	89.99
	Normalized Score	78.72	84.96	89.99
1 Random Row	Exact Match	52	67	75
	Score	90.50	94.10	95.50
	Normalized Score	90.50	94.10	95.50
3 Random Rows	Exact Match	35	35	46
	Score	92.45	92.55	93.47
	Normalized Score	92.45	92.55	93.47
1 Random Column	Exact Match	23	34	52
	Score	68.50	83.20	89.70
	Normalized Score	68.50	83.20	89.70
3 Random Columns	Exact Match	5	5	17
	Score	74.05	81.05	86.25
	Normalized Score	74.05	81.05	86.25
Full	Exact Match	0	0	4
	Score	70.09	72.90	83.30
	Normalized Score	70.09	72.90	83.30
Average	Exact Match	32.12	38.50	48.88
	Score	80.07	85.73	89.81
	Normalized Score	80.07	85.73	89.81
Difficulty: Hard				
1 Random Cell	Exact Match	77	83	88
	Score	74.00	83.00	88.00
	Normalized Score	77.00	83.00	88.00
5 Random Cells	Exact Match	38	50	62
	Score	81.20	84.60	88.20
	Normalized Score	81.20	84.60	88.20
10 Random Cells	Exact Match	13	33	42
	Score	77.61	86.67	89.19
	Normalized Score	77.61	86.67	89.19
1 Random Row	Exact Match	49	54	55
	Score	86.00	88.40	73.20
	Normalized Score	86.00	88.40	85.50
3 Random Rows	Exact Match	21	28	30
	Score	81.85	85.03	88.88
	Normalized Score	81.85	85.03	88.88
1 Random Column	Exact Match	21	34	38
	Score	54.20	76.70	70.60
	Normalized Score	54.30	76.70	76.70
3 Random Columns	Exact Match	7	8	8
	Score	69.00	73.05	80.55
	Normalized Score	69.00	73.05	80.55
Full	Exact Match	0	0	0
	Score	49.34	64.40	68.42
	Normalized Score	49.34	64.40	68.42
Average	Exact Match	28.25	36.25	40.38
	Score	71.65	80.23	80.88
	Normalized Score	72.04	80.23	83.18
<b>Overall Average</b>	<b>Exact Match</b>	30.19	37.38	44.62
	<b>Score</b>	75.86	82.98	85.35
	<b>Normalized Score</b>	76.05	82.98	86.50

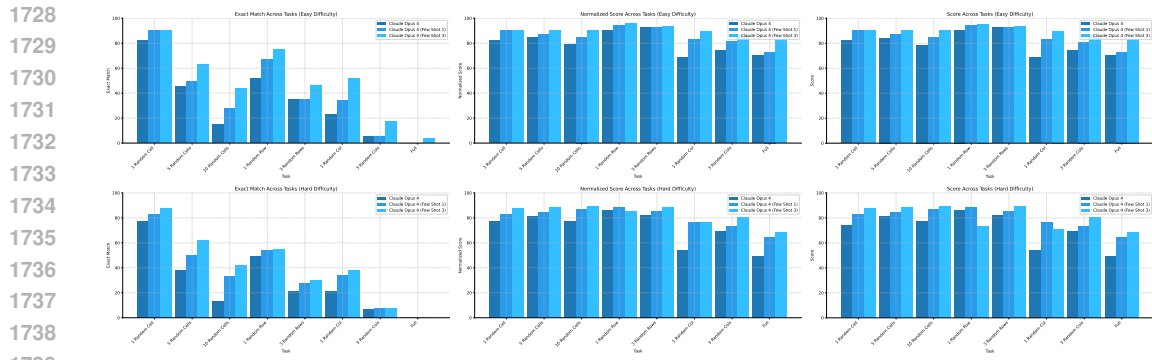


Figure 12: Few-Shot Experiment Results for easy (top) and hard (bottom) difficulty across all tasks for Claude Opus 4. The results table can be found in the Appendix E.

1728  
1729  
1730  
1731  
1732  
1733  
1734  
1735  
1736  
1737  
1738  
1739  
1740  
1741  
1742  
1743  
1744  
1745  
1746  
1747  
1748  
1749  
1750  
1751  
1752  
1753  
1754  
1755  
1756  
1757  
1758  
1759  
1760  
1761  
1762  
1763  
1764  
1765  
1766  
1767  
1768  
1769  
1770  
1771  
1772  
1773  
1774  
1775  
1776  
1777  
1778  
1779  
1780  
1781

1782 F ADDITIONAL PHYSICS-ENHANCED AND -NEUTRAL EXPERIMENT RESULTS

1783

1784

1785

1786 F.1 PHYSICS-NEUTRAL PROMPT & COMPLETION SAMPLE: 1 RANDOM CELL EASY

1787

1788

1789 **Prompt:** You are given a grid of cells. Each cell can have one of the following states:

1790 - 'L' indicates a special marker.

1791 - 'V' indicates an empty cell.

1792 - 'S' indicates a fixed marker.

1793

1794 The goal is to predict the correct grid configuration by filling in all 'V' cells with either '1' (solid)  
1795 or '0' (empty), based on the surrounding structure and general consistency rules.

1796

1797 Relevant rules for reasoning:

1798 - Filled cells should form continuous connections between relevant markers when necessary.

1799 - Avoid placing isolated filled cells that are not connected to anything meaningful.

1800 - Minimize the number of filled cells while maintaining a coherent and connected structure.

1801

1802 Below is the input grid with masked regions:

1803

1804 L L L L L 0 0 0 0 0

1805 0 1 1 1 0 0 0 0 0 0

1806 0 1 1 1 0 0 0 0 0 0

1807 0 0 1 0 0 0 0 0 0 0

1808 0 0 1 0 0 0 0 0 0 0

1809 0 0 1 0 0 0 0 0 0 0

1810 0 0 1 0 0 0 0 0 0 0

1811 V 0 1 1 1 0 0 0 0 0

1812 0 0 1 S S S 0 0 0 0

1813

1814 Please output the completed grid by replacing all 'V' cells with either '1' (solid) or '0' (empty).

1815 Maintain the same format as the input: one row per line, cells separated by spaces, and the total number of

1816 rows and columns unchanged. Return only the completed grid without any additional explanation.

1817

1818 **Model Completion (differences in red):**

1819 L L L L L 0 0 0 0 0

1820 0 1 1 1 0 0 0 0 0 0

1821 0 1 1 1 0 0 0 0 0 0

1822 0 0 1 0 0 0 0 0 0 0

1823 0 0 1 0 0 0 0 0 0 0

1824 0 0 1 0 0 0 0 0 0 0

1825 0 0 1 0 0 0 0 0 0 0

1826 0 0 1 0 0 0 0 0 0 0

1827 0 0 1 0 0 0 0 0 0 0

1828 0 0 1 S S S 0 0 0 0

1829 0 0 1 S S S 0 0 0 0

1830

1831 **Ground Truth:**

1832 L L L L L 0 0 0 0 0

1833 0 1 1 1 0 0 0 0 0 0

1834 0 1 1 1 0 0 0 0 0 0

1835 0 0 1 0 0 0 0 0 0 0

0 0 1 0 0 0 0 0 0 0

0 0 1 0 0 0 0 0 0 0

0 0 1 0 0 0 0 0 0 0

0 0 1 0 0 0 0 0 0 0

0 0 1 0 0 0 0 0 0 0

0 0 1 1 1 0 0 0 0 0

0 0 1 S S S 0 0 0 0

1836 F.2 PHYSICS-ENHANCED PROMPT & COMPLETION SAMPLE: 1 RANDOM CELL EASY  
 1837

1838 **Prompt:** You are given a structural material distribution represented as a grid. Each cell can have one of  
 1839 the following states:

- 1840 - 'L' indicates applied load.  
 1841 - 'V' indicates void.  
 1842 - 'S' indicates support.

1843  
 1844 The goal is to predict the correct material distribution by filling in all 'V' cells with either '1' (solid)  
 1845 or '0' (empty), based on the surrounding structure and implicit physical reasoning (such as load paths,  
 1846 supports, and forces).

1847 Relevant physical knowledge for reasoning:

- 1848 - Loads ('L') create forces that must be transferred through continuous material paths to supports ('S').  
 1849 - Stress follows the shortest stiff path from loads to supports.  
 1850 - Any material cell that is disconnected from both loads and supports does not carry force and should be  
 1851 avoided.  
 1852 - Structures should satisfy equilibrium: all loads must eventually be resisted by at least one support.  
 1853 - Use as little material as possible while still maintaining at least one valid load path.

1854  
 1855 Below is the input grid with masked regions:

1856  
 1857 0 0 L L L 0 0 0 0 0  
 1858 0 0 0 1 0 0 0 0 0  
 1859 0 0 0 V 0 0 0 0 0  
 1860 0 0 0 1 0 0 0 0 0  
 1861 0 0 0 1 0 0 0 0 0  
 1862 0 0 0 1 0 0 0 0 0  
 1863 0 0 1 1 1 0 0 0 0  
 1864 0 0 1 1 1 0 0 0 0  
 1865 S S S S S S S S S 0

1866  
 1867 Please output the completed grid by replacing all 'V' cells with either '1' (solid) or '0' (empty).  
 1868 Maintain the same format as the input: one row per line, cells separated by spaces, and the total number of  
 1869 rows and columns unchanged. Return only the completed grid without any additional explanation.

1870 **Model Completion (differences in red):**

1871 0 0 L L L 0 0 0 0 0  
 1872 0 0 0 1 0 0 0 0 0  
 1873 0 0 0 1 0 0 0 0 0  
 1874 0 0 0 1 0 0 0 0 0  
 1875 0 0 0 1 0 0 0 0 0  
 1876 0 0 0 1 0 0 0 0 0  
 1877 0 0 1 1 1 0 0 0 0  
 1878 0 0 1 1 1 0 0 0 0  
 1879 S S S S S S S S S 0

1880 **Ground Truth:**

1881 0 0 L L L 0 0 0 0 0  
 1882 0 0 0 1 0 0 0 0 0  
 1883 0 0 0 1 0 0 0 0 0  
 1884 0 0 0 1 0 0 0 0 0  
 1885 0 0 0 1 0 0 0 0 0  
 1886 0 0 0 1 0 0 0 0 0  
 1887 0 0 1 1 1 0 0 0 0  
 1888 0 0 1 1 1 0 0 0 0  
 1889 0 0 1 1 1 0 0 0 0  
 S S S S S S S S S 0

1890 F.3 PHYSICS-ENHANCED AND -NEUTRAL RESULTS FOR BEST PERFORMING MODEL:  
 1891 CLAUDE OPUS 4 ON ALL TASKS  
 1892  
 1893

1894 Table 6: Exact Match, Overlap- and Normalized Overlap-Scores across all difficulties and tasks for  
 1895 Gemini 2.5 Pro comparing default, physics-enhanced and -neutral prompt setups.

1896 Task	1896 Metric	1896 Gemini 2.5 Pro	1896 Gemini 2.5 Pro (Physics Enhanced Prompt)	1896 Gemini 2.5 Pro (Physics Neutral Prompt)
1897 Difficulty: Easy				
1898 1 Random Cell	Exact Match	38	77	80
	Score	-702.00	62.00	80.00
	Normalized Score	38.00	77.00	80.00
1899 5 Random Cells	Exact Match	18	44	37
	Score	-61.32	83.07	79.70
	Normalized Score	39.15	83.07	79.70
1901 10 Random Cells	Exact Match	8	13	11
	Score	12.23	80.34	78.70
	Normalized Score	44.21	80.34	79.30
1903 1 Random Row	Exact Match	30	46	60
	Score	41.20	90.10	91.50
	Normalized Score	61.00	90.10	91.50
1904 3 Random Rows	Exact Match	29	28	29
	Score	89.73	91.02	89.57
	Normalized Score	89.73	91.02	89.57
1906 1 Random Column	Exact Match	26	22	14
	Score	69.10	66.80	52.70
	Normalized Score	71.60	66.80	52.70
1908 3 Random Columns	Exact Match	3	5	4
	Score	69.08	74.77	65.23
	Normalized Score	69.73	74.77	65.23
1910 Full	Exact Match	0	0	0
	Score	71.59	68.81	73.84
	Normalized Score	71.59	68.81	73.84
1912 Average	Exact Match	19.00	29.38	29.38
	Score	-51.30	77.11	76.40
	Normalized Score	60.63	78.99	76.48
1914 Difficulty: Hard				
1915 1 Random Cell	Exact Match	75	72	71
	Score	33.00	32.00	66.00
	Normalized Score	75.00	72.00	71.00
1916 5 Random Cells	Exact Match	37	26	33
	Score	76.65	71.45	71.55
	Normalized Score	77.85	74.05	73.75
1918 10 Random Cells	Exact Match	14	7	7
	Score	75.97	73.43	59.97
	Normalized Score	75.97	73.87	70.49
1920 1 Random Row	Exact Match	46	42	38
	Score	79.80	79.20	75.40
	Normalized Score	79.80	79.20	76.80
1922 3 Random Rows	Exact Match	12	16	22
	Score	77.83	75.25	81.75
	Normalized Score	77.83	75.25	81.75
1924 1 Random Column	Exact Match	26	20	17
	Score	60.30	41.20	40.50
	Normalized Score	60.30	44.90	46.00
1925 3 Random Columns	Exact Match	3	7	0
	Score	58.02	62.90	58.40
	Normalized Score	58.02	62.90	58.40
1927 Full	Exact Match	0	0	0
	Score	40.48	37.50	50.83
	Normalized Score	40.56	37.52	50.88
1929 Average	Exact Match	26.62	23.75	23.50
	Score	62.76	59.12	63.05
	Normalized Score	68.17	64.96	66.13
1931 Overall Average	Exact Match	22.81	26.56	26.44
	Score	5.73	68.11	69.73
	Normalized Score	64.40	71.97	71.31

1934  
 1935  
 1936  
 1937  
 1938  
 1939  
 1940  
 1941  
 1942  
 1943



1944 **G MODEL SAMPLE PROMPTS & COMPLETIONS**

1945  
1946 **G.1 CLAUDE 3.7 SONNET - 1 RANDOM CELL EASY: SAMPLE 1**

1948 **Prompt:** You are given a structural material distribution represented as a grid. Each cell can have one of  
1949 the following states:  
1950 - 'L' indicates applied load.  
1951 - 'V' indicates void.  
1952 - 'S' indicates support.

1953  
1954 The goal is to predict the correct material distribution by filling in all 'V' cells with either '1' (solid)  
1955 or '0' (empty), based on the surrounding structure and implicit physical reasoning (such as load paths,  
1956 supports, and forces).

1957 **Important:** The completed structure should use as little material as possible while remaining stable and  
1958 plausible for carrying the applied forces. Minimize material usage unless necessary for structural support.

1959  
1960 Below is the input grid with masked regions:

1961  
1962 0 0 L L L 0 0 0 0 0  
1963 0 0 0 1 0 0 0 0 0 0  
1964 0 0 0 V 0 0 0 0 0 0  
1965 0 0 0 1 0 0 0 0 0 0  
1966 0 0 0 1 0 0 0 0 0 0  
1967 0 0 0 1 0 0 0 0 0 0  
1968 0 0 1 1 1 0 0 0 0 0  
1969 0 0 1 1 1 0 0 0 0 0  
1970 0 0 1 1 1 0 0 0 0 0  
1971 S S S S S S S S S 0

1972 Please output the completed grid by replacing all 'V' cells with either '1' (solid) or '0' (empty).  
1973 Maintain the same format as the input: one row per line, cells separated by spaces, and the total number of  
1974 rows and columns unchanged.  
1975 Return only the completed grid without any additional explanation.

1976 **Model Completion (differences in red):**

1977 0 0 L L L 0 0 0 0 0  
1978 0 0 0 1 0 0 0 0 0 0  
1979 0 0 0 1 0 0 0 0 0 0  
1980 0 0 0 1 0 0 0 0 0 0  
1981 0 0 0 1 0 0 0 0 0 0  
1982 0 0 1 1 1 0 0 0 0 0  
1983 0 0 1 1 1 0 0 0 0 0  
1984 0 0 1 1 1 0 0 0 0 0  
1985 S S S S S S S S S 0

1986 **Ground Truth:**

1987 0 0 L L L 0 0 0 0 0  
1988 0 0 0 1 0 0 0 0 0 0  
1989 0 0 0 1 0 0 0 0 0 0  
1990 0 0 0 1 0 0 0 0 0 0  
1991 0 0 0 1 0 0 0 0 0 0  
1992 0 0 1 1 1 0 0 0 0 0  
1993 0 0 1 1 1 0 0 0 0 0  
1994 0 0 1 1 1 0 0 0 0 0  
1995 S S S S S S S S S 0

1996  
1997

1998  
1999  
2000  
2001  
2002  
2003  
2004  
2005  
2006  
2007  
2008  
2009  
2010  
2011  
2012  
2013  
2014  
2015  
2016  
2017  
2018  
2019  
2020  
2021  
2022  
2023  
2024  
2025  
2026  
2027  
2028  
2029  
2030  
2031  
2032  
2033  
2034  
2035  
2036  
2037  
2038  
2039  
2040  
2041  
2042  
2043  
2044  
2045  
2046  
2047  
2048  
2049  
2050  
2051

## G.2 CLAUDE 3.7 SONNET - 1 RANDOM CELL EASY: SAMPLE 2

**Prompt:** You are given a structural material distribution represented as a grid. Each cell can have one of the following states:  
- 'L' indicates applied load.  
- 'V' indicates void.  
- 'S' indicates support.

The goal is to predict the correct material distribution by filling in all 'V' cells with either '1' (solid) or '0' (empty), based on the surrounding structure and implicit physical reasoning (such as load paths, supports, and forces).

Important: The completed structure should use as little material as possible while remaining stable and plausible for carrying the applied forces. Minimize material usage unless necessary for structural support.

Below is the input grid with masked regions:

```
0 0 0 0 0 0 L L L 0
0 0 0 0 0 0 0 1 0 0
0 0 0 0 0 0 0 1 0 0
0 0 0 0 0 0 0 1 0 0
0 0 0 0 0 0 0 1 0 0
0 0 0 0 0 0 0 1 0 0
0 0 0 0 1 1 1 1 0 0
0 0 0 1 1 1 1 1 0 0
0 0 1 1 1 1 1 0 0 0
0 0 1 1 1 1 1 0 0 0
0 1 1 1 1 1 V 0 0 0
S S S S S 1 0 0 0 0
```

Please output the completed grid by replacing all 'V' cells with either '1' (solid) or '0' (empty). Maintain the same format as the input: one row per line, cells separated by spaces, and the total number of rows and columns unchanged.  
Return only the completed grid without any additional explanation.

### Model Completion (differences in red):

```
0 0 0 0 0 0 L L L 0
0 0 0 0 0 0 0 1 0 0
0 0 0 0 0 0 0 1 0 0
0 0 0 0 0 0 0 1 0 0
0 0 0 0 0 0 0 1 0 0
0 0 0 0 1 1 1 1 0 0
0 0 0 1 1 1 1 1 0 0
0 0 1 1 1 1 1 0 0 0
0 1 1 1 1 1 1 0 0 0
S S S S S 1 0 0 0 0
```

### Ground Truth:

```
0 0 0 0 0 0 L L L 0
0 0 0 0 0 0 0 1 0 0
0 0 0 0 0 0 0 1 0 0
0 0 0 0 0 0 0 1 0 0
0 0 0 0 0 0 0 1 0 0
0 0 0 0 1 1 1 1 0 0
0 0 0 1 1 1 1 1 0 0
0 0 1 1 1 1 1 0 0 0
0 1 1 1 1 1 0 0 0 0
S S S S S 1 0 0 0 0
```

2052 **G.3 CLAUDE 3.7 SONNET - 1 RANDOM CELL HARD**

2053 **Prompt:** You are given a structural material distribution represented as a grid. Each cell can have one of  
 2054 the following states:  
 2055 - 'L' indicates applied load.  
 2056 - 'V' indicates void.  
 2057 - 'S' indicates support.

2058  
 2059 The goal is to predict the correct material distribution by filling in all 'V' cells with a floating point  
 2060 number between 0 and 1, with one decimal place (e.g., 0.0, 0.1, 0.2, ..., 1.0), based on the surrounding  
 2061 structure and implicit physical reasoning (such as load paths, supports, and forces).  
 2062

2063 **Important:** The completed structure should use as little material as possible while remaining stable and  
 2064 plausible for carrying the applied forces. Minimize material usage unless necessary for structural support.  
 2065

2066 Below is the input grid with masked regions:

2067  
 2068 0.0 0.0 0.0 L L L 0.0 0.0 0.0 0.0  
 2069 0.0 0.0 0.0 0.0 0.5 0.0 0.0 0.0 0.0 0.0  
 2070 0.0 0.0 0.0 0.0 0.5 0.0 0.0 0.0 0.0 0.0  
 2071 0.0 0.0 0.0 0.0 0.5 0.0 0.0 0.0 0.0 0.0  
 2072 0.0 0.0 0.0 0.0 0.5 0.0 0.0 0.0 0.0 0.0  
 2073 0.0 0.0 V 0.0 0.5 0.0 0.0 0.0 0.0 0.0  
 2074 0.0 0.0 0.0 0.0 0.6 0.0 0.0 0.0 0.0 0.0  
 2075 0.0 0.1 0.6 0.6 0.6 0.0 0.0 0.0 0.0 0.0  
 2076 0.0 0.7 0.2 0.8 0.0 0.0 0.0 0.0 0.0 0.0  
 2077 S S S 0.1 0.0 0.0 0.0 0.0 0.0 0.0

2078 Please output the completed grid by replacing all 'V' cells with a floating point number between 0 and 1,  
 2079 with one decimal place (e.g., 0.0, 0.1, 0.2, ..., 1.0).  
 2080 Maintain the same format as the input: one row per line, cells separated by spaces, and the total number of  
 2081 rows and columns unchanged.

2082 Return only the completed grid without any additional explanation.

2083 **Model Completion (differences in red):**

2084 0.0 0.0 0.0 L L L 0.0 0.0 0.0 0.0  
 2085 0.0 0.0 0.0 0.0 0.5 0.0 0.0 0.0 0.0 0.0  
 2086 0.0 0.0 0.0 0.0 0.5 0.0 0.0 0.0 0.0 0.0  
 2087 0.0 0.0 0.0 0.0 0.5 0.0 0.0 0.0 0.0 0.0  
 2088 0.0 0.0 0.3 0.0 0.5 0.0 0.0 0.0 0.0 0.0  
 2089 0.0 0.0 0.0 0.0 0.6 0.0 0.0 0.0 0.0 0.0  
 2090 0.0 0.1 0.6 0.6 0.6 0.0 0.0 0.0 0.0 0.0  
 2091 0.0 0.7 0.2 0.8 0.0 0.0 0.0 0.0 0.0 0.0  
 2092 S S S 0.1 0.0 0.0 0.0 0.0 0.0 0.0

2093 **Ground Truth:**

2094 0.0 0.0 0.0 L L L 0.0 0.0 0.0 0.0  
 2095 0.0 0.0 0.0 0.0 0.5 0.0 0.0 0.0 0.0 0.0  
 2096 0.0 0.0 0.0 0.0 0.5 0.0 0.0 0.0 0.0 0.0  
 2097 0.0 0.0 0.0 0.0 0.5 0.0 0.0 0.0 0.0 0.0  
 2098 0.0 0.0 0.0 0.0 0.5 0.0 0.0 0.0 0.0 0.0  
 2099 0.0 0.0 0.0 0.0 0.6 0.0 0.0 0.0 0.0 0.0  
 2100 0.0 0.1 0.6 0.6 0.6 0.0 0.0 0.0 0.0 0.0  
 2101 0.0 0.7 0.2 0.8 0.0 0.0 0.0 0.0 0.0 0.0  
 2102 S S S 0.1 0.0 0.0 0.0 0.0 0.0 0.0

2103  
 2104  
 2105

## 2106 G.4 CLAUDE 3.7 SONNET - FULL EASY

2107

2108 **Prompt:** You are given a structural material distribution represented as a grid. Each cell can have one of  
 2109 the following states:

2110 - 'L' indicates applied load.

2111 - 'V' indicates void.

2112 - 'S' indicates support.

2113

2114 The goal is to predict the correct material distribution by filling in all 'V' cells with either '1' (solid)  
 2115 or '0' (empty), based on the surrounding structure and implicit physical reasoning (such as load paths,  
 2116 supports, and forces).

2117

2118 Important: The completed structure should use as little material as possible while remaining stable and  
 2119 plausible for carrying the applied forces. Minimize material usage unless necessary for structural support.

2119

2120 Below is the input grid with masked regions:

2121

2122 0 L L L L L 0 0 0 0

2123 V V V V V V V V V V

2124 V V V V V V V V V V

2125 V V V V V V V V V V

2126 V V V V V V V V V V

2127 V V V V V V V V V V

2128 V V V V V V V V V V

2129 V V V V V V V V V V

2130 0 S S S S S 0 0 0 0

2131

2132 Please output the completed grid by replacing all 'V' cells with either '1' (solid) or '0' (empty).

2133 Maintain the same format as the input: one row per line, cells separated by spaces, and the total number of  
 2134 rows and columns unchanged.

2134

2135 Return only the completed grid without any additional explanation.

2136 **Model Completion (differences in red):**

2137 0 L L L L L 0 0 0 0

2138 0 1 1 1 1 1 0 0 0 0

2139 0 1 0 0 0 1 0 0 0 0

2140 0 1 0 0 0 1 0 0 0 0

2141 0 1 0 0 0 1 0 0 0 0

2142 0 1 0 0 0 1 0 0 0 0

2143 0 1 0 0 0 1 0 0 0 0

2144 0 1 1 1 1 1 0 0 0 0

2145 0 S S S S S 0 0 0 0

2146 **Ground Truth:**

2147 0 L L L L L 0 0 0 0

2148 0 0 1 1 1 0 0 0 0 0

2149 0 0 1 1 1 0 0 0 0 0

2150 0 0 0 1 0 0 0 0 0 0

2151 0 0 0 1 0 0 0 0 0 0

2152 0 0 0 1 0 0 0 0 0 0

2153 0 0 0 1 0 0 0 0 0 0

2154 0 0 0 1 0 0 0 0 0 0

2155 0 S S S S S 0 0 0 0

2156

2157

2158

2159

2160 **G.5 CLAUDE 3.7 SONNET - FULL HARD**

2161

2162 **Prompt:** You are given a structural material distribution represented as a grid. Each cell can have one of

2163 the following states:

2164 - 'L' indicates applied load.

2165 - 'V' indicates void.

2166 - 'S' indicates support.

2167

2168 The goal is to predict the correct material distribution by filling in all 'V' cells with a floating point

2169 number between 0 and 1, with one decimal place (e.g., 0.0, 0.1, 0.2, ..., 1.0), based on the surrounding

2170 structure and implicit physical reasoning (such as load paths, supports, and forces).

2171 **Important:** The completed structure should use as little material as possible while remaining stable and

2172 plausible for carrying the applied forces. Minimize material usage unless necessary for structural support.

2173

2174 Below is the input grid with masked regions:

2175

```
2176 0.0 L L L L L L L L L
2177 V V V V V V V V V V
2178 V V V V V V V V V
2179 V V V V V V V V V
2180 V V V V V V V V V
2181 V V V V V V V V V
2182 V V V V V V V V V
2183 V V V V V V V V V
2184 0.0 0.0 S S S S S 0.0 0.0 0.0
```

2185

2186 Please output the completed grid by replacing all 'V' cells with a floating point number between 0 and 1,

2187 with one decimal place (e.g., 0.0, 0.1, 0.2, ..., 1.0).

2188 Maintain the same format as the input: one row per line, cells separated by spaces, and the total number of

2189 rows and columns unchanged.

2190 Return only the completed grid without any additional explanation.

2191 **Model Completion (differences in red):**

```
2192 0.0 L L L L L L L L L
2193 0.0 0.0 0.1 0.2 0.6 0.6 0.2 0.1 0.0 0.0
2194 0.0 0.0 0.2 0.3 0.8 0.8 0.3 0.2 0.0 0.0
2195 0.0 0.0 0.3 0.5 1.0 1.0 0.5 0.3 0.0 0.0
2196 0.0 0.0 0.4 0.7 1.0 1.0 0.7 0.4 0.0 0.0
2197 0.0 0.0 0.5 0.9 1.0 1.0 0.9 0.5 0.0 0.0
2198 0.0 0.0 0.6 1.0 1.0 1.0 1.0 0.6 0.0 0.0
2199 0.0 0.0 0.7 1.0 1.0 1.0 1.0 0.7 0.0 0.0
2200 0.0 0.0 0.9 1.0 1.0 1.0 1.0 0.9 0.0 0.0
```

2201 **Ground Truth:**

```
2202 0.0 L L L L L L L L L
2203 0.0 0.0 0.2 0.3 0.0 0.3 0.0 0.3 0.2 0.0
2204 0.0 0.0 0.0 0.4 0.2 0.3 0.2 0.4 0.0 0.0
2205 0.0 0.0 0.0 0.2 0.4 0.0 0.4 0.2 0.0 0.0
2206 0.0 0.0 0.0 0.0 0.4 0.0 0.4 0.0 0.0 0.0
2207 0.0 0.0 0.0 0.0 0.2 0.5 0.2 0.0 0.0 0.0
2208 0.0 0.0 0.0 0.0 0.0 0.6 0.0 0.0 0.0 0.0
2209 0.0 0.0 0.0 0.0 0.0 0.6 0.0 0.0 0.0 0.0
2210 0.0 0.0 S S S S S 0.0 0.0 0.0
```

2211

2212

2213

2214 G.6 DEEPSEEK-R1 - 1 RANDOM CELL EASY: SAMPLE 1  
2215

2216 **Prompt:** You are given a structural material distribution represented as a grid. Each cell can have one of  
2217 the following states:

- 2218 - 'L' indicates applied load.
- 2219 - 'V' indicates void.
- 2220 - 'S' indicates support.

2221 The goal is to predict the correct material distribution by filling in all 'V' cells with either '1' (solid)  
2222 or '0' (empty), based on the surrounding structure and implicit physical reasoning (such as load paths,  
2223 supports, and forces).  
2224

2225 Important: The completed structure should use as little material as possible while remaining stable and  
2226 plausible for carrying the applied forces. Minimize material usage unless necessary for structural support.  
2227

2228 Below is the input grid with masked regions:

```
2229
2230 0 0 L L L 0 0 0 0 0
2231 0 0 0 1 0 0 0 0 0 0
2232 0 0 0 V 0 0 0 0 0 0
2233 0 0 0 1 0 0 0 0 0 0
2234 0 0 0 1 0 0 0 0 0 0
2235 0 0 0 1 0 0 0 0 0 0
2236 0 0 1 1 1 0 0 0 0 0
2237 0 0 1 1 1 0 0 0 0 0
2238 S S S S S S S S S 0
2239
```

2240 Please output the completed grid by replacing all 'V' cells with either '1' (solid) or '0' (empty).

2241 Maintain the same format as the input: one row per line, cells separated by spaces, and the total number of  
2242 rows and columns unchanged.

2243 Return only the completed grid without any additional explanation.

2244 **Model Completion (differences in red):**

```
2245 0 0 L L L 0 0 0 0 0
2246 0 0 0 1 0 0 0 0 0 0
2247 0 0 0 1 0 0 0 0 0 0
2248 0 0 0 1 0 0 0 0 0 0
2249 0 0 0 1 0 0 0 0 0 0
2250 0 0 1 1 1 0 0 0 0 0
2251 0 0 1 1 1 0 0 0 0 0
2252 0 0 1 1 1 0 0 0 0 0
2253 S S S S S S S S S 0
```

2254 **Ground Truth:**

```
2255 0 0 L L L 0 0 0 0 0
2256 0 0 0 1 0 0 0 0 0 0
2257 0 0 0 1 0 0 0 0 0 0
2258 0 0 0 1 0 0 0 0 0 0
2259 0 0 0 1 0 0 0 0 0 0
2260 0 0 1 1 1 0 0 0 0 0
2261 0 0 1 1 1 0 0 0 0 0
2262 0 0 1 1 1 0 0 0 0 0
2263 S S S S S S S S S 0
2264
2265
2266
2267
```

2268 **G.7 DEEPSEEK-R1 - 1 RANDOM CELL EASY: SAMPLE 2**

2269

2270 **Prompt:** You are given a structural material distribution represented as a grid. Each cell can have one of

2271 the following states:

2272 - 'L' indicates applied load.

2273 - 'V' indicates void.

2274 - 'S' indicates support.

2275

2276 The goal is to predict the correct material distribution by filling in all 'V' cells with either '1' (solid)

2277 or '0' (empty), based on the surrounding structure and implicit physical reasoning (such as load paths,

2278 supports, and forces).

2279 **Important:** The completed structure should use as little material as possible while remaining stable and

2280 plausible for carrying the applied forces. Minimize material usage unless necessary for structural support.

2281

2282 Below is the input grid with masked regions:

2283

2284 L L L L L L 0 0 0 0

2285 0 1 1 1 1 0 0 0 0 0

2286 0 0 1 1 0 0 0 0 0 0

2287 0 0 1 1 0 0 0 0 0 0

2288 0 0 1 1 1 0 0 0 0 0

2289 0 0 1 1 1 1 0 0 0 0

2290 V 0 0 1 1 1 1 0 0 0 0

2291 0 0 0 1 1 1 1 1 0 0

2292 0 0 0 0 1 1 1 1 0 0

2293 0 0 0 0 0 1 S S S 0

2294 Please output the completed grid by replacing all 'V' cells with either '1' (solid) or '0' (empty).

2295 Maintain the same format as the input: one row per line, cells separated by spaces, and the total number of

2296 rows and columns unchanged.

2297 Return only the completed grid without any additional explanation.

2298 **Model Completion (differences in red):**

2299 L L L L L L 0 0 0 0

2300 0 1 1 1 1 0 0 0 0 0

2301 0 0 1 1 0 0 0 0 0 0

2302 0 0 1 1 0 0 0 0 0 0

2303 0 0 1 1 1 0 0 0 0 0

2304 1 0 0 1 1 1 1 0 0 0

2305 0 0 0 1 1 1 1 1 0 0

2306 0 0 0 0 1 1 1 1 0 0

2307 0 0 0 0 0 1 S S S 0

2308 **Ground Truth:**

2309 L L L L L L 0 0 0 0

2310 0 1 1 1 1 0 0 0 0 0

2311 0 0 1 1 0 0 0 0 0 0

2312 0 0 1 1 0 0 0 0 0 0

2313 0 0 1 1 1 0 0 0 0 0

2314 0 0 0 1 1 1 1 0 0 0

2315 0 0 0 1 1 1 1 1 0 0

2316 0 0 0 0 1 1 1 1 0 0

2317 0 0 0 0 0 1 S S S 0

2318

2319

2320

2321

2322 **G.8 DEEPSEEK-R1 - 1 RANDOM CELL HARD**

2323

2324 **Prompt:** You are given a structural material distribution represented as a grid. Each cell can have one of

2325 the following states:

2326 - 'L' indicates applied load.

2327 - 'V' indicates void.

2328 - 'S' indicates support.

2329

2330 The goal is to predict the correct material distribution by filling in all 'V' cells with a floating point

2331 number between 0 and 1, with one decimal place (e.g., 0.0, 0.1, 0.2, ..., 1.0), based on the surrounding

2332 structure and implicit physical reasoning (such as load paths, supports, and forces).

2333 **Important:** The completed structure should use as little material as possible while remaining stable and

2334 plausible for carrying the applied forces. Minimize material usage unless necessary for structural support.

2335

2336 Below is the input grid with masked regions:

2337

2338 0.0 0.0 0.0 L L L 0.0 0.0 0.0 0.0

2339 0.0 0.0 0.0 0.0 0.5 0.0 0.0 0.0 0.0 0.0

2340 0.0 0.0 0.0 0.0 0.5 0.0 0.0 0.0 0.0 0.0

2341 0.0 0.0 0.0 0.0 0.5 0.0 0.0 0.0 0.0 0.0

2342 0.0 0.0 0.0 0.0 0.5 0.0 0.0 0.0 0.0 0.0

2343 0.0 0.0 V 0.0 0.5 0.0 0.0 0.0 0.0 0.0

2344 0.0 0.0 0.0 0.0 0.6 0.0 0.0 0.0 0.0 0.0

2345 0.0 0.1 0.6 0.6 0.6 0.0 0.0 0.0 0.0 0.0

2346 0.0 0.7 0.2 0.8 0.0 0.0 0.0 0.0 0.0 0.0

2347 S S S 0.1 0.0 0.0 0.0 0.0 0.0 0.0

2348 Please output the completed grid by replacing all 'V' cells with a floating point number between 0 and 1,

2349 with one decimal place (e.g., 0.0, 0.1, 0.2, ..., 1.0).

2350 Maintain the same format as the input: one row per line, cells separated by spaces, and the total number of

2351 rows and columns unchanged.

2352 Return only the completed grid without any additional explanation.

2353 **Model Completion (differences in red):**

2354 0.0 0.0 0.0 L L L 0.0 0.0 0.0 0.0

2355 0.0 0.0 0.0 0.0 0.5 0.0 0.0 0.0 0.0 0.0

2356 0.0 0.0 0.0 0.0 0.5 0.0 0.0 0.0 0.0 0.0

2357 0.0 0.0 0.0 0.0 0.5 0.0 0.0 0.0 0.0 0.0

2358 0.0 0.0 0.3 0.0 0.5 0.0 0.0 0.0 0.0 0.0

2359 0.0 0.0 0.0 0.0 0.6 0.0 0.0 0.0 0.0 0.0

2360 0.0 0.1 0.6 0.6 0.6 0.0 0.0 0.0 0.0 0.0

2361 0.0 0.7 0.2 0.8 0.0 0.0 0.0 0.0 0.0 0.0

2362 S S S 0.1 0.0 0.0 0.0 0.0 0.0 0.0

2363 **Ground Truth:**

2364 0.0 0.0 0.0 L L L 0.0 0.0 0.0 0.0

2365 0.0 0.0 0.0 0.0 0.5 0.0 0.0 0.0 0.0 0.0

2366 0.0 0.0 0.0 0.0 0.5 0.0 0.0 0.0 0.0 0.0

2367 0.0 0.0 0.0 0.0 0.5 0.0 0.0 0.0 0.0 0.0

2368 0.0 0.0 0.0 0.0 0.5 0.0 0.0 0.0 0.0 0.0

2369 0.0 0.0 0.0 0.0 0.6 0.0 0.0 0.0 0.0 0.0

2370 0.0 0.1 0.6 0.6 0.6 0.0 0.0 0.0 0.0 0.0

2371 0.0 0.7 0.2 0.8 0.0 0.0 0.0 0.0 0.0 0.0

2372 S S S 0.1 0.0 0.0 0.0 0.0 0.0 0.0

2373

2374

2375



2376 **G.9 DEEPSEEK-R1 - FULL EASY**  
 2377

2378 **Prompt:** You are given a structural material distribution represented as a grid. Each cell can have one of  
 2379 the following states:  
 2380 - 'L' indicates applied load.  
 2381 - 'V' indicates void.  
 2382 - 'S' indicates support.

2383  
 2384 The goal is to predict the correct material distribution by filling in all 'V' cells with either '1' (solid)  
 2385 or '0' (empty), based on the surrounding structure and implicit physical reasoning (such as load paths,  
 2386 supports, and forces).

2387 **Important:** The completed structure should use as little material as possible while remaining stable and  
 2388 plausible for carrying the applied forces. Minimize material usage unless necessary for structural support.  
 2389

2390 Below is the input grid with masked regions:

2391  
 2392 0 L L L L L 0 0 0 0  
 2393 V V V V V V V V V V  
 2394 V V V V V V V V V V  
 2395 V V V V V V V V V V  
 2396 V V V V V V V V V V  
 2397 V V V V V V V V V V  
 2398 V V V V V V V V V V  
 2399 V V V V V V V V V V  
 2400 0 S S S S S 0 0 0

2401  
 2402 Please output the completed grid by replacing all 'V' cells with either '1' (solid) or '0' (empty).  
 2403 Maintain the same format as the input: one row per line, cells separated by spaces, and the total number of  
 2404 rows and columns unchanged.  
 2405 Return only the completed grid without any additional explanation.

2406 **Model Completion (differences in red):**

2407 0 L L L L L 0 0 0 0  
 2408 1 1 1 1 1 1 1 1 1 1  
 2409 1 1 1 1 1 1 1 1 1 1  
 2410 1 1 1 1 1 1 1 1 1 1  
 2411 1 1 1 1 1 1 1 1 1 1  
 2412 1 1 1 1 1 1 1 1 1 1  
 2413 1 1 1 1 1 1 1 1 1 1  
 2414 1 1 1 1 1 1 1 1 1 1  
 2415 0 S S S S S 0 0 0 0

2416 **Ground Truth:**

2417 0 L L L L L 0 0 0 0  
 2418 0 0 1 1 1 0 0 0 0 0  
 2419 0 0 1 1 1 0 0 0 0 0  
 2420 0 0 0 1 0 0 0 0 0 0  
 2421 0 0 0 1 0 0 0 0 0 0  
 2422 0 0 0 1 0 0 0 0 0 0  
 2423 0 0 0 1 0 0 0 0 0 0  
 2424 0 0 0 1 0 0 0 0 0 0  
 2425 0 S S S S S 0 0 0 0

2426  
 2427  
 2428  
 2429

2430 **G.10 DEEPSEEK-R1 - FULL HARD**

2431

2432 **Prompt:** You are given a structural material distribution represented as a grid. Each cell can have one of

2433 the following states:

2434 - 'L' indicates applied load.

2435 - 'V' indicates void.

2436 - 'S' indicates support.

2437

2438 The goal is to predict the correct material distribution by filling in all 'V' cells with a floating point

2439 number between 0 and 1, with one decimal place (e.g., 0.0, 0.1, 0.2, ..., 1.0), based on the surrounding

2440 structure and implicit physical reasoning (such as load paths, supports, and forces).

2441 **Important:** The completed structure should use as little material as possible while remaining stable and

2442 plausible for carrying the applied forces. Minimize material usage unless necessary for structural support.

2443

2444 Below is the input grid with masked regions:

2445

2446 0.0 L L L L L L L L L L

2447 V V V V V V V V V V

2448 V V V V V V V V V V

2449 V V V V V V V V V V

2450 V V V V V V V V V V

2451 V V V V V V V V V V

2452 V V V V V V V V V V

2453 V V V V V V V V V V

2454 0.0 0.0 S S S S S 0.0 0.0 0.0

2455

2456 Please output the completed grid by replacing all 'V' cells with a floating point number between 0 and 1,

2457 with one decimal place (e.g., 0.0, 0.1, 0.2, ..., 1.0).

2458 Maintain the same format as the input: one row per line, cells separated by spaces, and the total number of

2459 rows and columns unchanged.

2460 Return only the completed grid without any additional explanation.

2461 **Model Completion (differences in red):**

2462 0.0 L L L L L L L L L L

2463 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0

2464 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0

2465 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0

2466 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0

2467 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0

2468 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0

2469 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0

2470 0.0 0.0 S S S S S 0.0 0.0 0.0

2471 **Ground Truth:**

2472 0.0 L L L L L L L L L L

2473 0.0 0.0 0.2 0.3 0.0 0.3 0.0 0.3 0.2 0.0

2474 0.0 0.0 0.0 0.4 0.2 0.3 0.2 0.4 0.0 0.0

2475 0.0 0.0 0.0 0.2 0.4 0.0 0.4 0.2 0.0 0.0

2476 0.0 0.0 0.0 0.0 0.4 0.0 0.4 0.0 0.0 0.0

2477 0.0 0.0 0.0 0.0 0.2 0.5 0.2 0.0 0.0 0.0

2478 0.0 0.0 0.0 0.0 0.0 0.6 0.0 0.0 0.0 0.0

2479 0.0 0.0 0.0 0.0 0.0 0.6 0.0 0.0 0.0 0.0

2480 0.0 0.0 S S S S S 0.0 0.0 0.0

2481

2482

2483

## 2484 G.11 GPT-4.1 - 1 RANDOM CELL EASY: SAMPLE 1

2485

2486 **Prompt:** You are given a structural material distribution represented as a grid. Each cell can have one of  
 2487 the following states:

2488 - 'L' indicates applied load.

2489 - 'V' indicates void.

2490 - 'S' indicates support.

2491

2492 The goal is to predict the correct material distribution by filling in all 'V' cells with either '1' (solid)  
 2493 or '0' (empty), based on the surrounding structure and implicit physical reasoning (such as load paths,  
 2494 supports, and forces).

2495

2496 Important: The completed structure should use as little material as possible while remaining stable and  
 2497 plausible for carrying the applied forces. Minimize material usage unless necessary for structural support.

2497

2498 Below is the input grid with masked regions:

2499

2500 0 0 L L L 0 0 0 0 0

2501 0 0 0 1 0 0 0 0 0 0

2502 0 0 0 V 0 0 0 0 0 0

2503 0 0 0 1 0 0 0 0 0 0

2504 0 0 0 1 0 0 0 0 0 0

2505 0 0 0 1 0 0 0 0 0 0

2506 0 0 1 1 1 0 0 0 0 0

2507 0 0 1 1 1 0 0 0 0 0

2508 0 0 1 1 1 0 0 0 0 0

2509 S S S S S S S S S 0

2510

Please output the completed grid by replacing all 'V' cells with either '1' (solid) or '0' (empty).

2511

Maintain the same format as the input: one row per line, cells separated by spaces, and the total number of  
 2512 rows and columns unchanged.

2513

Return only the completed grid without any additional explanation.

2514

**Model Completion (differences in red):**

2515

0 0 L L L 0 0 0 0 0

2516

0 0 0 1 0 0 0 0 0 0

2517

0 0 0 1 0 0 0 0 0 0

2518

0 0 0 1 0 0 0 0 0 0

2519

0 0 0 1 0 0 0 0 0 0

2520

0 0 1 1 1 0 0 0 0 0

2521

0 0 1 1 1 0 0 0 0 0

2522

0 0 1 1 1 0 0 0 0 0

2523

S S S S S S S S S 0

2524

**Ground Truth:**

2525

0 0 L L L 0 0 0 0 0

2526

0 0 0 1 0 0 0 0 0 0

2527

0 0 0 1 0 0 0 0 0 0

2528

0 0 0 1 0 0 0 0 0 0

2529

0 0 0 1 0 0 0 0 0 0

2530

0 0 1 1 1 0 0 0 0 0

2531

0 0 1 1 1 0 0 0 0 0

2532

0 0 1 1 1 0 0 0 0 0

2533

S S S S S S S S S 0

2534

2535

2536

2537

## 2538 G.12 GPT-4.1 - 1 RANDOM CELL EASY: SAMPLE 2

2539

2540 **Prompt:** You are given a structural material distribution represented as a grid. Each cell can have one of  
 2541 the following states:

2542 - 'L' indicates applied load.

2543 - 'V' indicates void.

2544 - 'S' indicates support.

2545

2546 The goal is to predict the correct material distribution by filling in all 'V' cells with either '1' (solid)  
 2547 or '0' (empty), based on the surrounding structure and implicit physical reasoning (such as load paths,  
 2548 supports, and forces).

2549

2550 Important: The completed structure should use as little material as possible while remaining stable and  
 2551 plausible for carrying the applied forces. Minimize material usage unless necessary for structural support.

2551

2552 Below is the input grid with masked regions:

2553

2554 L L L L L L 0 0 0 0

2555 0 1 1 1 1 0 0 0 0 0

2556 0 0 1 1 0 0 0 0 0 0

2557 0 0 1 1 0 0 0 0 0 0

2558 0 0 1 1 1 0 0 0 0 0

2559 0 0 1 1 1 1 0 0 0 0

2560 V 0 0 1 1 1 1 0 0 0 0

2561 0 0 0 1 1 1 1 1 0 0

2562 0 0 0 0 1 1 1 1 0 0

2563 0 0 0 0 0 1 S S S 0

2564

2565 Please output the completed grid by replacing all 'V' cells with either '1' (solid) or '0' (empty).

2566 Maintain the same format as the input: one row per line, cells separated by spaces, and the total number of  
 2567 rows and columns unchanged.

2568 Return only the completed grid without any additional explanation.

2569 **Model Completion (differences in red):**

2570 L L L L L L 0 0 0 0

2571 0 1 1 1 1 0 0 0 0 0

2572 0 0 1 1 0 0 0 0 0 0

2573 0 0 1 1 0 0 0 0 0 0

2574 0 0 1 1 1 0 0 0 0 0

2575 0 0 1 1 1 1 0 0 0 0

2576 0 0 0 1 1 1 1 1 0 0

2577 0 0 0 0 0 1 S S S 0

2578 0 0 0 0 0 1 S S S 0

2579 0 0 0 0 0 1 S S S 0

2580 0 0 0 0 0 1 S S S 0

2581 0 0 0 0 0 1 S S S 0

2582 0 0 0 0 0 1 S S S 0

2583 0 0 0 0 0 1 S S S 0

2584 0 0 0 0 0 1 S S S 0

2585 0 0 0 0 0 1 S S S 0

2586 0 0 0 0 0 1 S S S 0

2587 0 0 0 0 0 1 S S S 0

2588 0 0 0 0 0 1 S S S 0

2589 0 0 0 0 0 1 S S S 0

2590 0 0 0 0 0 1 S S S 0

2591 0 0 0 0 0 1 S S S 0

2591

2592

2593

## 2592 G.13 GPT-4.1 - 1 RANDOM CELL HARD

2593

2594 **Prompt:** You are given a structural material distribution represented as a grid. Each cell can have one of  
 2595 the following states:

2596 - 'L' indicates applied load.

2597 - 'V' indicates void.

2598 - 'S' indicates support.

2599

2600 The goal is to predict the correct material distribution by filling in all 'V' cells with a floating point  
 2601 number between 0 and 1, with one decimal place (e.g., 0.0, 0.1, 0.2, ..., 1.0), based on the surrounding  
 2602 structure and implicit physical reasoning (such as load paths, supports, and forces).

2603

2604 Important: The completed structure should use as little material as possible while remaining stable and  
 2605 plausible for carrying the applied forces. Minimize material usage unless necessary for structural support.

2606

2607 Below is the input grid with masked regions:

2608

2609 0.0 0.0 0.0 L L L 0.0 0.0 0.0 0.0

2610 0.0 0.0 0.0 0.0 0.5 0.0 0.0 0.0 0.0 0.0

2611 0.0 0.0 0.0 0.0 0.5 0.0 0.0 0.0 0.0 0.0

2612 0.0 0.0 0.0 0.0 0.5 0.0 0.0 0.0 0.0 0.0

2613 0.0 0.0 V 0.0 0.5 0.0 0.0 0.0 0.0 0.0

2614 0.0 0.0 0.0 0.0 0.6 0.0 0.0 0.0 0.0 0.0

2615 0.0 0.1 0.6 0.6 0.6 0.0 0.0 0.0 0.0 0.0

2616 0.0 0.7 0.2 0.8 0.0 0.0 0.0 0.0 0.0 0.0

2617 S S S 0.1 0.0 0.0 0.0 0.0 0.0 0.0

2618

2619 Please output the completed grid by replacing all 'V' cells with a floating point number between 0 and 1,  
 2620 with one decimal place (e.g., 0.0, 0.1, 0.2, ..., 1.0).

2621

2622 Maintain the same format as the input: one row per line, cells separated by spaces, and the total number of  
 2623 rows and columns unchanged.

2624

2625 Return only the completed grid without any additional explanation.

2626

2627 **Model Completion (differences in red):**

2628 0.0 0.0 0.0 L L L 0.0 0.0 0.0 0.0

2629 0.0 0.0 0.0 0.0 0.5 0.0 0.0 0.0 0.0 0.0

2630 0.0 0.0 0.0 0.0 0.5 0.0 0.0 0.0 0.0 0.0

2631 0.0 0.0 0.0 0.0 0.5 0.0 0.0 0.0 0.0 0.0

2632 0.0 0.0 0.3 0.0 0.5 0.0 0.0 0.0 0.0 0.0

2633 0.0 0.0 0.0 0.0 0.6 0.0 0.0 0.0 0.0 0.0

2634 0.0 0.1 0.6 0.6 0.6 0.0 0.0 0.0 0.0 0.0

2635 0.0 0.7 0.2 0.8 0.0 0.0 0.0 0.0 0.0 0.0

2636 S S S 0.1 0.0 0.0 0.0 0.0 0.0 0.0

2637

2638 **Ground Truth:**

2639 0.0 0.0 0.0 L L L 0.0 0.0 0.0 0.0

2640 0.0 0.0 0.0 0.0 0.5 0.0 0.0 0.0 0.0 0.0

2641 0.0 0.0 0.0 0.0 0.5 0.0 0.0 0.0 0.0 0.0

2642 0.0 0.0 0.0 0.0 0.5 0.0 0.0 0.0 0.0 0.0

2643 0.0 0.0 0.0 0.0 0.5 0.0 0.0 0.0 0.0 0.0

2644 0.0 0.0 0.0 0.0 0.6 0.0 0.0 0.0 0.0 0.0

2645 0.0 0.1 0.6 0.6 0.6 0.0 0.0 0.0 0.0 0.0

0.0 0.7 0.2 0.8 0.0 0.0 0.0 0.0 0.0 0.0

S S S 0.1 0.0 0.0 0.0 0.0 0.0 0.0

2643

2644

2645

2646 G.14 GPT-4.1 - FULL EASY  
2647

2648 **Prompt:** You are given a structural material distribution represented as a grid. Each cell can have one of  
2649 the following states:  
2650 - 'L' indicates applied load.  
2651 - 'V' indicates void.  
2652 - 'S' indicates support.

2653 The goal is to predict the correct material distribution by filling in all 'V' cells with either '1' (solid)  
2654 or '0' (empty), based on the surrounding structure and implicit physical reasoning (such as load paths,  
2655 supports, and forces).  
2656

2657 Important: The completed structure should use as little material as possible while remaining stable and  
2658 plausible for carrying the applied forces. Minimize material usage unless necessary for structural support.  
2659

2660 Below is the input grid with masked regions:

2661  
2662 0 L L L L L 0 0 0 0  
2663 V V V V V V V V V V  
2664 V V V V V V V V V V  
2665 V V V V V V V V V V  
2666 V V V V V V V V V V  
2667 V V V V V V V V V V  
2668 V V V V V V V V V V  
2669 V V V V V V V V V V  
2670 0 S S S S S 0 0 0  
2671

2672 Please output the completed grid by replacing all 'V' cells with either '1' (solid) or '0' (empty).  
2673 Maintain the same format as the input: one row per line, cells separated by spaces, and the total number of  
2674 rows and columns unchanged.

2675 Return only the completed grid without any additional explanation.

2676 **Model Completion (differences in red):**

2677 0 L L L L L 0 0 0 0  
2678 1 1 1 1 1 1 1 0 0 0  
2679 1 0 0 0 0 0 1 0 0 0  
2680 1 0 0 0 0 0 1 0 0 0  
2681 1 0 0 0 0 0 1 0 0 0  
2682 1 0 0 0 0 0 1 0 0 0  
2683 1 0 0 0 0 0 1 0 0 0  
2684 1 1 1 1 1 1 1 0 0 0

2685 0 S S S S S 0 0 0 0

2686 **Ground Truth:**

2687 0 L L L L L 0 0 0 0  
2688 0 0 1 1 1 0 0 0 0 0  
2689 0 0 1 1 1 0 0 0 0 0  
2690 0 0 0 1 0 0 0 0 0 0  
2691 0 0 0 1 0 0 0 0 0 0  
2692 0 0 0 1 0 0 0 0 0 0  
2693 0 0 0 1 0 0 0 0 0 0  
2694 0 0 0 1 0 0 0 0 0 0  
2695 0 S S S S S 0 0 0 0

2696  
2697  
2698  
2699

## 2700 G.15 GPT-4.1 - FULL HARD

2701 **Prompt:** You are given a structural material distribution represented as a grid. Each cell can have one of  
 2702 the following states:  
 2703 - 'L' indicates applied load.  
 2704 - 'V' indicates void.  
 2705 - 'S' indicates support.

2706  
 2707 The goal is to predict the correct material distribution by filling in all 'V' cells with a floating point  
 2708 number between 0 and 1, with one decimal place (e.g., 0.0, 0.1, 0.2, ..., 1.0), based on the surrounding  
 2709 structure and implicit physical reasoning (such as load paths, supports, and forces).  
 2710

2711 **Important:** The completed structure should use as little material as possible while remaining stable and  
 2712 plausible for carrying the applied forces. Minimize material usage unless necessary for structural support.  
 2713

2714 Below is the input grid with masked regions:

2715  
 2716 0.0 L L L L L L L L L L  
 2717 V V V V V V V V V V  
 2718 V V V V V V V V V V  
 2719 V V V V V V V V V V  
 2720 V V V V V V V V V V  
 2721 V V V V V V V V V V  
 2722 V V V V V V V V V V  
 2723 V V V V V V V V V V  
 2724 0.0 0.0 S S S S S 0.0 0.0 0.0

2725  
 2726 Please output the completed grid by replacing all 'V' cells with a floating point number between 0 and 1,  
 2727 with one decimal place (e.g., 0.0, 0.1, 0.2, ..., 1.0).  
 2728 Maintain the same format as the input: one row per line, cells separated by spaces, and the total number of  
 2729 rows and columns unchanged.  
 2730 Return only the completed grid without any additional explanation.

2731 **Model Completion (differences in red):**

2732 0.0 L L L L L L L L L L  
 2733 0.0 0.4 0.6 0.8 1.0 1.0 0.8 0.6 0.4 0.0  
 2734 0.0 0.3 0.5 0.7 0.9 0.9 0.7 0.5 0.3 0.0  
 2735 0.0 0.2 0.4 0.6 0.8 0.8 0.6 0.4 0.2 0.0  
 2736 0.0 0.1 0.3 0.5 0.7 0.7 0.5 0.3 0.1 0.0  
 2737 0.0 0.1 0.2 0.3 0.5 0.5 0.3 0.2 0.1 0.0  
 2738 0.0 0.1 0.2 0.2 0.4 0.4 0.2 0.2 0.1 0.0  
 2739 0.0 0.1 0.1 0.1 0.3 0.3 0.1 0.1 0.1 0.0  
 2740 0.0 0.0 S S S S S 0.0 0.0 0.0

2741 **Ground Truth:**

2742 0.0 L L L L L L L L L L  
 2743 0.0 0.0 0.2 0.3 0.0 0.3 0.0 0.3 0.2 0.0  
 2744 0.0 0.0 0.0 0.4 0.2 0.3 0.2 0.4 0.0 0.0  
 2745 0.0 0.0 0.0 0.2 0.4 0.0 0.4 0.2 0.0 0.0  
 2746 0.0 0.0 0.0 0.0 0.4 0.0 0.4 0.0 0.0 0.0  
 2747 0.0 0.0 0.0 0.0 0.2 0.5 0.2 0.0 0.0 0.0  
 2748 0.0 0.0 0.0 0.0 0.0 0.6 0.0 0.0 0.0 0.0  
 2749 0.0 0.0 0.0 0.0 0.0 0.6 0.0 0.0 0.0 0.0  
 2750 0.0 0.0 S S S S S 0.0 0.0 0.0

2751  
 2752  
 2753

2754  
2755  
2756  
2757  
2758  
2759  
2760  
2761  
2762  
2763  
2764  
2765  
2766  
2767  
2768  
2769  
2770  
2771  
2772  
2773  
2774  
2775  
2776  
2777  
2778  
2779  
2780  
2781  
2782  
2783  
2784  
2785  
2786  
2787  
2788  
2789  
2790  
2791  
2792  
2793  
2794  
2795  
2796  
2797  
2798  
2799  
2800  
2801  
2802  
2803  
2804  
2805  
2806  
2807

**G.16 GEMINI 2.5 PRO - 1 RANDOM CELL EASY: SAMPLE 1**

**Prompt:** You are given a structural material distribution represented as a grid. Each cell can have one of the following states:  
 - 'L' indicates applied load.  
 - 'V' indicates void.  
 - 'S' indicates support.

The goal is to predict the correct material distribution by filling in all 'V' cells with either '1' (solid) or '0' (empty), based on the surrounding structure and implicit physical reasoning (such as load paths, supports, and forces).

Important: The completed structure should use as little material as possible while remaining stable and plausible for carrying the applied forces. Minimize material usage unless necessary for structural support.

Below is the input grid with masked regions:

```
0 0 L L L 0 0 0 0 0
0 0 0 1 0 0 0 0 0 0
0 0 0 V 0 0 0 0 0 0
0 0 0 1 0 0 0 0 0 0
0 0 0 1 0 0 0 0 0 0
0 0 0 1 0 0 0 0 0 0
0 0 1 1 1 0 0 0 0 0
0 0 1 1 1 0 0 0 0 0
0 0 1 1 1 0 0 0 0 0
S S S S S S S S S 0
```

Please output the completed grid by replacing all 'V' cells with either '1' (solid) or '0' (empty). Maintain the same format as the input: one row per line, cells separated by spaces, and the total number of rows and columns unchanged.

Return only the completed grid without any additional explanation.

**Model Completion (differences in red):**

```
0 0 L L L 0 0 0 0 0
0 0 0 1 0 0 0 0 0 0
0 0 0 1 0 0 0 0 0 0
0 0 0 1 0 0 0 0 0 0
0 0 0 1 0 0 0 0 0 0
0 0 0 1 0 0 0 0 0 0
0 0 0 1 0 0 0 0 0 0
0 0 0 1 0 0 0 0 0 0
0 0 1 1 1 0 0 0 0 0
0 0 1 1 1 0 0 0 0 0
0 0 1 1 1 0 0 0 0 0
0 0 1 1 1 0 0 0 0 0
0 0 1 1 1 0 0 0 0 0
```

**Ground Truth:**

```
0 0 L L L 0 0 0 0 0
0 0 0 1 0 0 0 0 0 0
0 0 0 1 0 0 0 0 0 0
0 0 0 1 0 0 0 0 0 0
0 0 0 1 0 0 0 0 0 0
0 0 0 1 0 0 0 0 0 0
0 0 0 1 0 0 0 0 0 0
0 0 1 1 1 0 0 0 0 0
0 0 1 1 1 0 0 0 0 0
0 0 1 1 1 0 0 0 0 0
S S S S S S S S S 0
```



## 2808 G.17 GEMINI 2.5 PRO - 1 RANDOM CELL EASY: SAMPLE 2

2809

2810 **Prompt:** You are given a structural material distribution represented as a grid. Each cell can have one of  
 2811 the following states:

2812 - 'L' indicates applied load.

2813 - 'V' indicates void.

2814 - 'S' indicates support.

2815

2816 The goal is to predict the correct material distribution by filling in all 'V' cells with either '1' (solid)  
 2817 or '0' (empty), based on the surrounding structure and implicit physical reasoning (such as load paths,  
 2818 supports, and forces).

2819

2820 Important: The completed structure should use as little material as possible while remaining stable and  
 2821 plausible for carrying the applied forces. Minimize material usage unless necessary for structural support.

2821

2822 Below is the input grid with masked regions:

2823

2824 L L L L L L 0 0 0 0

2825 0 1 1 1 1 0 0 0 0 0

2826 0 0 1 1 0 0 0 0 0 0

2827 0 0 1 1 0 0 0 0 0 0

2828 0 0 1 1 1 0 0 0 0 0

2829 0 0 1 1 1 1 0 0 0 0

2830 V 0 0 1 1 1 1 0 0 0 0

2831 0 0 0 1 1 1 1 1 0 0

2832 0 0 0 0 1 1 1 1 0 0

2833 0 0 0 0 0 1 S S S 0

2834

2835 Please output the completed grid by replacing all 'V' cells with either '1' (solid) or '0' (empty).

2836 Maintain the same format as the input: one row per line, cells separated by spaces, and the total number of  
 2837 rows and columns unchanged.

2838 Return only the completed grid without any additional explanation.

2839 **Model Completion (differences in red):**

2840 L L L L L L 0 0 0 0

2841 0 1 1 1 1 0 0 0 0 0

2842 0 0 1 1 0 0 0 0 0 0

2843 0 0 1 1 0 0 0 0 0 0

2844 0 0 1 1 1 0 0 0 0 0

2845 0 0 0 1 1 1 1 1 0 0

2846 0 0 0 0 1 1 1 1 0 0

2847 0 0 0 0 0 1 S S S 0

2848 **Ground Truth:**

2849 L L L L L L 0 0 0 0

2850 0 1 1 1 1 0 0 0 0 0

2851 0 0 1 1 0 0 0 0 0 0

2852 0 0 1 1 0 0 0 0 0 0

2853 0 0 1 1 1 0 0 0 0 0

2854 0 0 0 1 1 1 1 1 0 0

2855 0 0 0 0 1 1 1 1 0 0

2856 0 0 0 0 0 1 S S S 0

2857

2858

2859

2860

2861

## 2862 G.18 GEMINI 2.5 PRO - 1 RANDOM CELL HARD

2863

2864 **Prompt:** You are given a structural material distribution represented as a grid. Each cell can have one of  
 2865 the following states:

2866 - 'L' indicates applied load.

2867 - 'V' indicates void.

2868 - 'S' indicates support.

2869

2870 The goal is to predict the correct material distribution by filling in all 'V' cells with a floating point  
 2871 number between 0 and 1, with one decimal place (e.g., 0.0, 0.1, 0.2, ..., 1.0), based on the surrounding  
 2872 structure and implicit physical reasoning (such as load paths, supports, and forces).

2873

2874 Important: The completed structure should use as little material as possible while remaining stable and  
 2875 plausible for carrying the applied forces. Minimize material usage unless necessary for structural support.

2876

2877 Below is the input grid with masked regions:

2878

```
2878 0.0 L V L L L L L L 0.0
2879 0.0 0.0 0.2 0.3 0.1 0.1 0.3 0.2 0.0 0.0
2880 0.0 0.0 0.1 0.4 0.0 0.0 0.4 0.1 0.0 0.0
2881 0.0 0.0 0.0 0.4 0.0 0.0 0.4 0.0 0.0 0.0
2882 0.0 0.0 0.0 0.4 0.0 0.0 0.4 0.0 0.0 0.0
2883 0.0 0.0 0.0 0.4 0.0 0.0 0.4 0.0 0.0 0.0
2884 0.0 0.0 0.0 0.4 0.1 0.1 0.4 0.0 0.0 0.0
2885 0.0 0.0 0.0 0.4 0.1 0.1 0.4 0.0 0.0 0.0
2886 0.0 S S S S S S S S 0.0
```

2887

2888 Please output the completed grid by replacing all 'V' cells with a floating point number between 0 and 1,  
 2889 with one decimal place (e.g., 0.0, 0.1, 0.2, ..., 1.0).

2890

2891 Maintain the same format as the input: one row per line, cells separated by spaces, and the total number of  
 2892 rows and columns unchanged.

2892

2893 Return only the completed grid without any additional explanation.

2894 **Model Completion (differences in red):**

2895

```
2895 0.0 L 0.2 L L L L L L 0.0
2896 0.0 0.0 0.2 0.3 0.1 0.1 0.3 0.2 0.0 0.0
2897 0.0 0.0 0.1 0.4 0.0 0.0 0.4 0.1 0.0 0.0
2898 0.0 0.0 0.0 0.4 0.0 0.0 0.4 0.0 0.0 0.0
2899 0.0 0.0 0.0 0.4 0.0 0.0 0.4 0.0 0.0 0.0
2900 0.0 0.0 0.0 0.4 0.1 0.1 0.4 0.0 0.0 0.0
2901 0.0 0.0 0.0 0.4 0.1 0.1 0.4 0.0 0.0 0.0
2902 0.0 S S S S S S S S 0.0
```

2903

2904 **Ground Truth:**

2905

```
2905 0.0 L L L L L L L L 0.0
2906 0.0 0.0 0.2 0.3 0.1 0.1 0.3 0.2 0.0 0.0
2907 0.0 0.0 0.1 0.4 0.0 0.0 0.4 0.1 0.0 0.0
2908 0.0 0.0 0.0 0.4 0.0 0.0 0.4 0.0 0.0 0.0
2909 0.0 0.0 0.0 0.4 0.0 0.0 0.4 0.0 0.0 0.0
2910 0.0 0.0 0.0 0.4 0.1 0.1 0.4 0.0 0.0 0.0
2911 0.0 0.0 0.0 0.4 0.1 0.1 0.4 0.0 0.0 0.0
2912 0.0 S S S S S S S S 0.0
```

2913

2914

2915

2916 **G.19 GEMINI 2.5 PRO - FULL EASY**

2917

2918 **Prompt:** You are given a structural material distribution represented as a grid. Each cell can have one of

2919 the following states:

2920 - 'L' indicates applied load.

2921 - 'V' indicates void.

2922 - 'S' indicates support.

2923

2924 The goal is to predict the correct material distribution by filling in all 'V' cells with either '1' (solid)

2925 or '0' (empty), based on the surrounding structure and implicit physical reasoning (such as load paths,

2926 supports, and forces).

2927

2928 **Important:** The completed structure should use as little material as possible while remaining stable and

2929 plausible for carrying the applied forces. Minimize material usage unless necessary for structural support.

2930

2931 Below is the input grid with masked regions:

2932

```

2933 0 0 0 L L L L 0 0 0
2934 V V V V V V V V V
2935 V V V V V V V V V
2936 V V V V V V V V V
2937 V V V V V V V V V
2938 V V V V V V V V V
2939 V V V V V V V V V
2940 0 0 0 0 S S S 0 0 0

```

2941

2942 Please output the completed grid by replacing all 'V' cells with either '1' (solid) or '0' (empty).

2943 Maintain the same format as the input: one row per line, cells separated by spaces, and the total number of

2944 rows and columns unchanged.

2945 Return only the completed grid without any additional explanation.

2946 **Model Completion (differences in red):**

```

2947 0 0 0 L L L L 0 0 0
2948 0 0 0 1 1 1 1 0 0 0
2949 0 0 0 0 1 1 1 0 0 0
2950 0 0 0 0 1 1 1 0 0 0
2951 0 0 0 0 1 1 1 0 0 0
2952 0 0 0 0 1 1 1 0 0 0
2953 0 0 0 0 1 1 1 0 0 0
2954 0 0 0 0 1 1 1 0 0 0
2955 0 0 0 0 S S S 0 0 0

```

2956 **Ground Truth:**

```

2957 0 0 0 L L L L 0 0 0
2958 0 0 0 0 1 1 0 0 0 0
2959 0 0 0 0 1 1 0 0 0 0
2960 0 0 0 0 1 1 0 0 0 0
2961 0 0 0 0 1 1 0 0 0 0
2962 0 0 0 0 1 1 0 0 0 0
2963 0 0 0 0 1 1 0 0 0 0
2964 0 0 0 0 1 1 0 0 0 0
2965 0 0 0 0 S S S 0 0 0

```

2966

2967

2968

2969

2970 **G.20 GEMINI 2.5 PRO - FULL HARD**

2971

2972 **Prompt:** You are given a structural material distribution represented as a grid. Each cell can have one of

2973 the following states:

2974 - 'L' indicates applied load.

2975 - 'V' indicates void.

2976 - 'S' indicates support.

2977

2978 The goal is to predict the correct material distribution by filling in all 'V' cells with a floating point

2979 number between 0 and 1, with one decimal place (e.g., 0.0, 0.1, 0.2, ..., 1.0), based on the surrounding

2980 structure and implicit physical reasoning (such as load paths, supports, and forces).

2981 **Important:** The completed structure should use as little material as possible while remaining stable and

2982 plausible for carrying the applied forces. Minimize material usage unless necessary for structural support.

2983

2984 Below is the input grid with masked regions:

2985

```
2986 0.0 0.0 L L L 0.0 0.0 0.0 0.0 0.0
2987 V V V V V V V V V V
2988 V V V V V V V V V V
2989 V V V V V V V V V V
2990 V V V V V V V V V V
2991 V V V V V V V V V V
2992 V V V V V V V V V V
2993 V V V V V V V V V V
2994 S S S S S S S S 0.0
```

2995

2996 Please output the completed grid by replacing all 'V' cells with a floating point number between 0 and 1,

2997 with one decimal place (e.g., 0.0, 0.1, 0.2, ..., 1.0).

2998 Maintain the same format as the input: one row per line, cells separated by spaces, and the total number of

2999 rows and columns unchanged.

3000 Return only the completed grid without any additional explanation.

3001 **Model Completion (differences in red):**

```
3002 0.0 0.0 L L L 0.0 0.0 0.0 0.0 0.0
3003 0.0 0.1 1.0 1.0 1.0 0.1 0.0 0.0 0.0 0.0
3004 0.0 0.2 1.0 1.0 1.0 0.2 0.0 0.0 0.0 0.0
3005 0.1 0.3 1.0 1.0 1.0 0.3 0.1 0.0 0.0 0.0
3006 0.2 0.4 1.0 1.0 1.0 0.4 0.2 0.0 0.0 0.0
3007 0.3 0.5 1.0 1.0 1.0 0.5 0.3 0.1 0.0 0.0
3008 0.4 0.6 1.0 1.0 1.0 0.6 0.4 0.2 0.0 0.0
3009 0.5 0.7 1.0 1.0 1.0 0.7 0.5 0.3 0.1 0.0
3010 0.6 0.8 1.0 1.0 1.0 0.8 0.6 0.4 0.2 0.0
3011 S S S S S S S S 0.0
```

3012 **Ground Truth:**

```
3013 0.0 0.0 L L L 0.0 0.0 0.0 0.0 0.0
3014 0.0 0.0 0.0 0.9 0.0 0.0 0.0 0.0 0.0 0.0
3015 0.0 0.0 0.0 0.8 0.0 0.0 0.0 0.0 0.0 0.0
3016 0.0 0.0 0.0 0.8 0.0 0.0 0.0 0.0 0.0 0.0
3017 0.0 0.0 0.0 0.9 0.0 0.0 0.0 0.0 0.0 0.0
3018 0.0 0.0 0.1 0.8 0.1 0.0 0.0 0.0 0.0 0.0
3019 0.0 0.0 0.1 0.8 0.1 0.0 0.0 0.0 0.0 0.0
3020 0.0 0.0 0.2 0.8 0.2 0.0 0.0 0.0 0.0 0.0
3021 S S S S S S S S 0.0
```

3022

3023